

Aus dem Universitätsklinikum Münster

Institut für Hygiene

- Direktor: Univ.-Prof. Dr. rer. nat. Dr. h. c. Helge Karch -

**ReSeqLIMS: Laborinformations- und Managementsystem
für die Hochdurchsatz-Resequenzierung**

INAUGURAL - DISSERTATION

zur

Erlangung des doctor rerum medicinalium

der Medizinischen Fakultät

der Westfälischen Wilhelms-Universität Münster

vorgelegt von David Johannes Schmitz-Hübsch
aus Steinfurt

2012

Gedruckt mit Genehmigung der Medizinischen Fakultät der Westfälischen Wilhelms-
Universität Münster

Dekan: Univ.-Prof. Dr. med. Wilhelm Schmitz

1. Berichterstatter: Univ.-Prof. Dr. med. Dag Harmsen
2. Berichterstatter: Univ.-Prof. Dr. rer. nat. Dr. h. c. Helge Karch

Tag der mündlichen Prüfung: 05.07.2012

Aus dem Universitätsklinikum Münster
Institut für Hygiene

- Direktor: Univ.-Prof. Dr. rer. nat. Dr. h. c. Helge Karch -
Referent: Univ.-Prof. Dr. med. Dag Harmsen
Koreferent: Univ.-Prof. Dr. rer. nat. Dr. h. c. Helge Karch

Zusammenfassung

ReSeqLIMS: Laborinformations- und Managementsystem
für die Hochdurchsatz-Resequenzierung

Schmitz-Hübsch, David Johannes

Die molekulare Typisierung von Mikroorganismen ist zur Erkennung von Infektionsausbrüchen und langfristiger evolutionärer Trends von entscheidender Bedeutung. Hierzu haben sich in den letzten Jahren zunehmend Methoden basierend auf Amplikon-Resequenzierungen etabliert. Neben den laborseitigen Anforderungen für die Hochdurchsatz-Resequenzierung stehen insbesondere leistungsfähige Laborinformations- und Managementsysteme (LIMS) im Fokus, um die Laborabläufe sicher zu verwalten und die Daten für weiterführende Analysen bereitzustellen.

Da gegenwärtig für die Hochdurchsatz-Resequenzierung nur wenige und sehr teure LIMS verfügbar sind, wurde im Rahmen der vorliegenden Arbeit das so genannte ReSeqLIMS programmiert. Dieses ReSeqLIMS wurde im Rahmen des vom BMBF geförderten Projektes Phylogenie, Bio-Informatik und Amplikon-Resequenzierung von Zoonose-Erregern (PBA-Zoo) als Webanwendung in einer Drei-Schichten-Architektur implementiert. Die Präsentationsschicht wurde mit Hilfe von JavaServer Faces (JSF), JavaScript und Facelets, die Logikschicht mittels Java programmiert. Zudem wurde Hibernate als Middleware zwischen Logikschicht und Datenbank verwendet, um eine Datenbankunabhängigkeit zu garantieren. Hier wurde eine PostgreSQL-Datenbank als Datenbasis für das ReSeqLIMS ausgewählt.

Die Funktionalität des ReSeqLIMS kann in Module differenziert werden. Das Kundenmodul umfasst das Anwender-, Konditions- und Rechnungsmanagement und ermöglicht u.a. die Registrierung und Auftragsgenerierung, automatisiert die E-Mail-Kommunikation und stellt Ergebnisse zum Download zur Verfügung. Das Workflowmanagement ist Teil des Labormoduls. Hier sind die einzelnen Arbeitsschritte in einen qualitätsüberwachten Workflow integriert, in dem gerätespezifische Arbeits- und Ausgabelisten verarbeitet und Proben inkl. der Verbrauchsmaterialien protokolliert werden. Gerätespezifische Parameter wie z.B. die Anzahl der Kapillaren des DNA-Sequenzierers oder die Arbeits- und Ausgabelisten des Pipettierroboters werden im Gerätemodul verwaltet.

Mit dem ReSeqLIMS ist es möglich, den Datenfluss während der verschiedenen Laborprozesse im Hochdurchsatz zu steuern und die Resequenzierungsdaten für nachfolgende Downstream-Analysen bereitzustellen. Die ReSeqLIMS-Webanwendung ist unter der Adresse <http://www.pba-zoo.de> zu erreichen.

Tag der mündlichen Prüfung: 05.07.2012

Inhaltsverzeichnis

1	Einleitung	1
2	Material und Methoden	9
2.1	Webanwendung und Java	9
2.2	Softwareentwicklung	11
2.2.1	Grundlagen der Softwareentwicklung	11
2.2.2	Die Datenhaltungsschicht	13
2.2.3	Die Logikschicht	16
2.2.4	Die Präsentationsschicht	16
2.3	Entwicklungsumgebung und Versionshinweise verwendeter Software und Bibliotheken für die ReSeqLIMS-Implementierung	20
3	Ergebnisse	22
3.1	Das ReSeqLIMS-Datenmodell	22
3.1.1	Datenmodelle des ReSeqLIMS-Labormoduls	22
3.1.2	Datenmodelle des ReSeqLIMS-Kundenmoduls	31
3.1.3	Datenmodell des ReSeqLIMS-Gerätemoduls	37
3.1.4	Modulübergreifende Datenmodelle	38
3.2	Das ReSeqLIMS	40
3.2.1	Die ReSeqLIMS-Webanwendung	40
3.2.2	Das ReSeqLIMS-Kundenmodul	42
3.2.3	Das ReSeqLIMS-Labormodul	49
3.2.4	Das ReSeqLIMS-Gerätemodul	59
3.2.5	Die ReSeqLIMS-Statusberechnung	59
3.2.6	Die ReSeqLIMS-Dokumente	60
4	Diskussion	67
5	Literaturverzeichnis	78
6	Abbildungsverzeichnis	85
7	Tabellenverzeichnis	88
8	Abkürzungsverzeichnis	89
	Lebenslauf	91
	Danksagung	92
	Anhang	I

1 Einleitung

Nach Schätzungen der World Health Organization (WHO) sind ca. 25% der weltweiten jährlichen Todesfälle des Menschen auf Infektionskrankheiten zurückzuführen [1]. Durch das Anhaften oder Eindringen von Erregern in den Wirtsorganismus und der dortigen Vermehrung entstehen Infekte, die bei Vorliegen von klinischen Symptomen wie z.B. Fieber oder Eiter als Infektionskrankheiten bezeichnet werden [29]. Zu den Erregern von Infektionskrankheiten zählen Mikroorganismen, die u.a. in Bakterien, Pilze, Viren, Würmer, Prionen und Protozoen eingeteilt werden können. Etwa zwei Drittel aller Infektionskrankheiten sind Zoonosen, die von Erregern durch die wechselseitige Übertragung zwischen Tier und Mensch verursacht werden [1]. Als präventive Maßnahmen gegen Infektionskrankheiten haben sich Hygienestandards und Impfungen, soweit diese für einen Erreger vorhanden sind, etabliert. Liegt bereits eine Infektion vor werden Antibiotika (gegen Bakterien), Antimykotika (gegen Pilze) oder Virostatika (gegen Viren) zur Behandlung eingesetzt. Für die Bestimmung einer geeigneten Therapie ist eine genaue Diagnose mit der Bestimmung des Erregers und seiner Eigenschaften von grundlegender Bedeutung. Die Entwicklung präventiver Maßnahmen, Therapien und Wirkstoffe erfordert detaillierte Kenntnisse sowohl der Anpassungsmechanismen der Erreger im Wirt und auf spezifische Wirkstoffe als auch der Infektionswege.

Die vermehrte Personenmobilität und der großräumige Transport von Lebensmitteln und Tieren tragen dazu bei, dass sich Erreger extrem schnell verteilen. In Krankenhäusern treten Erreger wie z.B. Methicillin-resistente *Staphylococcus aureus* (MRSA) [27,38] auf, die bei Patienten zu lebensgefährlichen Infektionen führen können und durch Dritte (z.B. Pfleger, Ärzte und Besucher) übertragen werden. Im Falle des Ausbruchs einer Infektionskrankheit müssen der verursachende Erregerstamm und dessen Infektionswege schnellstmöglich zurückverfolgt werden, um geeignete Maßnahmen ergreifen zu können. Zudem hilft die Kenntnis der Infektionswege und -mechanismen

bei der Entwicklung neuer präventiver Maßnahmen. In diesem Zusammenhang ist die ausschließliche Identifizierung des Erregers nicht ausreichend, da sich die Erregerstämme einer Art erheblich, u.a. in ihrem Wirtsspektrum, den durch sie verursachenden Symptomen und Krankheitsverläufen und der Inkubationszeit unterscheiden. Um aber Infektionswege verfolgen und Maßnahmen ergreifen zu können, ist daher die weitere Subtypisierung des Erregerstammes meist mittels genotypischer Verfahren erforderlich. Die Methoden der Genotypisierung generieren durch die Analyse des Genoms oder einzelner Abschnitte hieraus ein für den jeweiligen Erreger eindeutiges, charakteristisches Ergebnis, den so genannten genetischen Fingerabdruck. Dadurch können primär lokale Geschehen wie z.B. Infektionsausbrüche innerhalb eines kurzen Zeitraums identifiziert und die Infektionswege verfolgt werden [22]. Die Protokollierung periodisch auftretender Infektionsausbrüche ermöglicht die Vorhersage längerfristiger Entwicklungstrends über Jahre bzw. Jahrzehnte und eine Risikoabschätzung über die Verbreitung und Pathogenitätsentwicklung von Krankheitserregern [2,66].

Seit Mitte der 80er Jahre des vergangenen Jahrhunderts wurden Genotypisierungen v.a. mit der Pulsed Field Gelelectrophoresis (PFGE)-Methode [57] durchgeführt, die sich zu einem de-facto Standard entwickelt hatte. Hierbei wird das Genom mit spezifischen, selten schneidenden Restriktionsenzymen in DNA-Fragmente zerstückelt, die auf einem Agarosegel entsprechend ihrer Länge aufgetrennt werden [17] und das entstehende Bandenmuster anschließend als Charakteristikum für einen Erreger ausgewertet. Während Banden der Gelelektrophorese für DNA-Fragmente ab einer Länge von 30 Kilobasenpaaren (kb) eine einzige dicke Bande zeigen, können mit der PFGE DNA-Fragmente bis zu einer Länge von 2.000 kb [57] sehr gut aufgelöst und damit eine sehr hohe Auflösung von Arten und Subgruppen erzielt werden [22]. Die PFGE-Methode ist jedoch labortechnisch sehr aufwendig und zeitintensiv, weshalb sie für den klinischen Alltag nur bedingt geeignet ist. Hinzu kommt, dass Methoden, die auf der Erstellung eines durch Gelelektrophorese generierten genetischen Fingerabdruckes basieren, schwierig zu standardisieren sind und die Ergebnisse über die Zeit und damit sowohl innerhalb eines Labores als auch zwischen verschiedenen Laboren nur schwierig zu reproduzieren und zu vergleichen sind [2,17,45].

Für eine exakte Bestimmung des Genotyps eines Erregers bzw. zur Identifizierung der Unterschiede zwischen Genomen, wäre im optimalen Fall die DNA-Sequenzierung vollständiger Genome erforderlich. Im klinischen Alltag ist die Bestimmung des Geno-

typs eines Erregers auf Basis der DNA-Sequenzierung eines vollständigen Genoms aus zeitlichen, technologischen und wirtschaftlichen Gründen bisher noch nicht praktikabel. Aktuelle Methoden der Genotypisierung, die eine DNA-Sequenzierung beinhalten, beschränken sich stattdessen auf wenige Gene bzw. Genfragmente. Die Multilocus Sequence Typing (MLST)-Methode definiert ca. 450 Basenpaare (bp) lange DNA-Fragmente von i.d.R. sieben Haushaltsgenen, die zur Genotypisierung von Bakterien verwendet werden [15,17,22,73]. Haushaltsgene sind nicht regulative Gene, die unabhängig vom Zelltyp und von äußeren Einflüssen exprimiert werden, da sie für grundlegende metabolische Funktionen eines Erregers verantwortlich sind. Die Anzahl der Haushaltsgene und welche Haushaltsgene für die MLST-Methode verwendet werden ist abhängig von der zu analysierenden Spezies [17]. Nach der Amplifizierung der DNA-Fragmente mittels der Polymerase Chain Reaction (PCR) und der anschließenden DNA-Sequenzierung werden die Konsensussequenzen der jeweiligen DNA-Fragmente bestimmt. Die Konsensussequenzen definieren die Allele der analysierten Gene. Durch den Vergleich der Konsensussequenzen mit Einträgen in MLST-Referenzdatenbanken bereits sequenzierter und dokumentierter Allele, die jeweils mit einer Allelnummer abgespeichert wurden, ergibt sich für den analysierten Organismus das Allelprofil, eine Kombination von z.B. sieben Allelnummern. Der MLST-Sequenztyp, der zur Aufklärung der verwandtschaftlichen Beziehungen zwischen Bakterienarten und zur Differenzierung von Bakterien-Subgruppen verwendet werden kann wird wiederum durch das Allelprofil definiert [22]. Neue Allele oder Sequenztypen werden nach einer Verifikation in die Referenzdatenbanken übernommen. Die standardisierten Protokolle der MLST-Methode, die Verwendung der DNA-Sequenz anstelle von Bandenmustern auf einem Agarosegel und die Normierung der Ergebnisse auf der Basis von Referenzdatenbanken machen die Ergebnisse der MLST-Methode reproduzierbar und innerhalb sowie zwischen verschiedenen Laboren vergleichbar [17,41]. Ergebnisse von Methoden, die auf der Aufklärung der DNA-Sequenz basieren, sind aufgrund des aus A, C, G und T bestehenden DNA-Alphabets, das die vier DNA-Basen Adenin, Cytosin, Guanin und Thymin kodiert, einfach in Referenzdatenbanken zu hinterlegen und zu verwalten. Die MLST-Methode ist durch die Verwendung der PCR und der DNA-Sequenzierung weitestgehend automatisierbar und wird als Hochdurchsatz-Methode im klinischen Alltag eingesetzt. Durch die Fokussierung der MLST-Methode auf Haushaltsgene und

nicht auf Gene, die mit der Virulenz der Erreger in Verbindung gebracht werden können, ist die Auflösung von Subgruppen jedoch begrenzt [41].

Einen anderen Ansatz als die MLST-Methode verfolgt die Multi Loci Variable Number Tandem Repeat (MLVA)-Methode [10,63], die auf dem Phänomen basiert, dass eine Vielzahl von kurzen Nukleotidabfolgen, die Minisatelliten mit einer Länge von ca. 10 bp bis 100 bp, im Genom als tandemartige Wiederholungen (Tandem-Repeats) auftreten. Die Anzahl der Wiederholungen kann zur Differenzierung von Arten und Subgruppen von Bakterien verwendet werden und lässt sich gut in Referenzdatenbanken verwalten [3,72]. Im Vergleich mit der PFGE-Methode ist die MLVA-Methode wesentlich einfacher und weniger aufwendig durchzuführen. Ein Nachteil der MLVA-Methode ergibt sich aus den hohen Mutationsraten bei Bakterien und den damit verbundenen Homoplasien, die insbesondere bei der Erstellung von phylogenetischen Stammbäumen zu Fehlern führen können [2]. Zusätzlich sind die Ergebnisse der MLVA-Methode z.T. schwierig zu reproduzieren und zwischen Laboren zu vergleichen, da nur die Länge und nicht die DNA-Sequenz der Tandem-Repeats bestimmt wird. Während die MLST-Methode die DNA-Sequenz und die MLVA-Methode die Anzahl der Tandem-Repeats spezifischer DNA-Fragmente für eine Genotypisierung analysiert, basiert die Single Nucleotide Polymorphisms (SNPs)-Methode auf der Detektion einzelner Nukleotidvariationen in einem Genom [41,43,52]. SNPs sind in großer Anzahl ungleichmäßig, aber in bestimmten Regionen konzentriert auftretend, über das gesamte Genom eines Individuums verteilt. In SNP-Analysen werden Nukleotidvariationen aus kodierenden und nicht-kodierenden DNA-Sequenzen berücksichtigt. Der Vorteil der SNP-Analysen ergibt sich aus der einfachen und schnellen Bestimmung sowie der guten Verwaltbarkeit der SNPs in Referenzdatenbanken. Die Verwendung der beschriebenen Methoden zur Genotypisierung ist abhängig von der Fragestellung und der zu analysierenden Spezies.

Das wiederholte Sequenzieren von bereits sequenzierten Genomen bzw. von DNA-Fragmenten eines Organismus, z.B. in MLST- oder SNP-Analysen, wird häufig als Resequenzierung bezeichnet. Während in den Anfängen der DNA-Sequenzierung die Aufklärung der Genomsequenz eines Organismus das primäre Ziel war, zeigten Forschungsarbeiten schon früh, dass das Sequenzieren eines Genoms zur Charakterisierung einer Art nicht ausreichend ist. Je nach analysierter Subgruppe ergeben sich z.T. erhebliche Unterschiede in der DNA-Sequenz der Genome und der genetischen Ausstattung.

Um die Funktion der Gene eines Erregers, dessen Anpassungsmechanismen oder Reaktionen auf verschiedene Substanzen verstehen zu können, ist daher das Pangenom, das die gesamte genetische Information einer Art umfasst, entscheidend [41]. Dies ist zumindest approximativ nur durch umfassende Resequenzierungsprojekte zu erhalten, die in Variationsanalysen und Genotypisierungen differenziert werden können. Während in Variationsanalysen bisher nicht bekannte Mutationen bzw. DNA-Sequenzunterschiede wie z.B. SNPs und InDels (Insertionen und Deletionen) analysiert werden, werden mit Methoden der Genotypisierung, die auf einer Resequenzierung basieren (z.B. MLST- oder SNP-Analysen), bekannte Mutationen identifiziert. Die Ergebnisse von Methoden, die auf der DNA-Sequenzierung basieren, sind einfach zu vergleichen, zu reproduzieren und in Referenzdatenbanken zu verwalten. Der Goldstandard der DNA-Sequenzierung ist die DNA-Sequenzierung nach Sanger, die qualitativ hohe DNA-Fragmente mit einer Länge von bis zu 1.100 bp sequenzieren kann [14,55]. Die DNA-Fragmentlänge der neuen DNA-Sequenzierungsverfahren wie z.B. der Pyrosequenzierung [40,53], der Sequencing-By-Ligation [59] oder der Sequencing-By-Synthesis [11] sind mit ca. 450 bp bei der Pyrosequenzierung und ca. 100 bp bzw. ca. 25-75 bp bei den anderen Verfahren deutlich kürzer. Aufgrund des deutlich höheren Durchsatzes pro Sequenzierungslauf und den damit verbundenen geringeren Kosten pro DNA-Base im Vergleich zur DNA-Sequenzierung nach Sanger [18,37] sind die neuen Verfahren aber insbesondere für Resequenzierungsprojekte, in denen die DNA-Sequenz bereits sequenzierter Erreger als Referenzsequenz vorliegt, sehr gut geeignet [11].

Seit Anfang 2008 wird das Institut für Hygiene des Universitätsklinikum Münster (UKM) vom Bundesministerium für Bildung und Forschung (BMBF) bei der Etablierung einer interdisziplinären Bioinformatikplattform im Rahmen des Phylogenie, Bioinformatik und Amplikon-Resequenzierung von Zoonose-Erregern (PBA-Zoo) Projektes gefördert. Zentraler Bestandteil des PBA-Zoo Projektes ist der Aufbau eines Amplicon Re-Sequencing Centers, das Netzwerkpartner bei der Planung und Durchführung von Resequenzierungsprojekten, z.B. mit der MLST-Methode, unterstützt. Aufgrund des zu erwartenden hohen Probenaufkommens wurde zur Optimierung der Sequenzierungsabläufe und zur Automatisierung von phylogenetischen und funktionellen Resequenzierungsfragestellungen ein Pipettierroboter in den Laborworkflow integriert. Bei den erforderlichen Restrukturierungsmaßnahmen, die zum Aufbau eines Hochdurchsatzlabores notwendig waren, musste der Ausbau und die Implementierung einer leistungsfähigen

gen Informationssystem-Infrastruktur berücksichtigt werden. Die Informationssystem-Infrastruktur umfasst die Hardware, Software, Netzwerke, Daten, Dienstleistungen und die menschliche Arbeitskraft, die erforderlich ist, um z.B. in einem Labor die für das Erreichen definierter Ziele erforderlichen Prozesse informationstechnisch bestmöglichst zu unterstützen, zu automatisieren und zu steuern, um sie effizient und effektiv durchführen zu können. Die Zentrale Komponente in der Informationssystem-Infrastruktur des PBA-Zoo Projektes ist ein Laborinformations- und Managementsystem (LIMS).

An ein LIMS im Kontext der Resequenzierung und insbesondere an das LIMS im Rahmen des PBA-Zoo Projektes lassen sich spezifische Anforderungen definieren. Zur DNA-Sequenzierung eines einzelnen Genomes im Shotgun-Sequencing Ansatz wird die DNA-Probe kloniert und anschließend sequenziert. Für die Qualität der durch Fragment-Assembling hergestellten Konsensussequenz ist die Qualität der Ausgangsproben und die Qualität bzw. Anzahl der sequenzierten DNA-Fragmente entscheidend. Hier ist eine Verfolgung und Protokollierung der einzelnen DNA-Proben bzw. der DNA-Fragmente innerhalb der Prozesse des Laborworkflows nicht zwingend notwendig. In Resequenzierungsprojekten liegt dagegen i.d.R. eine Vielzahl von DNA-Proben vor, die z.B. mit der MLST-Methode analysiert werden sollen. Für jedes der z.B. sieben definierten Haushaltsgene wird während der PCR aus einer DNA-Probe ein genspezifischer Probenansatz hergestellt. Für die DNA-Sequenzierung werden die genspezifischen Probenansätze in zwei Probenansätze aufgeteilt, um je einen Probenansatz mit Forward-Primer und einen Probenansatz mit Reverse-Primer herzustellen. Im Ergebnis generiert die MLST-Methode mit sieben zu analysierenden Haushaltsgenen für jede Ausgangsprobe vierzehn Chromatogramme, die den Ausgangsproben und den entsprechenden Genen eindeutig zuzuordnen sind. Zur Einhaltung der im Labor geltenden Normen und Standards sowie zur Sicherstellung der Qualität der Ergebnisse, sind auf sämtlichen Stufen des Laborworkflows die verwendeten Geräte, Reagenzien und Verbrauchsmaterialien für jede einzelne Probe zu protokollieren. Werden während der Probenbearbeitung qualitative Mängel festgestellt, können anhand der vorliegenden Qualitätsdaten die Ursachen identifiziert und ausgewählte Proben ab einer spezifischen Stufe im Laborworkflow wiederholt werden. Daher ist im Gegensatz zum Shotgun-Sequencing Ansatz im Rahmen von Gesamtgenomsequenzierungen bei der Durchführung von Resequenzierungen einzelner DNA-Fragmente die eindeutige Identifizierung, Protokollierung und Rückverfolgung der Proben über sämtliche Stufen des Laborworkflows zwingend erforder-

derlich. Eine zusätzliche besondere Anforderung an das LIMS im Rahmen des PBA-Zoo Projektes ergab sich aus der notwendigen Integration des Pipettierroboters. Die Module des Pipettierroboters wie z.B. die DNA-Normalisierung und das PCR-Setup wurden als unabhängige Module implementiert. Das führt dazu, dass im Gegensatz zu der Implementierung einer Prozesskette, die aus mehreren, zusammenhängenden Modulen besteht, unabhängige Module in neuen Methoden wiederverwendet und zu neuen Prozessketten zusammengesetzt werden können. Die Spezifikation der Reihenfolge und der Informationsfluss zwischen den Modulen muss dafür vom LIMS übernommen werden.

Da nach Analyse kommerzieller LIMS- und Open-Source-Softwareprodukte kein Produkt identifiziert werden konnte, das den speziellen Anforderungen insbesondere an das Datenmanagement des Pipettierroboters sowie an die Unterstützung der optimierten Laborprozesse bereits durchgeführter Resequenzierungen und dem kalkulierten Kostenrahmen des PBA-Zoo Projektes entsprach, wurde hier eine Eigenentwicklung beschlossen. Für diese Eigenentwicklung von Software war eine detaillierte Anforderungsanalyse erforderlich, in der der zu unterstützende Anwendungsbereich definiert, Anwendergruppen identifiziert, eine geeignete Softwarearchitektur spezifiziert wurde. Darüber hinaus wurde festgelegt, ob das LIMS als Desktop- oder Webanwendung zu implementieren ist.

Ziel der vorliegenden Dissertation mit dem Titel „ReSeqLIMS: Laborinformations- und Managementsystem für die Hochdurchsatz-Resequenzierung“ war die Bearbeitung folgender Fragestellungen:

1. Wie muss die Datenbasis des ReSeqLIMS für das Amplicon Re-Sequencing Center aussehen?
2. Wie sind die Module des ReSeqLIMS zu implementieren, um die optimale Integration von Personen, Laborprozessen und Geräten des Labores zu gewährleisten?
 - Welcher Software-Typ (Desktop- oder Webanwendung) ist für das ReSeqLIMS unter Berücksichtigung der Zielgruppen besonders geeignet?
 - Wie können den Kunden der Zugang und die intuitive Nutzung des ReSeqLIMS ermöglicht und die Mitarbeiter in der Erledigung ihrer Forschungs- und Routinarbeiten optimal unterstützt werden?
 - Was ist bei der Implementierung des ReSeqLIMS zu berücksichtigen, um die existierenden Qualitätsstandards und Normen des Labores sicherzustellen und die Erweiterbarkeit bzgl. neuer Geräte und Laborworkflows bzw. deren Veränderungen und Anpassungen zu ermöglichen?

2 Material und Methoden

2.1 Webanwendung und Java

Das Internet ist ein weltweites Rechnernetzwerk, in dem Rechner und lokale Rechnernetzwerke verbunden sind. Während unter dem Begriff Internet allein die Infrastruktur zu verstehen ist, ermöglicht erst die Bereitstellung und Nutzung von Internetdiensten wie z.B. des World Wide Webs (WWW) sowie Electronic-Mail (E-Mail) und des File Transfer Protocols (FTP) einen Austausch von Daten [12]. Die Kommunikation in Rechnernetzwerken erfolgt über standardisierte Internetprotokolle, im WWW z.B. Protokolle der Transmission Control Protocol/Internet Protocol (TCP/IP)-Familie. Entsprechend dem Client-Server-Modell werden in einem Rechnernetzwerk Server (Dienstanbieter) und Clients (Dienstnutzer) unterschieden (siehe Abb. 1) [48].

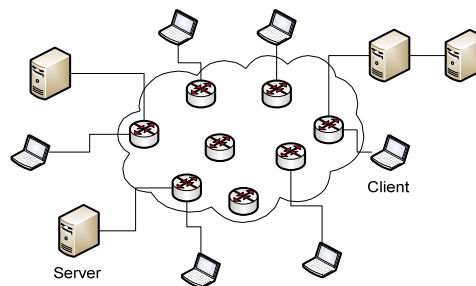


Abb. 1: Im Client-Server-Modell werden weltweit verteilte Clients und Server über das Internet miteinander verbunden.

Ein Dienst kann u.a. ein Prozess, ein Programm oder auch ein Betriebsmittel sein. Die Anforderung eines Dienstes durch den Client wird als Request bezeichnet, die Bereitstellung des Dienstes durch den Server an den Client als Response. Zur erfolgreichen Kommunikation zwischen Server und Client ist die eindeutige Identifikation eines Rechners in einem Rechnernetzwerk erforderlich. Hierfür wird im Internet der Uniform

Resource Locator (URL), die global eindeutige Adresse von Rechnern, Ressourcen oder Dokumenten, verwendet (siehe Abb. 2) [12].

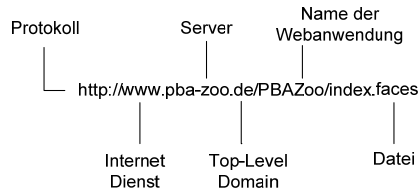


Abb. 2: Die URL wird zur eindeutigen Identifikation von Rechnern, Ressourcen oder Dokumenten im Internet verwendet.

Eine Webanwendung ist ein Softwaresystem, das auf einem Server ausgeführt wird und entsprechend der World Wide Web Consortium (W3C)-Spezifikationen [30] implementiert ist [35]. Die von der Webanwendung bereitgestellten Inhalte und Dienste können vom Client unter Verwendung des Internets über ein User-Interface genutzt werden. Als User-Interface verwendet der Client i.d.R. einen Webbrowser, ein spezielles Computerprogramm zur Anzeige von Dokumenten und Daten im WWW unter Verwendung des Hypertext Transfer Protocols (HTTP). Bei der Entwicklung von Webanwendungen ist zu berücksichtigen, dass das HTTP-Protokoll zustandslos ist und aufeinander folgende Requests eines Clients als unabhängige Transaktionen bearbeitet werden. Daher ist ein Zugriff auf Informationen vorangegangener Requests ohne besondere Vorkehrungen nicht möglich, da ein Request übergreifender Informationsaustausch die Identifikation des Clients am Server erfordert. Innerhalb einer Session, dem Zeitraum der bestehenden Verbindung eines Clients und eines Servers, könnten durch die Identifikation des Clients dessen Requests erkannt und einer Session zugeordnet werden. Zur Identifikation des Clients wird eine Session-ID verwendet, die entweder direkt in der Request-URL oder im HTTP-Header an den Server übertragen werden kann. Die Übertragung der Session-ID im HTTP-Header erfordert die clientseitige Verwendung von Cookies, einer Textdatei, die nach dem Senden des Responses vom Server auf der Clientseite gespeichert wird, wenn im HTTP-Header eine Cookie-Zeile übertragen wird. In einem Cookie können neben Informationen des Webservers auch benutzerspezifische Einstellungen gespeichert werden. Stellt der Client nun einen Request an einen Server, werden seiten- und serverspezifische Cookies im HTTP-Header mit übertragen und ermöglichen die serverseitige Client-Identifikation. Die Verwendung von Cookies muss

clientseitig erlaubt sein entweder generell oder für spezifische Webseiten. Responses des Servers werden in Hypertext Markup Language (HTML)-Dateien im Webbrowser dargestellt. Die textbasierte Auszeichnungssprache HTML eignet sich insbesondere für die strukturierte Darstellung von Texten, Multimediaelementen und Hyperlinks.

Zur Implementierung einer Webanwendung können eine Vielzahl von Programmiersprachen verwendet werden. Die Programmiersprache Java [32] wurde seit dem Jahr 1991 von der Firma Sun Microsystems entwickelt und steht seit 1995 zum freien Download zur Verfügung. Die Plattformunabhängigkeit ist ein großer Vorteil von Java. Einzig die Installation der Java-Runtime-Environment (JRE) ist zur Ausführung von Java-Programmen erforderlich, die aus der Java Virtual Machine (JVM), in der die Ausführung der Java Programme erfolgt und überwacht wird, sowie den Java-Klassenbibliotheken besteht. Die Plattformunabhängigkeit von Java wird dadurch erreicht, dass der Java-Quellcode zunächst mit einem Compiler in plattformunabhängigen Bytecode übersetzt wird. Anschließend wird der Bytecode in der JVM mit einem plattformspezifischen Interpreter in ausführbaren Programmcode übersetzt. Das objektorientierte Entwurfparadigma der Programmiersprache Java kapselt Daten bzw. Eigenschaften und Methoden in einem Objekt. Die Eigenschaften eines Objektes sind von anderen Objekten nicht direkt, sondern ausschließlich über die Methoden des Objektes manipulierbar. Komplexe Systeme bestehen aus einer Vielzahl von Objekten, die durch das Versenden und Empfangen von Nachrichten miteinander kommunizieren. Das abstrakte Konstrukt der Klasse abstrahiert von einem Objekt, auch als Instanz einer Klasse bezeichnet, und stellt den Bauplan eines Objektes dar. Im Vergleich z.B. mit der Programmiersprache C wurde in Java auf die fehleranfällige Zeigerarithmetik verzichtet und die Speicherverwaltung vom Garbage-Collector übernommen. Dadurch konnten sowohl die Programmierung von Software vereinfacht als auch Schwachstellen beseitigt werden, über die Schadcode in Programme eingefügt werden könnte.

2.2 Softwareentwicklung

2.2.1 Grundlagen der Softwareentwicklung

Die Entwicklung und Wartung komplexer Software erfordert einen hohen zeitlichen und finanziellen Aufwand. Die Komplexität von Prozessen und Systemen kann durch die Zerlegung in Subprozesse bzw. -systeme mit definierten Abhängigkeiten erheblich reduziert werden [7]. In der Softwareentwicklung hat sich zur Komplexitätsre-

duzierung ein strukturierter Prozess mit den Kernprozessen bzw. Phasen Planung, Analyse, Entwurf, Implementierung sowie Validierung und Tests etabliert [7]. Innerhalb des Softwareentwicklungsprozesses sind die Ergebnisse vorangegangener Phasen Grundlage und Rahmen für die folgenden Phasen. Rückkopplungen über die Ergebnisse in einer Phase führen zu Veränderungen in den vorangegangenen Phasen, welche wiederum zu Anpassungen in den nachfolgenden Phasen führen können. Je später im Entwicklungsprozess Änderungen in frühen Phasen durchgeführt werden müssen, umso zeitaufwendiger und teurer sind die daraus resultierenden Änderungen in nachfolgenden Phasen. Daher liegt der Fokus vor der eigentlichen Implementierung der Software auf einer ausführlichen und detaillierten Planung und Analyse, um spätere Änderungen in den Dokumenten und Modellen dieser Phasen möglichst vollständig zu vermeiden.

Das zu implementierende Softwaresystem ist, ebenso wie der Softwareentwicklungsprozess, gekennzeichnet durch einen hohen Grad an Komplexität. Um die Komplexität von Softwaresystemen zu reduzieren, werden Softwarearchitekturen verwendet, die die strukturierte und/oder hierarchische Anordnung von Softwaresystemkomponenten und deren Beziehungen innerhalb eines Softwaresystems beschreiben [7]. Eine der populärsten Softwarearchitekturen ist die Drei-Schichten-Architektur (siehe Abb. 3) [48].

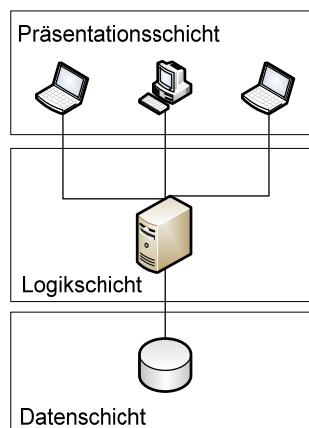


Abb. 3: Die Drei-Schichten-Architektur strukturiert ein Softwaresystem in eine Präsentations-, Logik- und Datenschicht.

Die Strukturierung eines Softwaresystems erfolgt in der Drei-Schichten-Architektur in eine Präsentations-, Logik- und Datenhaltungsschicht. Die Kommunika-

tion erfolgt nur zwischen unmittelbar benachbarten Schichten und Softwaresystemkomponenten können nur exakt einer Schicht zugeordnet werden. Änderungen einzelner Softwaresystemkomponenten können aufgrund der Verwendung von Schnittstellen zwischen den Schichten ohne Änderungen in den benachbarten Schichten vorgenommen werden.

2.2.2 Die Datenhaltungsschicht

Arbeitsabläufe in einem Unternehmen bestehen aus einer definierten Abfolge von Aktivitäten, um eine spezifische Arbeitsaufgabe zu erreichen. Für einen Arbeitsablauf ist das räumliche und zeitliche Zusammenwirken von Mensch und Betriebsmitteln charakteristisch, durch das eine Eingabe in eine Ausgabe überführt wird [49]. Ein- und Ausgaben sind z.B. Arbeitsgegenstände, Informationen und Energien. Eine große Herausforderung in vielen Unternehmen ist die effiziente, aktuelle und konsistente Bereitstellung von Informationen in Arbeitsabläufen, die in beliebigen Bereichen von unterschiedlichen Personen in einer Vielzahl von Formaten generiert werden. Effiziente Arbeitsabläufe basieren auf einer effizienten Datenorganisation, die ohne vorangegangene Datenmodellierung nicht möglich ist [8]. Das Ziel der Erstellung eines Datenmodells ist die redundanzfreie, vollständige und funktionsunabhängige Beschreibung der im relevanten Unternehmensbereich auftretenden Datenstrukturen [8]. In einem dreistufigen Prozess wird daher zunächst das semantische Datenmodell bzw. das Fachkonzept definiert, das in das Datenverarbeitungs (DV)-Konzept überführt wird, welches wiederum die Basis für die anschließende Implementierung ist [8,56]. Das bekannteste und ein leicht verständliches Modell zur Modellierung von Daten auf fachkonzeptioneller Ebene ist das Entity Relationship Model (ERM) von Chen [16,61], das zwischen Entities, den realen Gegenständen bzw. Objekten des betrachteten Anwendungsbereichs, und den Relationships, den Beziehungen zwischen den Entities, unterscheidet (siehe Abb. 4). Gleichartige Entities werden zu Entitytypen, gleichartige Relationships zu Relationshiptypen zusammengefasst. Die Darstellung der Entitytypen erfolgt im ERM durch Rechtecke, während Rauten die Relationshiptypen repräsentieren. Die Studenten einer Universität sind demnach als Entities aufzufassen, die z.B. zu den Professoren in Beziehung stehen, da Studenten die Vorlesungen von Professoren besuchen (siehe Abb. 4, Beispiel). Die Entities Studenten können im Entitytypen „Student“ zusammengefasst werden. Die Relationships zwischen Studenten und Professoren werden zum Relationship-

typen Professoren-Student-Zuordnung zusammengefasst. Entities besitzen Eigenschaften bzw. Attribute, z.B. Matrikelnummer und Namen. Jedes Attribut eines Entities ordnet dem Entitytypen Werte einer Domäne zu, wobei die Domäne alle Werte zusammenfasst, die eine Eigenschaft bzw. das Attribut eines Objektes annehmen kann.

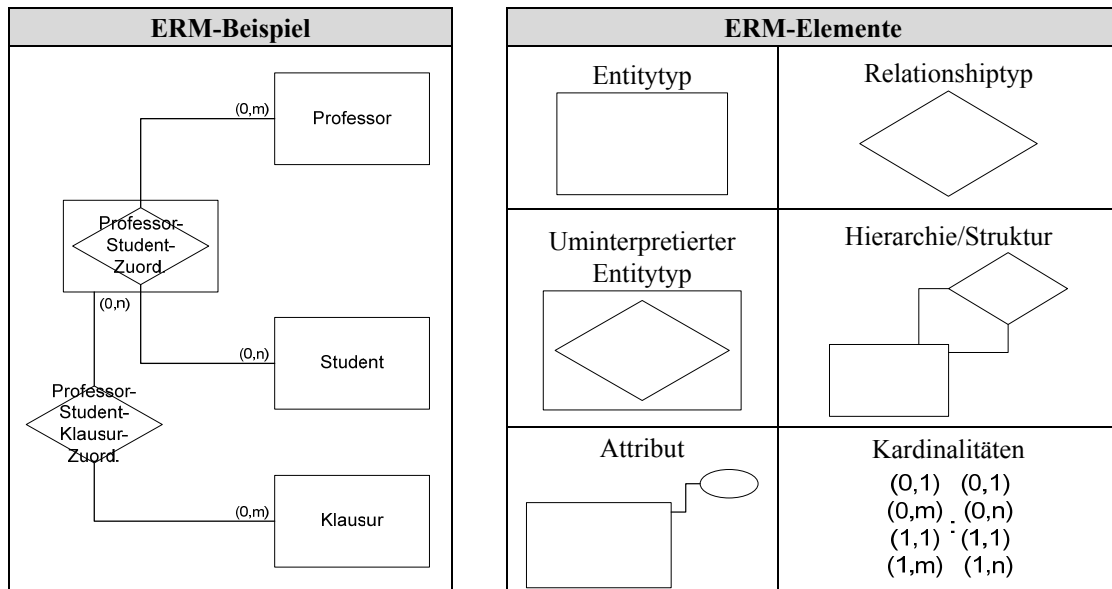


Abb. 4: Beispiel und ERM-Notation [8,56].

Wird ein Objekt eindeutig durch eine Eigenschaft identifiziert, ist das Attribut durch Unterstreichen als Schlüsselattribut zu kennzeichnen. Eignen sich mehrere Attribute als Schlüsselattribut, ist zu analysieren, welches Attribut als Schlüsselattribut verwendet werden kann. Attribute werden im ERM durch Ovale repräsentiert. Für die Relationshiptypen können Komplexitätsgrade (Kardinalitäten) angegeben werden. Kardinalitäten definieren, wie oft ein Entitytyp in einem Relationshiptyp auftreten kann bzw. muss. Die Kardinalität wird häufig in der Max-Notation (z.B. 1:m) angegeben, obwohl die Min-Max-Notation (z.B. (0,1):(0,m)) aufgrund der exakteren Abbildung der Beziehungen zwischen Entitytypen vorzuziehen ist [36]. Wird vereinfachend angenommen, dass ein Professor nur eine Vorlesung anbietet, bedeutet die (0,n):(0,m) Kardinalität zwischen den Entitytypen Student und Professor, dass ein Student keine oder bis zu n Vorlesungen besucht und die Vorlesung eines Professors von keinem oder von bis zu m Studenten besucht wird. Entities können auch Relationships zu Entities des gleichen Entitytypen eingehen. Je nach Kardinalität dieser Relationships können Strukturen oder Hierarchien unterschieden werden. Während in einer Struktur ein Enti-

tytyp mehrere Vorgänger besitzen kann, hat in einer Hierarchie jedes Entity exakt einen Vorgänger. Ein Sonderfall tritt auf, wenn Entities mit den Relationships anderer Entities in Beziehung gesetzt werden. Da Relationships keine Beziehungen eingehen können, ist der Relationshiptyp als uminterpretierter Entitytyp zu modellieren. Im Beispiel der Abb. 4 sind Prüfungen nur für die Vorlesungen sinnvoll, für die sich ein Student eingeschrieben bzw. angemeldet hat. Die Menge der von einem Studenten besuchten Vorlesungen befindet sich im Relationshiptyp zwischen Student und Professor, der daher als uminterpretierter Entitytyp modelliert werden muss, um eine Verbindung mit dem Entitytyp Klausur eingehen zu können.

Ist das Datenmodell auf Fachkonzeptebene erstellt, erfolgt die Überführung in das DV-Konzept, das auch als Datenbankmodell bezeichnet wird [8]. Das bekannteste Datenbankmodell ist das Relationenmodell, das Grundlage der weit verbreiteten relationalen Datenbanken ist [36]. Eine Relation ist die mathematische Beschreibung einer zweidimensionalen Tabelle. Jede Zeile der Tabelle beschreibt einen Datensatz, die Entities, mit einer festen Anzahl von Attributausprägungen bzw. Spalten, den Eigenschaften der Entities. Für die Überführung des ERMs in das relationale Datenbankmodell gilt, dass Entity- und Relationshiptypen von (m:n)-Beziehungen jeweils zu Relationen werden, während bei (1:m)-Beziehungen der Entitytyp, dessen Entity mit der Kardinalität m im entsprechenden Relationshiptypen auftreten, das Schlüsselattribut des anderen Entitytypen als Nichtschlüsselattribut in die Relation übernimmt (siehe Abb. 5) [36]. Der beteiligte Relationshiptyp wird nicht in eine Relation umgesetzt.

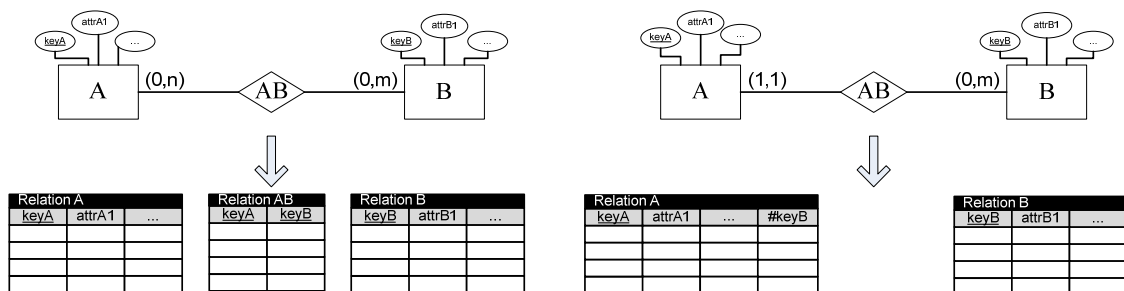


Abb. 5: Transformation eines ERMs in ein relationales Datenbankmodell am Beispiel einer (0,n):(0,m)- und einer (1,1):(0,m)-Beziehung.

Die Implementierung der Datenbank erfolgt auf Basis des Datenbankmodells in der Sprache eines spezifischen Datenbanksystems. Die Datenbanksprachen relationaler

Datenbanken basieren auf dem Structured Query Language (SQL)-Standard [70]. Der SQL-Standard beinhaltet Data Definition Language (DDL)-Befehle zur Beschreibung und Erstellung der Datenbank, Data Manipulation Language (DML)-Befehle zur Manipulation von Daten sowie weiterer Befehle, um SQL-Befehle in eine Programmiersprache zu integrieren [8,36].

2.2.3 Die Logikschicht

In der Logikschicht erfolgen die Berechnungen der Anwendung. Durch die Trennung der Logik von der Präsentations- und Datenschicht können Änderungen, z.B. an der Benutzerschnittstelle und auch an der Datenbank, ohne Änderungen in der Programmlogik vorgenommen werden. Die Berechnungen der Logikschicht basieren auf Eingaben sowie der Interaktion mit den Anwendern und der zugrundeliegenden Datenbasis. Die Implementierung der Logikschicht kann generell in allen Programmiersprachen wie Perl, C und Java erfolgen. Bei der Kommunikation zwischen dem objektorientierten Java und einer auf Tabellen basierenden relationalen Datenbank besteht allerdings ein Paradigmenbruch, da Objekte mit Attributen im relationalen Modell als Zeile bzw. Tupel abgebildet werden, in der die Spalten den Attributen entsprechen. Der Paradigmenbruch zwischen Tupel und Objekt kann durch die Transformation von Objekten auf Tabellen mit datenbankspezifischen Programmiersprachenkonstrukten überwunden werden. Um nicht nur die Logik von der Datenhaltung zu trennen, sondern auch Datenbankunabhängigkeit zu garantieren, steht mit Hibernate [28] ein Open-Source Persistenz and Mapping Framework zur Verfügung. Dieses stellt Methoden bereit, die die Transformation von Objekten auf Tabellen für die gängigsten relationalen Datenbanken übernehmen. Dazu werden in Extensible Markup Language (XML)-Dateien die Attribute der Objekte den entsprechenden Spalten der relationalen Tabellen zugeordnet. Die Definition und Spezifikation der verwendeten Datenbank erfolgt ausschließlich zentral in der Hibernate-Konfigurationsdatei. Ein Austausch der Datenbank erfolgt bei Verwendung von Hibernate allein durch Anpassungen in der zentralen Konfigurationsdatei, ohne Änderungen am Java-Quellcode vornehmen zu müssen.

2.2.4 Die Präsentationsschicht

In der Präsentationsschicht erfolgt die Implementierung des Graphical User Interfaces (GUI), der Benutzerschnittstelle bzw. Weboberfläche zur Interaktion der

Webanwendung und des Clients. Das Front-End auf Clientseite ist der Webbrowser zur Anzeige der serverseitig erzeugten HTML-Dateien. Da HTML ein statisches Anzeigeformat ist, können nach dem Laden der HTML-Seiten im Browser zunächst keine Änderung mehr an Layout und Daten erfolgen. Das dynamische Verändern bereits geladener HTML-Seiten ist erst durch Weiterentwicklungen wie JavaScript und Asynchronous JavaScript and XML (AJAX) möglich. JavaScript ist eine clientseitige Skriptsprache, die sich insbesondere für die Überprüfung von Benutzereingaben, Anpassungen des Layouts und weiterer Anzeigeeffekte wie Popups eignet. Der Quellcode von JavaScript-Programmen wird in den HTML-Quelltext integriert und durch den Webbrowser ausgeführt. Bei der Implementierung von JavaScript-Programmen ist darauf zu achten, dass kein einheitlicher JavaScript-Standard für alle Browser existiert. AJAX wird ebenfalls von den gängigsten Webbrowsern unterstützt und ermöglicht eine Kommunikation von Server und Client, ohne dass die Webseite neu geladen werden muss bzw. es werden nur die Bestandteile der Seite nachgeladen, deren Daten sich geändert haben. Dieses führt zu reduzierten Lade- und Wartezeiten, da durch die asynchrone Kommunikation während des Sendens und Empfangens der Anwender weiter mit der Benutzeroberfläche interagieren kann. Durch die Verwendung von AJAX kann auch bei Webanwendungen ein desktopähnliches Verhalten erzielt werden.

Serverseitig werden Servlets zur Erzeugung von dynamischen Webseiten verwendet, die im Servlet-Container ausgeführt werden. Servlets sind Java-Klassen, die Anfragen von Clients empfangen, Berechnungen ausführen und die Ergebnisse im HTML-Format an die Clients zurücksenden. Das Servlet übernimmt die Erzeugung und Übertragung der Session-ID und speichert benutzerbezogene Daten der aktuell gültigen Session im Hauptspeicher des Servers. Wird die Session nach zu langer Inaktivität oder nach Löschung ungültig, werden die Daten wieder entfernt. Bei der Verwendung von Servlets ist zu beachten, dass zwischen Darstellung und Programmlogik nicht getrennt wird. Änderungen an der Ergebnispräsentation erfordern daher Änderungen am Servlet, wozu Programmierkenntnisse, z.B. beim Webdesigner, erforderlich sind. Die Web-Programmiersprache JavaServer Pages (JSP) erlaubt im Gegensatz zu Servlets die Trennung von Layout und Programmlogik zur Generierung von dynamischen Webseiten. JSP-Seiten sind HTML- oder XML-Dateien, in die direkt Java-Code eingebettet werden kann. Um insbesondere die Lesbarkeit zu erhöhen, kann durch das Verwenden einer Tag-Bibliothek der Java-Code einer JSP-Seite ausgelagert werden. Die Tag-

Bibliothek der JSP-Spezifikation definiert Tags, die in der HTML/XML-Seite die Ergebnisposition und/oder das Layout der Ergebnispräsentation bestimmen. Die Verarbeitung der JSP-Tags mit evtl. angegebenen Parametern ist in den Tag-Klassen implementiert. JSPs trennen Programmlogik und Präsentation in unterschiedliche Dateien. Die Übersetzung der JSP-Seiten in Java-Quellcode erfolgt durch den JSP-Compiler, bei Bedarf unter Verwendung der Tag-Klassen. Der erzeugte Java-Quellcode entspricht dem eines Servlets und kann vom Webserver weiterverarbeitet werden.

Werden die HTML-Seiten von Webanwendungen betrachtet, können eine Vielzahl von Standardfunktionalitäten identifiziert werden. Die Interaktion mit den Anwendern erfolgt z.B. über Webformulare und die Eingabe von Benutzerdaten erfolgt in Eingabefelder, die auf ihre Richtigkeit bzgl. Semantik und Syntax geprüft werden. Die Ergebnisse aus Datenbankabfragen werden häufig in Tabellen dargestellt, die Inhalte von Komponenten wie z.B. Dropdown-Listen basieren auf den Daten einer Datenbank. Die Navigation in einer Webanwendung wird durch Hyperlinks und Benutzerinteraktionen festgelegt. Das JavaServer Faces (JSF)-Framework ist ein Programmierstandard, der den Entwurfsrahmen und die Anwendungsarchitektur von Softwarekomponenten für die Benutzeroberfläche einer Webanwendung definiert. Die in JSF bereits implementierten Komponenten wie z.B. Dropdown-Listen, Tabellen, Formularfelder inklusive Überprüfungsmechanismen und der Verbindung zu Datenquellen können zur Zusammenstellung einer Webseite durch einfaches Einbinden in HTML-Seiten verwendet werden. Hierfür stehen Tags aus Tag-Bibliotheken zur Verfügung. Die Strukturierung der JSF-Komponenten einer Seite erfolgt in einer Baumstruktur, die auch die Bearbeitungsreihenfolge beeinflusst. Für die Anzeige von JSF-Seiten sind komponentenspezifische Renderer zur Trennung von Präsentation und Logik definiert. Die Ergebnisseiten werden in HTML oder in Extensible HTML (XHTML) an den Client zurückgesendet. Die Navigationsregeln werden zentral in einer Konfigurationsdatei erfasst. Das JSF-Framework ermöglicht die eigene Implementierung und Registrierung von Komponenten im JSF-Framework. Die Verwendung von in JSF vorgefertigten Komponenten entspricht dem Prinzip der Wiederverwendbarkeit Coderedundanzen zu vermeiden und durch das Verwenden bereits implementierter Komponenten Zeit und Kosten zu sparen. Das JSF-Framework sieht die Verwendung von AJAX nicht vor. Um trotzdem die Vorteile von AJAX nutzen zu können, sind Komponentenbibliotheken verfügbar, die JSF-Komponenten mit AJAX Funktionalitäten bereitstellen.

Die Verwendung von JSP und JSF ist aufgrund der Unterschiede u.a. im Life-Cycle z.T. mit Problemen behaftet. JSF nutzt JSP zur Erstellung der Ergebnisdateien. Im Unterschied zu JSF verwendet JSP aber zur Strukturierung der Komponenten in einer Seite keine Baumdarstellung. Stattdessen werden die Elemente vielmehr entsprechend ihrer Position in der Seite von oben nach unten abgearbeitet, um auf den Request einer Komponente einen unmittelbaren Response geben zu können. JSF dagegen verwendet einen Komponentenbaum zur Strukturierung der Komponenten und trennt in einem komplizierteren Life-Cycle klar zwischen der Erzeugung und dem Rendern der Komponente in unterschiedlichen Life-Cycle-Phasen. Daher kann Facelets [46] JSP als View-Handler ersetzen, da Facelets auf der Erstellung eines eigenen Komponentenbaumes aus den JSF Komponenten basiert. Ein weiterer Vorteil bei der Verwendung von Facelets ist die Unabhängigkeit von einem spezifischen Web-Container, da JSF und JSP auch ohne Java Platform Enterprise Edition (JEE)5 und JSP 2.1 verwendet werden können. Im Vergleich zu JSP erzeugt und kompiliert Facelets bei einem ersten Aufruf einer Webseite im Hintergrund keinen Java-Bytecode, wodurch sich für die Seitenerstellung gegenüber JSP ein Zeitvorteil ergibt. Des Weiteren unterstützen JSP und JSF das in Facelets implementierte Konzept des Templating nicht, welches die einfache Strukturierung einer Webseite z.B. mit einer einheitlichen Kopf-, Fuß und Navigationsleiste ermöglicht. Zusätzlich bietet Facelets eine verbesserte Debug-Möglichkeit mit einer Spezifizierung der Fehlerlokalisierung. Durch die ausschließliche Verwendung von XHTML als Eingabeformat wird bei der Verwendung von Facelets allerdings die Markup-Unabhängigkeit von JSF aufgegeben. Damit können für die JSF-Komponenten nur Facelets-spezifische Renderer verwendet werden. Das Verwenden von JSF-Komponenten mit Renderern alternativer Anbieter ist mit Facelets nicht mehr möglich.

Die Eingabe von Daten über Webformulare ist ab einer bestimmten Größenordnung nicht mehr komfortabel und praktikabel, insbesondere wenn die Erfassung der Daten z.T. bereits in Excel- oder Word-Dateien erfolgte. Das Auslesen von Informationen aus Dokumenten und die Transformation in ein zur weiteren Verarbeitung geeignetes Format für z.B. eine Webanwendung erfolgt durch ein Verarbeitungsprozess, der als Parsen bezeichnet wird. Computerprogramme, die diesen Prozess durchführen, werden als Parser bezeichnet. In der Open-Source-Software Apache Poor Obfuscation Implementation (POI) steht eine Softwarebibliothek zur Verfügung, die Methoden zum Parsen (Lesen) und Schreiben der gängigsten Microsoft (MS)-Dokumente wie z.B. MS-

Word, MS-Excel, MS-PowerPoint und MS-Outlook Dateien beinhaltet. Apache POI ist in Java implementiert und kann als Application Programming Interface (API) in eine Webanwendung integriert werden.

2.3 Entwicklungsumgebung und Versionshinweise verwendeter Software und Bibliotheken für die ReSeqLIMS-Implementierung

Die Entwicklung von Software kann durch die Verwendung einer Integrated Development Environment (IDE) unterstützt werden. Eine IDE integriert einen Texteditor zur Bearbeitung der Quellcodedateien sowie neben Compiler und Debugger je nach Programmiersprache evtl. auch Interpreter und Linker. Zusätzlich stehen eine Vielzahl von Formatierungs- und Strukturierungsfunktionen zur Verfügung. Durch die Verwendung einer IDE werden sowohl die Lesbarkeit des Quellcodes als auch die Ausführung und das Testen der Software erheblich vereinfacht. Für die Entwicklung von Software mit der Programmiersprache Java eignet sich insbesondere Eclipse [20]. Eclipse ist ein Open-Source-Software Projekt, das für die gängigsten Betriebssysteme erhältlich ist und über eine Vielzahl von Programmiersprachen spezifischen Plug-Ins verfügt. Die Entwicklung von dynamischen Webanwendungen wird durch Tools aus dem Web Tools Platform (WTP)-Projekt [71] unterstützt und vereinfacht. Für die Versionskontrolle der Webanwendung wird Subversion [21] verwendet.

Das ReSeqLIMS ist eine dynamische Java-Webanwendung. Für die Implementierung wurde Eclipse in der Version 3.4.2 und Java in der Version 1.6 verwendet. Als Webserver wurde der Apache Webserver 2.2.10 [64] in Verbindung mit Apache Tomcat 6.0 [6] als Servlet-Container aufgesetzt. Die Benutzer- bzw. Weboberfläche sollte als ein benutzerfreundliches, intuitives und komfortables GUI implementiert werden. Dazu wurde nicht die in JEE5 integrierte JSF-Implementierung, sondern Apache MyFaces 1.2.5 [46] verwendet. Der Vorteil von MyFaces gegenüber der JSF-Implementierung in JEE5 sind zusätzliche Komponenten wie z.B. die im ReSeqLIMS verwendete Kalenderkomponente. Um auch Komponenten mit AJAX-Unterstützung verwenden zu können, wurde zusätzlich zu MyFaces die umfangreiche Komponentenbibliothek RichFaces 3.3.2 [50] integriert. Zur Nutzung des Templatings wird Facelets 1.1.15 eingesetzt [33]. Die Grundlage des ReSeqLIMS ist eine umfangreiche Datenbank, in der Labordaten und Kundeninformationen persistent gespeichert werden. Hier wird die PostgreSQL-Version 8.1.15 [47] verwendet, obwohl zum Zeitpunkt der Implementierung bereits

PostgreSQL in der Version 8.4 verfügbar war. Um Datenbankunabhängigkeit zu erreichen, wurde die aktuellste Hibernate Version 3.5.5 verwendet, die nur bis zur PostgreSQL Version 8.1. getestet wurde. Der Versuch Hibernate 3.5.5 [28] mit der PostgreSQL Version 8.4 zu nutzen, wurde nach auftretenden Instabilitäten wieder verworfen. Neben der Datenbank ist eine konfigurierbare Ordnerstruktur auf dem Webserver zur Speicherung der generierten Excel-, PDF- und Chromatogramm-Dateien erforderlich (siehe Anh. A, Abb. 33). Zum Einlesen und Schreiben von Excel-Dateien, z.B. der Auftragsdatei und der Arbeits- und Ausgabelisten des Pipettierroboters, wird Apache POI 3.5 beta 4 [5] verwendet. Die Arbeitsliste für den DNA-Sequenzierer wird mit Hilfe von Standard Java-Methoden im TXT-Format generiert. Das Versenden der systemseitig generierten E-Mails erfolgt mit Hilfe von Sendmail [58] in der Version 8.13.8, das in den Apache Webserver integriert wurde, und der Javamail-Bibliothek 1.4.2. Für das Generieren der Portable Document Format (PDF)-Dokumente, wie z.B. der Rechnung und dem Barcodeaufkleber-Anschreiben, wurde die iText-Bibliothek 2.1.7 [31] verwendet. Die Erstellung der Flash-Tutorials der Hilfeseite erfolgte mit dem Programm ViewletBuilder 4.5.8 x511 [68].

Die Webanwendung wurde mit den Standard-Webbrowsern Mozilla Firefox (Version 3.6.3), Microsoft Internet Explorer (8.0.6001.18702), Apple Safari (Version 4.0.5) und Google Chrome (Version 5.0.375.99) getestet.

3 Ergebnisse

3.1 Das ReSeqLIMS-Datenmodell

Vor der Erstellung des Datenmodells erfolgt zunächst die Analyse des Anwendungsbereiches, in der durch Abstraktion die Objekte bzw. Daten des Anwendungsbereiches identifiziert und mit Hilfe des ERMs von Chen modelliert werden, die für das ReSeqLIMS von grundlegender Bedeutung sind. In den folgenden Kapiteln wird das ReSeqLIMS-ERM systematisch aufgebaut, aus Gründen der Übersichtlichkeit werden aber jeweils nur die neuen ERM-Bereiche grafisch dargestellt und beschrieben. Die Beschreibung der ERMs in den folgenden Kapiteln erfolgt synchron zu der modularen Einteilung des ReSeqLIMS in Kap. 3.2, um die modulare Struktur der ReSeqLIMS-Webanwendung bereits hier zu betonen. Die definierten ERMs und Kardinalitäten sind nicht als allgemeingültig zu verstehen, da das ReSeqLIMS-Datenmodell spezifisch an die Anforderungen der Amplikon-Resequenzierung im Rahmen des PBA-Zoo Projektes angepasst wurde. Auf eine Darstellung der Primärschlüssel und Attribute wird in den ERMs des Kap. 3.1 zugunsten einer besseren Lesbarkeit der ERMs nicht mit aufgeführt.

Die Überführung der fachkonzeptionellen Darstellung der ReSeqLIMS-Datenbank in das Datenbankmodell sowie der Implementierung des letzteren wird in der vorliegenden Dissertation nicht behandelt. Stattdessen wird in Kap. 3.2 die Benutzerschnittstelle und Funktionalität des ReSeqLIMS näher beschrieben.

3.1.1 Datenmodelle des ReSeqLIMS-Labormoduls

3.1.1.1 Methoden und Labororganisation

Ein Labor umfasst Arbeitsplätze naturwissenschaftlicher Disziplinen, an denen unter Einhaltung definierter Normen, Regeln und Vorschriften wissenschaftliche Forschungen und Experimente durchgeführt werden. Um ein definiertes Ziel zu erreichen, werden in einem Labor u.a. Methoden wie z.B. die MLST-Methode angewendet. Methoden, die auf der Bestimmung der DNA-Sequenz basieren sind gemeinsam, dass mit

der DNA als Ausgangsmaterial anhand der PCR eine Amplifizierung des Ausgangsmaterials bzw. spezifischer Zielfragmente durchgeführt wird. Vor der DNA-Sequenzierung werden dem Reaktionsgemisch Sequenzierprimer hinzugefügt, die für die DNA-Sequenzierung der Zielfragmente verwendet werden. Ein Primer(paar), das zur Sequenzierung sämtlicher Zielfragmente einer Methode verwendet werden kann, wird als Universal-Primer(paar) bezeichnet, fragmentspezifische Primer demgegenüber als individuelle Primer. Einer Methode muss mindestens ein Mastermix zugeordnet werden, der das in der PCR zu amplifizierende Fragment definiert (siehe Abb. 6). Die Anzahl der Mastermixe einer Methode ist vorgegeben durch die Anzahl der in der PCR zu amplifizierenden DNA-Sequenzbereiche. Für die PCR einer MLST-Methode mit sieben zu analysierenden Haushaltsgenen sind sieben Mastermixe erforderlich. Den methodenspezifischen Mastermixen (Methoden-Mastermix) muss jeweils mindestens ein Primer für die Sequenzierung des DNA-Fragmentes zugeordnet werden (Methoden-Mastermix-Primer).

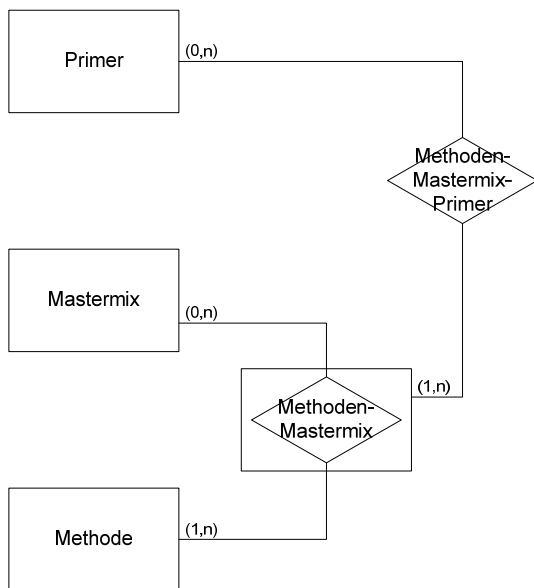


Abb. 6: Eine Methode im Bereich der DNA-Sequenzierung benötigt mindestens einen Mastermix für die PCR und mindestens einen Primer für die Sequenzierung.

Methoden können, vergleichbar mit einem Rezept, in einzelne Arbeitsschritte differenziert werden. In den einzelnen Arbeitsschritten werden z.B. die Mengen und die Qualität des eingesetzten Ausgangsmaterials sowie der verwendeten Verbrauchsmaterialien des jeweiligen Arbeitsschrittes festgelegt. Die Laborstation definiert den inhaltli-

chen Rahmen eines Arbeitsschrittes, wobei von konkreten Werten abstrahiert wird (siehe Abb. 7). Für die MLST-Methode können die Laborstationen DNA-Quantifizierung, DNA-Normalisierung, PCR-Setup, PCR-Produktaufreinigung, Sequenzierungssetup, Sequenzierungsproduktaufreinigung und die abschließende DNA-Sequenzierung unterschieden werden. Die Reihenfolge der Laborstationen wird durch die paarweise Anordnung von Vorgänger- und Nachfolger-Laborstation, die als Laborstationspaar bezeichnet werden, festgelegt. Durch die Modellierung der Reihenfolge als Struktur und nicht als Sequenz sind für eine Laborstation mehrere Vorgänger und mehrere Nachfolger möglich.

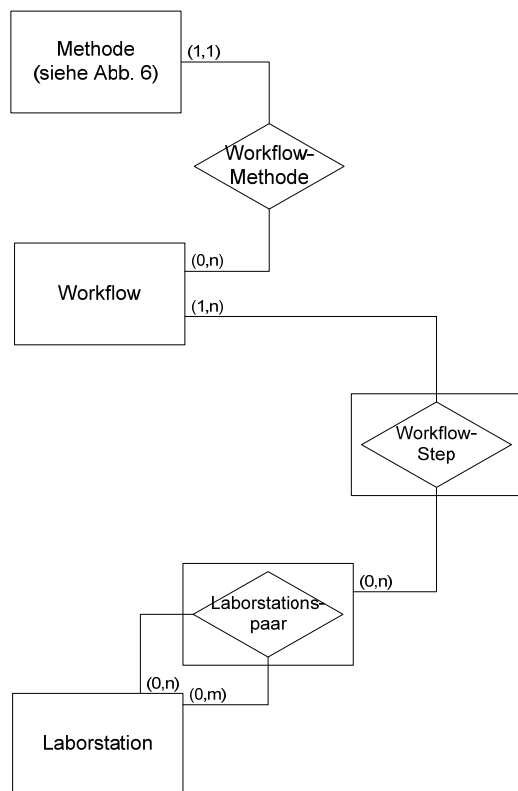


Abb. 7: Methoden bestehen aus einer definierten Abfolge von Arbeitsschritten, die jeweils an Laborstationen abgearbeitet werden.

Eine definierte Abfolge von Laborstationen zur Erreichung eines spezifischen Zieles, das durch eine Methode definiert wird, wird als Workflow bezeichnet. Einem Workflow muss mindestens ein Laborstationspaar zugeordnet werden (Workflow-Step). Durch das weitere Hinzufügen von Laborstationspaaren können beliebige Workflows modelliert werden (siehe Abb. 8). Methoden, deren Arbeitsschritte auf eine identische

Abfolge von Laborstationen abgebildet werden können, werden dem entsprechenden Workflow eindeutig zugeordnet (Workflow-Methode).

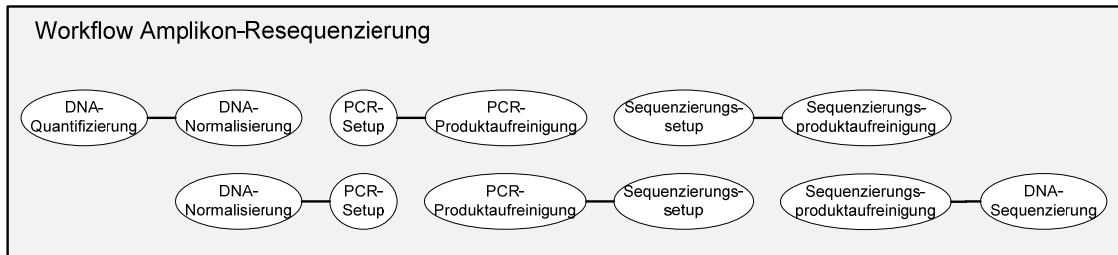


Abb. 8: Der Amplikon-Resequenzierung Workflow kann durch das paarweise Anordnen von Laborstationspaaren zusammengesetzt werden.

3.1.1.2 Materialien, Optionen und Trägermedien für Methoden

Für ein Labor gilt, dass die verwendeten Materialien in bestimmten Zuständen oder mit bestimmten Eigenschaften in Abhängigkeit u.a. von den Verarbeitungs- und Aufbewahrungsmöglichkeiten in einem Labor auftreten können. Zur differenzierten Betrachtung eines Materials mit verschiedenen Eigenschaften muss den Materialien mindestens eine Option zugeordnet werden (Material-Option, siehe Abb. 9). Jedes Material benötigt mindestens ein Trägermedium (Plastikware) auf dem es im Labor transportiert, verarbeitet und gelagert werden kann. Die Material-Option kann Auswirkungen auf die wählbare Plastikware haben, so dass die Bestimmung des Trägermediums durch Zuordnung der Material-Option zu mindestens einer Plastikware erfolgen muss (Material-Option-Plastikware). Eine Methode ist beschränkt auf die Verarbeitung bestimmter Materialien, die evtl. zudem mit bestimmten Eigenschaften und auf gewünschten Plastikwaren vorliegen bzw. eingesendet werden sollen (Methoden-Option-Plastikware).

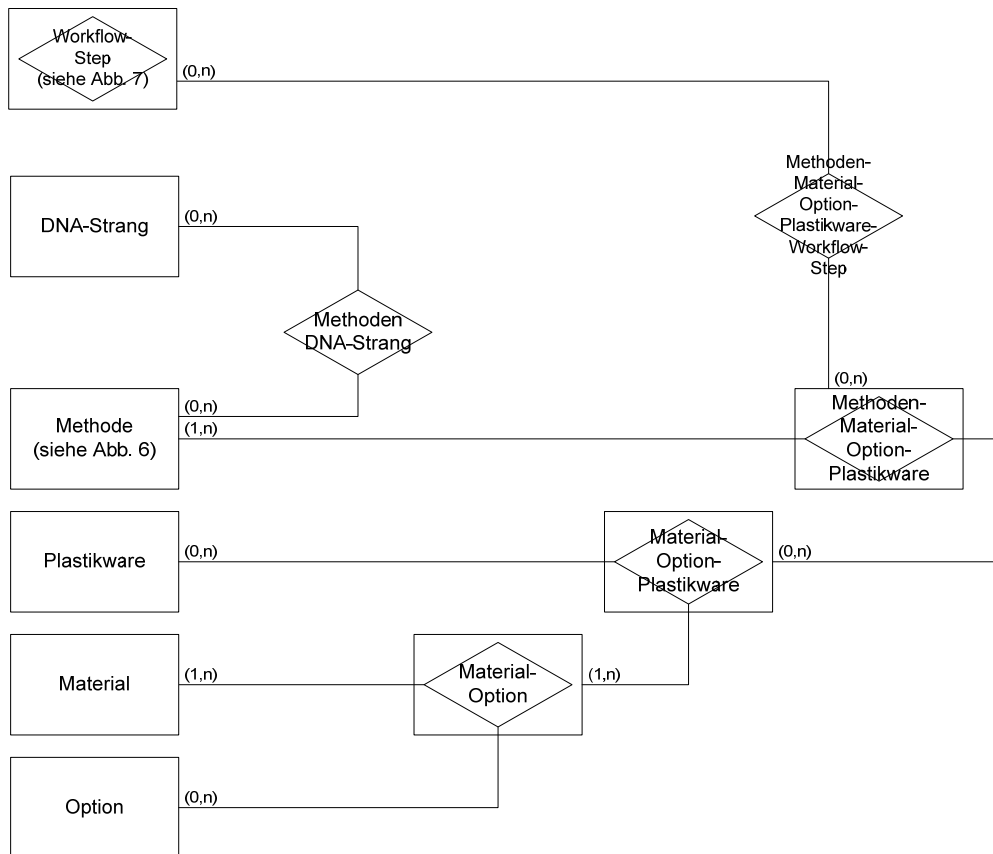


Abb. 9: Methoden lassen sich auf Material-Option-Plastikware Kombinationen und optional auf eine Auswahl zu sequenzierender DNA-Stränge einschränken.

Bei einer DNA-Sequenzierung können Forward-, Reverse- oder beide Stränge sequenziert werden. In der Zuordnung eines DNA-Stranges zu einer Methode wird festgelegt, welche DNA-Stränge eine Methode zur Sequenzierung anbietet. Werden der Methode keine DNA-Stränge zugeordnet, handelt es sich um Methoden wie z.B. die Fragmentlängen-Analyse, die nicht mit dem Arbeitsschritt der DNA-Sequenzierung beendet werden. Für die Modellierung der Methoden-DNA-Strang Zuordnung gilt, dass nicht nur zwischen Forward- und Reverse-Strang unterschieden werden kann, sondern das im ERM aus Abb. 9 auch die Möglichkeit von DNA-Strängen einzelner Gene zur Auswahl bei einer Methode definiert ist. Den Workflow-spezifischen Laborstationspaaren können Methoden-Material-Option-Plastikware Kombinationen zugeordnet werden. Dadurch werden alle möglichen Workflow-spezifischen Laborstationspaare festgelegt, an denen mit der Bearbeitung einer Methoden-Material-Option-Plastikware Kombinationen begonnen werden kann (Methoden-Material-Option-Plastikware-Workflow-Step).

3.1.1.3 Verbrauchsmaterialien

An den Laborstationen werden zur Bearbeitung der Laborproben (siehe Kap. 3.1.1.4) Verbrauchsmaterialien eingesetzt (siehe Abb. 10).

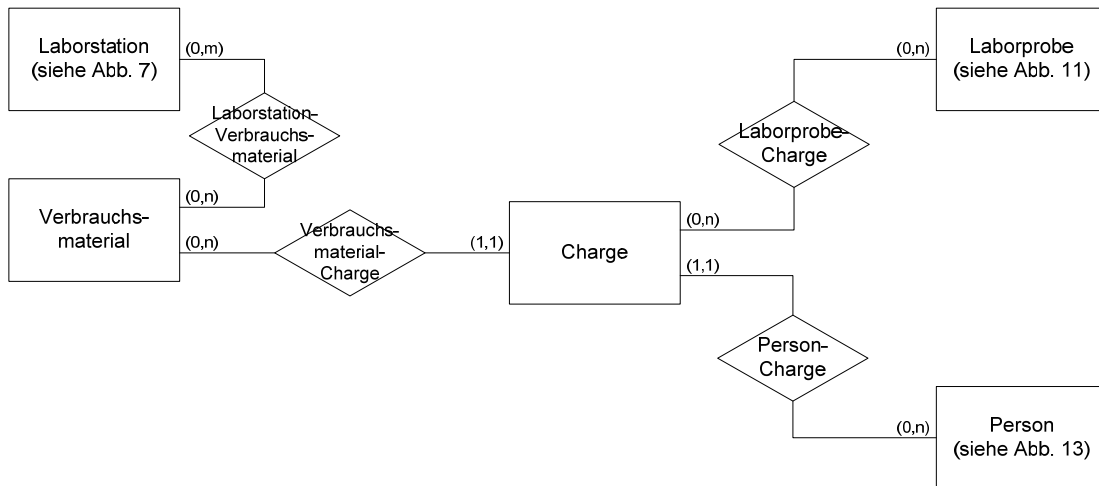


Abb. 10: An einer Laborstation werden Verbrauchsmaterialien für die Bearbeitung der Laborproben verwendet. Für die Qualitätskontrolle werden die Verbrauchsmaterialien in Chargen eingeteilt, die den Laborproben verbrauchsbezogen zugeordnet werden.

Die Zuordnung Laborstation-Verbrauchsmaterial definiert, welche Verbrauchsmaterialien an welchen Laborstationen eingesetzt werden. Zur Qualitätssicherung werden die Verbrauchsmaterialien weiter in Chargen eingeteilt (siehe Kap. 3.2.3.5). Chargen sind i.d.R. definiert als ein oder mehrere Reaktionsgefäße der angesetzten Verbrauchsmaterialien (Verbrauchsmaterial-Charge). Die den Chargen zugeordnete Person (siehe Kap. 3.1.2.1) ist für die jeweiligen Chargen verantwortlich (Person-Charge).

3.1.1.4 Proben und Laborproben im ReSeqLIMS

Das ReSeqLIMS unterscheidet zwischen Proben, die von einer Person eingesetzt werden, und Laborproben, die im Labor bearbeitet werden (siehe Abb. 11).

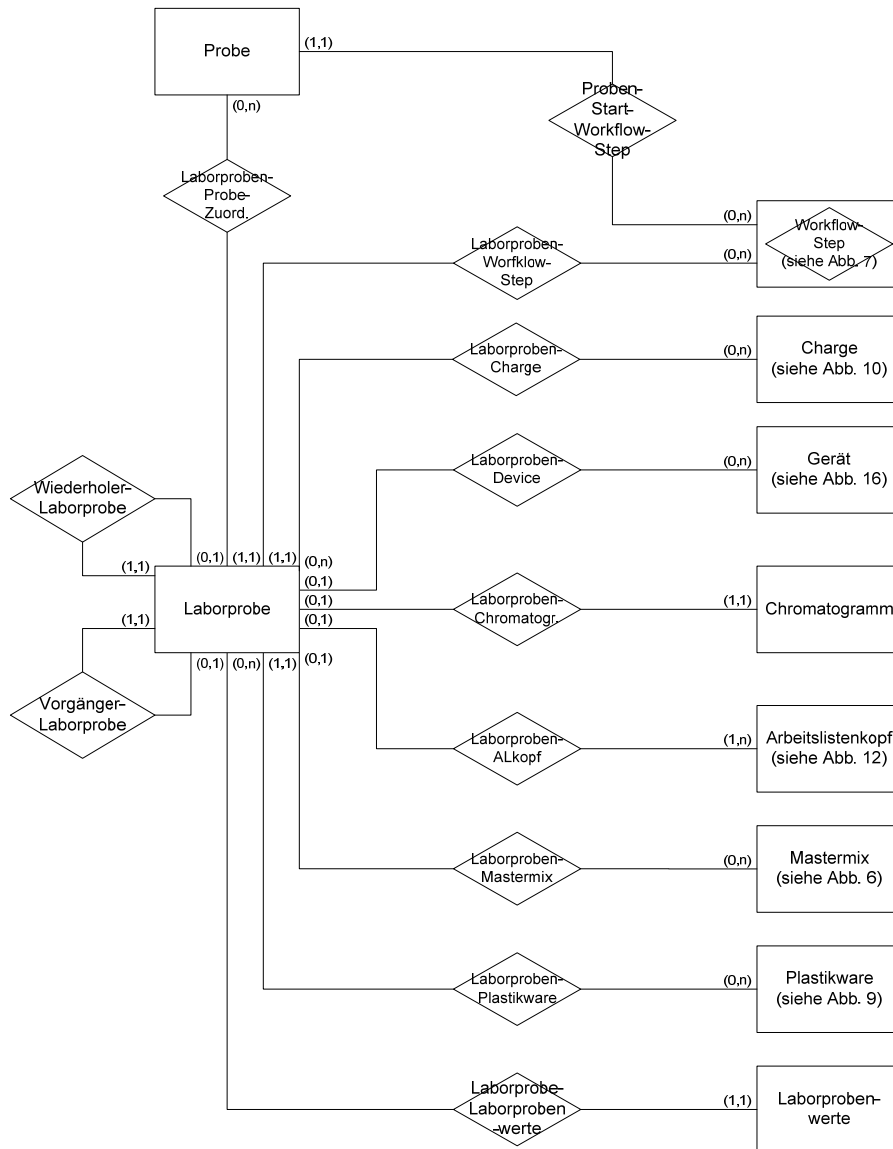


Abb. 11: Eine Probe besitzt immer einen startenden Workflow-Step, an dem die Bearbeitung im Labor beginnt. Die Laborproben durchlaufen Workflow-Steps mit einem definierten Ziel, z.B. der DNA-Sequenzierung.

Dadurch erfolgt eine strikte Trennung zwischen den Auftrags- und den Labordaten. Die Zuordnung von genau einem Workflow-spezifischen Laborstationspaar zu einer Probe definiert den Probe-Start-Workflow-Step, also die Laborstation, an der die Bearbeitung der Probe beginnt. Die möglichen Workflow-Steps, an denen die Bearbeitung der Probe beginnt, müssen in der Menge der Methoden-Material-Option-Plastikware-Workflow-Step Kombinationen aus Abb. 9 enthalten sein. Proben, die zur Bearbeitung im Labor freigegeben werden, werden als Laborproben bezeichnet. Jede Laborprobe kann eindeutig der (Ausgangs-)Probe zugeordnet werden (Probe-Laborprobe-Zuordnung). Der aktuelle Workflow-Step, der dem Bearbeitungsstand der Laborprobe

entspricht, wird durch die eindeutige Zuordnung der Laborprobe zu einem Workflow-Step definiert (Laborproben-Workflow-Step). An jedem Workflow-Step wird zu einer Laborprobe die Vorgänger-Laborprobe vermerkt, die nur im Falle der Start-Laborprobe in einem Workflow vorhanden ist. Jede Laborprobe befindet sich zu jeder Zeit in einem Workflow auf einer Plastikware (Laborprobe-Plastikware). Wird eine Laborprobe zur weiteren Verarbeitung ausgewählt und eine Arbeitsliste generiert, werden der zur Laborprobe gehörende Arbeitslistenkopf (siehe Kap. 3.1.1.5) sowie das Gerät (siehe Kap. 3.1.3) mit dem die Bearbeitung erfolgt, zugeordnet. Der Mastermix bzw. das Chromatogramm einer Laborprobe kann nur an spezifischen Workflow-Steps zugeordnet werden, da eine Laborprobe nur im PCR-Setup mit einem bzw. mehreren Mastermix verarbeitet wird bzw. die Zuordnung der Chromatogramme nur zu den sequenzierten Laborproben erfolgt. In beiden Fällen ist die Angabe von Mastermix und Chromatogramm zu der jeweiligen Laborprobe eindeutig. Zur Bearbeitung der Laborproben werden an den Laborstationen Verbrauchsmaterialien eingesetzt, die aus Gründen der Qualitätssicherung protokolliert werden müssen. Dieses erfolgt durch Zuordnung der aktuellen Charge eines Verbrauchsmaterials zu der jeweiligen Laborprobe. Die Verwendung von z.T. auch mehreren Verbrauchsmaterialien an einer Laborprobe ist optional, da z.B. keine Zuordnung erfolgt, wenn die Protokollierung von Verbrauchsmaterialien an einer Laborstation nicht notwendig ist. Treten dennoch qualitative oder technische Mängel während der Laborprozesse auf, kann das Wiederholen von Laborproben an spezifischen Workflow-Steps erforderlich sein (siehe Kap. 3.2.3.4). Für die Wiederholung von Laborproben muss die zu wiederholende Laborprobe, z.B. die bereits sequenzierte Laborprobe nach der DNA-Sequenzierung, und die zu reaktivierende Laborprobe an der Laborstation, ab der die Wiederholung erfolgen soll, unterschieden werden. Die zu wiederholende Laborprobe und die zu reaktivierende Laborprobe müssen über die Vorgänger-Laborproben Verbindung evtl. auch über mehrere Workflow-Steps miteinander in Beziehung stehen. An der reaktivierten Laborprobe wird die zu wiederholende Laborprobe vermerkt (Wiederholer-Laborprobe). Die Laborprobenwerte der reaktivierten Laborprobe(n) werden extra gesichert, da bei der anstehenden Wiederholung die alten Laborprobenwerte der reaktivierten Laborprobe mit neuen Laborwerten überschrieben werden.

3.1.1.5 Arbeitslisten

Die Bearbeitung der Laborproben erfordert i.d.R. den Einsatz von Geräten, Materialien und Verbrauchsmaterialien. Die Informationen über die Proben, die Reagenzien und Verbrauchsmaterialien, die be- und verarbeitet werden sollen, sowie die Mengen und das Bearbeitungsziel wird in Workflow-Step-spezifischen Arbeitslisten festgehalten. Eine Arbeitsliste besteht aus einem Arbeitslistenkopf, der allgemeine Informationen der Arbeitsliste enthält, und den Arbeitslistenpositionen (siehe Abb. 12). Einem Arbeitslistenkopf muss mindestens eine Arbeitslistenposition zugeordnet werden, eine Arbeitslistenposition kann nur in exakt einem Arbeitslistenkopf vorkommen. Eine Arbeitslistenposition ist eine Menge von Laborproben auf einer spezifischen Plastikware (Arbeitslistenposition-Plastikware). Es gilt, dass auch Teilmengen von Laborproben einer Plastikware in eine Arbeitsliste aufgenommen werden können, eine Arbeitslistenposition aber nicht auf mehrere Plastikwaren verteilt sein kann. Die Arbeitslistenpositionen einer Arbeitsliste können von unterschiedlichen Plastikwaren (Ausgangsplastikwaren) stammen. Die Arbeitslisten generierende Person ist definiert durch die Arbeitslistenkopf-Person Zuordnung. Neben methodenunabhängigen Laborstationen, wie z.B. der DNA-Normalisierung, ist an anderen Laborstationen wie z.B. dem PCR-Setup eine Unterscheidung der Laborproben nach den Methoden entsprechend der die Laborproben bearbeitet werden sollen notwendig, da sich z.B. die eingesetzten Reagenzien und Pipettiermuster je nach verwendeter Methode gravierend unterscheiden können. Wird eine Arbeitsliste an einer methodenspezifischen Laborstation generiert, wird durch die Arbeitslistenkopf-Methode die Methode der in einer Arbeitsliste enthaltenen Laborproben festgehalten. Die Arbeitslistenkopf-Plastikware definiert die Zielplastikware, auf der die Laborproben nach dem aktuellen Arbeitsschritt vorliegen werden. Es gilt, dass die Laborproben nach einem Arbeitsschritt nur auf einem Typ Plastikware, aber auf mehreren Plastikwaren desselben Typs vorliegen können.

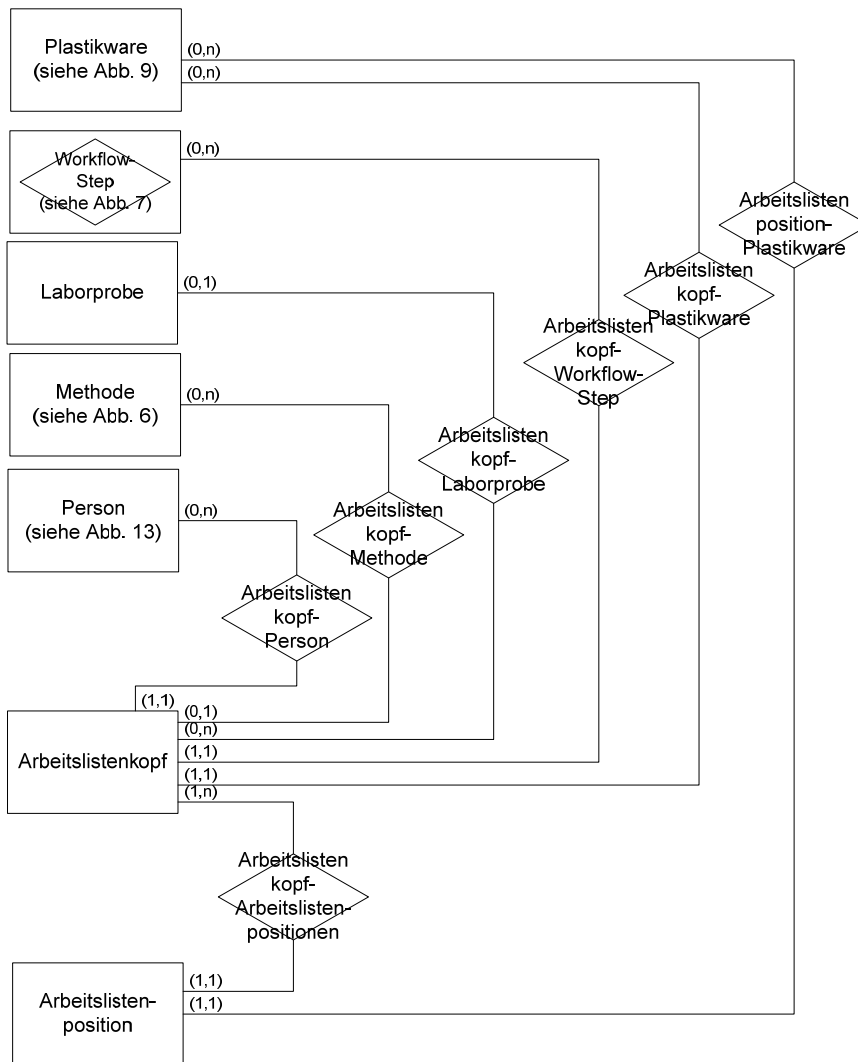


Abb. 12: Zur Bearbeitung von Workflow-Step-spezifischen Laborproben müssen Arbeitslisten generiert werden.

3.1.2 Datenmodelle des ReSeqLIMS-Kundenmoduls

3.1.2.1 Die Person als ein zentrales Objekt des ReSeqLIMS

Die Person ist ein zentrales Objekt des ReSeqLIMS. Eine Person muss sich u.a. an der Webanwendung anmelden, Aufträge (siehe Kap. 3.1.2.3) generieren und Arbeitslisten zusammenstellen können. Personen werden in externe und interne Personen differenziert. Während die externen Personen (Kunden) nur auf einen stark eingeschränkten Funktionsumfang Zugriff erhalten, stehen den internen Personen (Mitarbeitern) weitere Funktionalitäten zur Verfügung, die insbesondere das Labormodul betreffen. Die Mitarbeiter lassen sich bzgl. der zugeordneten Funktionalitäten in weitere Untergruppen ein-

teilen, je nachdem ob es sich z.B. um Doktoranden oder um technische Assistenten handelt. Generell werden zu einer Person Stammdaten wie z.B. Name, Vorname, Benutzername und Passwort abgespeichert (siehe Abb. 13). Die Adresse wird als eigenes Entity modelliert, um jeder Person neben einer Adresse eine optionale Rechnungsadresse zuzuordnen zu können. Jede Adresse kann exakt einer Person zugeordnet werden. Eine Person besitzt mindestens eine, im Falle der optionalen Rechnungsadresse maximal zwei Adressen. Durch die zwingend erforderliche Einordnung in exakt eine Nutzergruppe werden einer Person spezifische Zugriffsrechte auf Funktionalitäten zugeordnet (Personen-Nutzergruppe). Die Angabe exakt einer Organisation und optionaler Projekte zu einer Person erfolgt zur Zuteilung von Material-Option Kombinationen, Methoden und Konditionen (siehe Kap. 3.1.2.2). Die Organisation wird als Organisationshierarchie modelliert, die z.B. für die Zuordnung von Konditionen verwendet werden. Neben den Methoden und Material-Optionen, die für sämtliche Personen verfügbar sind, kann der Fall eintreten, dass bestimmte Methoden und Material-Optionen nur für eine spezifische Personengruppe oder für einzelne Personen verfügbar sein sollen. Personenspezifische Methoden und Material-Optionen sind den Personen direkt zuzuordnen, während für Personengruppen die Zuordnung zu Projekten erfolgt, da Projekte als Konstrukte zur Verwaltung organisationsübergreifender bzw. eingeschränkter Personengruppen verwendet werden. Die Methoden und Material-Optionen, die einem Projekt zugeordnet wurden, stehen sämtlichen Personen zur Verfügung, die wiederum dem Projekt zugeordnet wurden. Um Aufträge generieren zu können sind Barcodeaufkleber erforderlich, die zur Auszeichnung der einzusendenden Plastikware zu verwenden sind. Einer Person können beliebig viele Barcodes zugeordnet werden (Personen-Barcode).

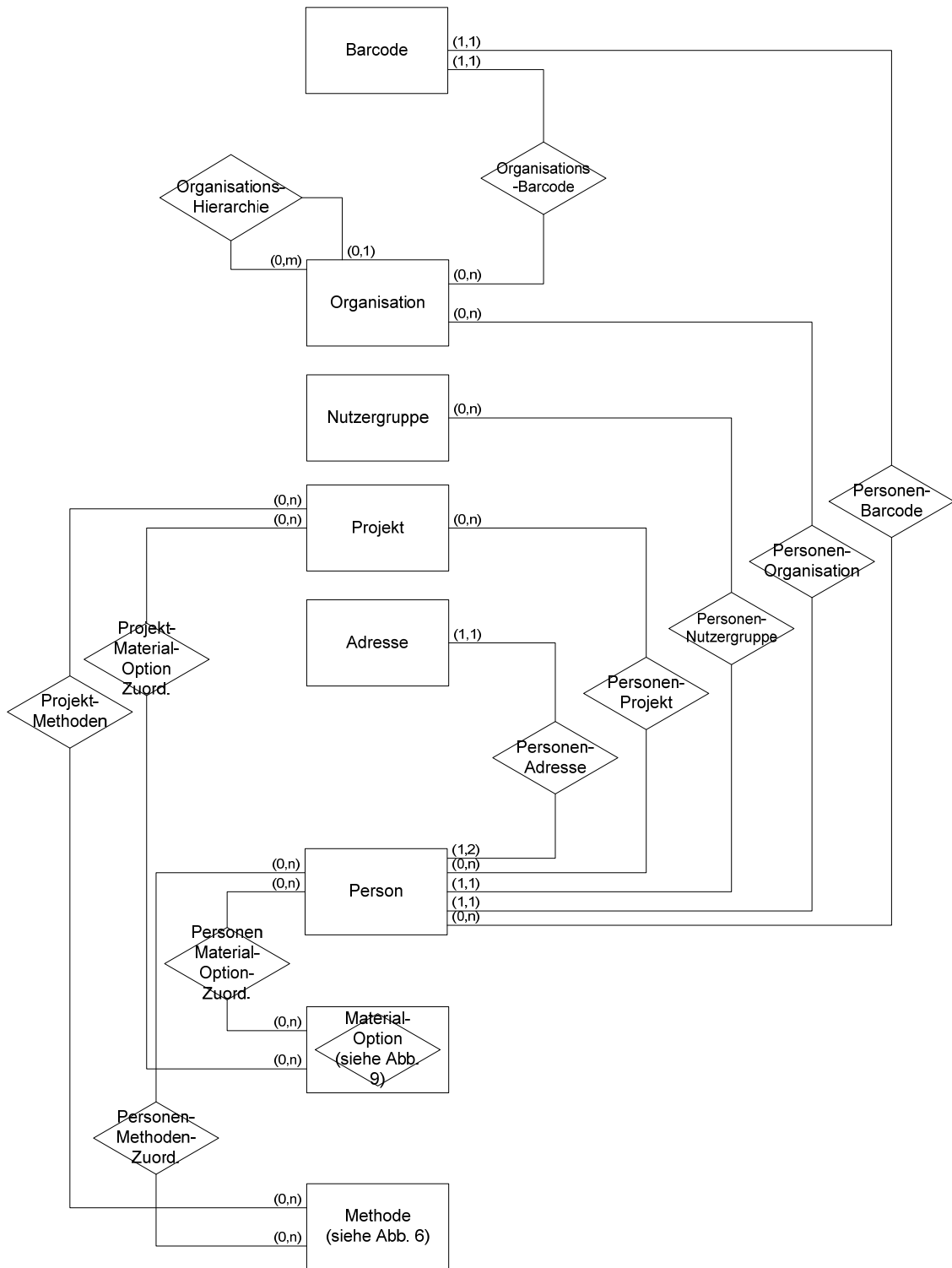


Abb. 13: Die Person ist ein zentrales Objekt des ReSeqLIMS.

3.1.2.2 Die Konditionen des ReSeqLIMS

Konditionen können auf den drei Ebenen Organisation, Projekt und Person angeben werden.

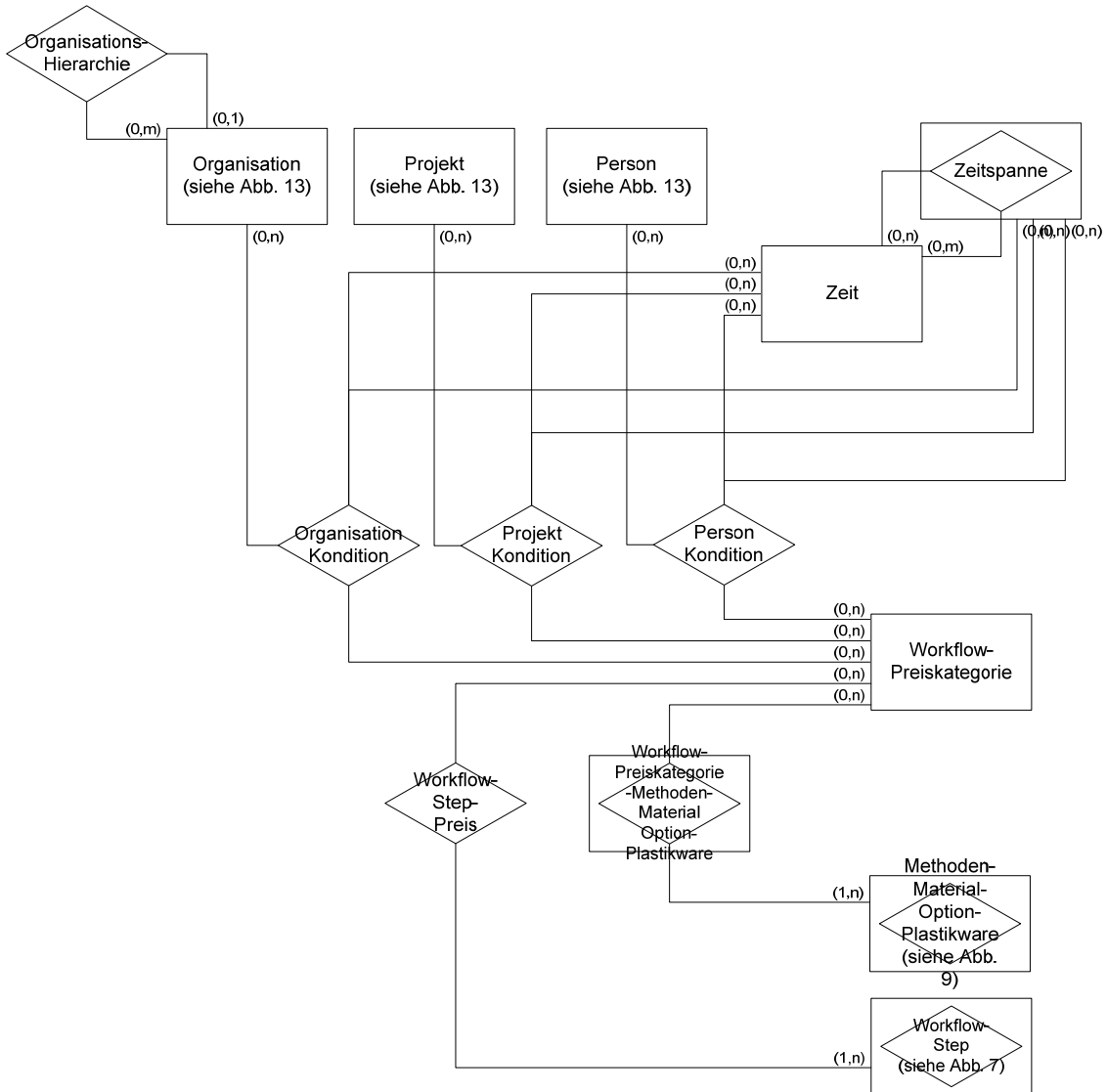


Abb. 14: Konditionen können auf den Ebenen Organisation, Projekt und Person definiert werden und beziehen sich i.d.R. auf spezifische Methoden-Material-Option-Plastikware Kombinationen.

Während Konditionen einer Organisation sämtlichen Personen dieser Organisation zur Verfügung stehen, wird das Konstrukt des Projektes dazu benutzt einer evtl. auch organisationsübergreifenden Personengruppe eine Kondition zuzuordnen. Konditionen, die einer Person zugeordnet werden, gelten nur für diese Person. Eine Kondition wird

definiert durch die Zuordnung einer Organisation, eines Projektes oder einer Person zu einer Workflow-Preiskategorie zu einer bestimmten Zeit (siehe Abb. 14). Die Modellierung der Zeit ermöglicht neben der Modellierung des Datums der Konditionserstellung auch die Angabe einer Zeitspanne, die als Struktur modelliert ist und die die Konditionsgültigkeit definiert. Die Abrechnungsmöglichkeiten einer Methoden-Material-Option-Plastikware Kombination werden durch die Zuordnung von mindestens einer Workflow-Preiskategorie definiert. Die exakten Preise einer Workflow-Preiskategorie (Workflow-Step-Preise) werden festgelegt durch Zuordnung der Workflow-Steps zu einer Workflow-Preiskategorie. Die Aufsummierung der Preise für die einzelnen Workflow-Steps von dem startenden bis zum abschließenden Workflow-Step bzgl. der Laborproben, die den Proben eines Auftrages zuzuordnen sind, ergibt den Gesamtpreis eines Auftrages.

3.1.2.3 Der Auftrag des ReSeqLIMS

Die Person-Labor-Interaktion beginnt mit einem Auftrag, der sowohl von den Kunden als auch von den Mitarbeitern generiert werden muss, um Proben analysieren zu können. Der Auftragskopf, der allgemeine Informationen eines Auftrages enthält, die für alle Auftragspositionen gelten, ist definiert durch die Beziehung von einer Person und der Zeit (siehe Abb. 15). Dadurch sind Auftragszeitpunkt und auftragsgenerierende Person eindeutig definiert. Aufträge können von Mitarbeitern für Kunden und für andere Mitarbeiter generiert werden. Die Person, die einen Auftrag für eine andere Person generiert, wird durch die optionale Zuordnung des Auftragskopfes zu einer Person festgelegt (Auftragskopf-Person). Eine Person verfügt maximal über zwei Adressen (siehe Kap. 3.1.2.1). Die für einen Auftrag zu verwendende Rechnungsadresse ergibt sich durch die obligate Auftragskopf-Adresse Zuordnung. Die Angabe von E-Mail-Adressen, die zusätzlich in die Labor-Person-Kommunikation bzgl. eines spezifischen Auftrages mit einbezogen werden sollen, ist optional. Jeder Auftrag enthält mindestens eine Auftragsposition. Eine Auftragsposition ist definiert als Verbindung von Auftragskopf und einer Methoden-Material-Option-Plastikware Kombination. Damit ist eindeutig festgelegt, welches Material mit welcher Eigenschaft auf welchem Trägermedium mit welcher Methode bearbeitet werden soll. Die Angabe des DNA-Stranges bzw. von DNA-Strängen, die zu sequenzieren sind, ist optional (Auftragspositions-DNA-Strang). Für Methoden, die auf der DNA-Sequenzierung basieren, ist für alle eingesendeten Ma-

materialien, die nicht direkt sequenziert werden können, der bzw. die DNA-Stränge anzugeben. Proben die direkt sequenziert werden können enthalten die Sequenzierprimer bereits. Damit ist die Angabe der zu sequenzierenden DNA-Stränge auch bei Methoden, die auf der DNA-Sequenzierung basieren, obsolet.

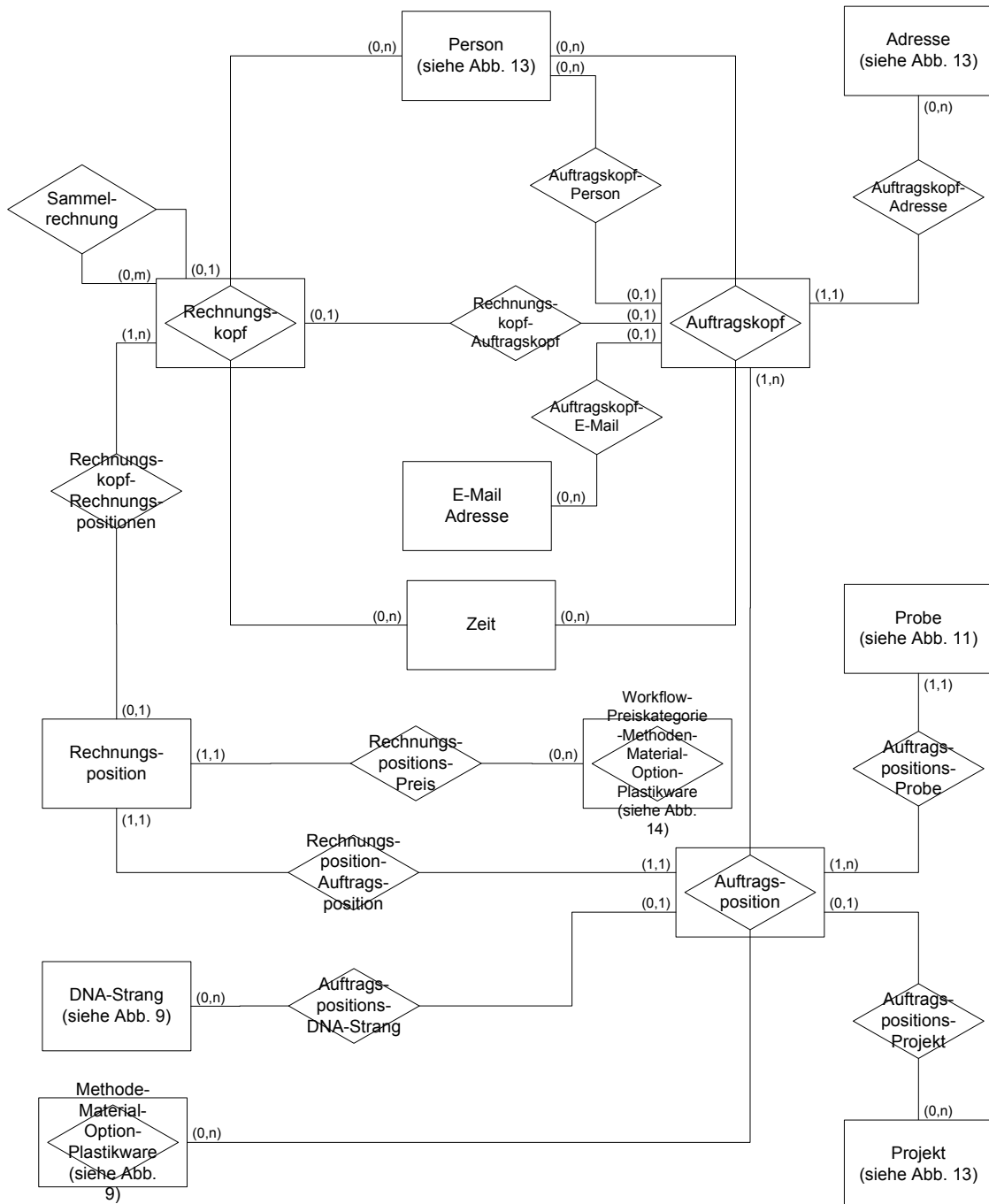


Abb. 15: Der ReSeqLIMS-Auftrage ist der Auslöser der Kunden-Labor Interaktion.

Die Zuordnung eines Projektes ist optional und wird aus abrechnungstechnischen Gründen vorgenommen (Auftragspositions-Projekt) (siehe Kap. 3.1.2.2). Jeder Auftragsposition muss mindestens eine Probe zugeordnet werden (Auftragspositions-Probe), die durch ihren Barcode, bei Mikrotiterplatten zusätzlich durch ihre Plattenposition, eindeutig identifiziert werden kann.

Eine Rechnung wird für einen vollständig bearbeiteten Auftrag generiert. Der Rechnungskopf ist dem Auftragskopf entsprechend definiert durch die Beziehung einer Person und eines Zeitpunktes. Ein Rechnungskopf kann max. einem Auftragskopf zugeordnet werden, ein Auftragskopf wird max. einem Rechnungskopf zugeordnet. Eine Sammelrechnung wird durch das Zusammenfassen mehrerer Rechnungsköpfe in einen gemeinsamen Rechnungskopf erstellt. Während einem Rechnungskopf mindestens eine Rechnungsposition zugeordnet ist, steht eine Rechnungsposition in Beziehung zu exakt einer Auftragsposition, da eine Auftragsposition nur über genau eine Rechnungsposition abgerechnet wird. Zur Definition der gewährten Kondition für eine Auftragsposition erfolgt die Zuordnung einer Rechnungsposition zu einer Workflow-Preiskategorie-Methode-Material-Option-Plastikware Kombination (Rechnungspositions-Preis).

3.1.3 Datenmodell des ReSeqLIMS-Gerätmoduls

Das Gerätmodul des ReSeqLIMS verwaltet die zur Bearbeitung von Methoden eingesetzten Geräte eines Labores (siehe Abb. 16). Ein Gerät muss genau einer Gerätegruppe zugeordnet werden, eine Gerätegruppe kann mehrere Geräte beinhalten. In der Zuordnung Gerätegruppe zu einer bestimmten Laborstation wird festgelegt, welche Geräte an einer Laborstation verwendet werden können (Gerätegruppe-Laborstation). Je nach verwendetem Gerät an einer Laborstation wird in der Geräte-Plastikware-Laborstation Beziehung definiert, welche und wie viele einer bestimmten Plastikware an einer Laborstation auf einem Gerät als Ausgangs- und Zielplastikware eingesetzt werden können. Einem Gerät können optional Protokolle zugeordnet werden, die bei einem Geräteeinsatz zur Verwendung kommen und ausgewählt werden können.

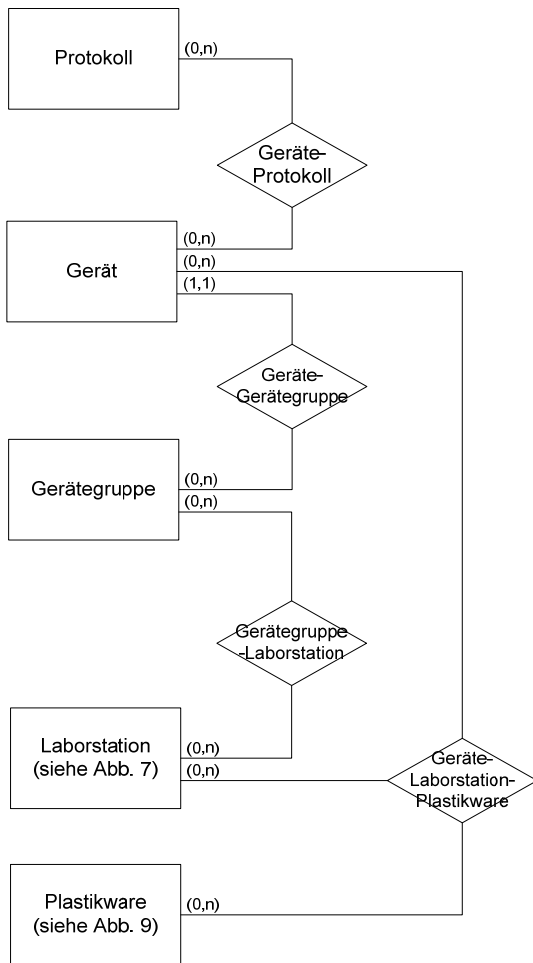


Abb. 16: Das ReSeqLIMS-Gerätemodul verwaltet gerätespezifische Eigenschaften.

3.1.4 Modulübergreifende Datenmodelle

3.1.4.1 Die Verwendung von Status zur Fortschrittsanzeige

Um den Fortschritt der Prozessierung von Aufträgen, Proben, Laborproben oder Rechnungen einfach und übersichtlich verfolgen zu können, erhalten die Objekte im ReSeqLIMS einen kontextsensitiven Status, der dem aktuellen Bearbeitungsstand der Objekte entspricht (siehe Anh. B). Der Status ist als eigenes Entity definiert (siehe Abb. 17). Jedes Objekt, das einen Status besitzt, kann immer nur einen spezifischen Status besitzen. Eine Ausnahme bilden die Laborstationen. Die Zuordnung von exakt zwei Status zu einer Laborstation definiert einen Status für Laborproben, die noch prozessiert werden müssen, sowie einen Status für Laborproben, die an der entsprechenden Laborstation bereits prozessiert wurden.

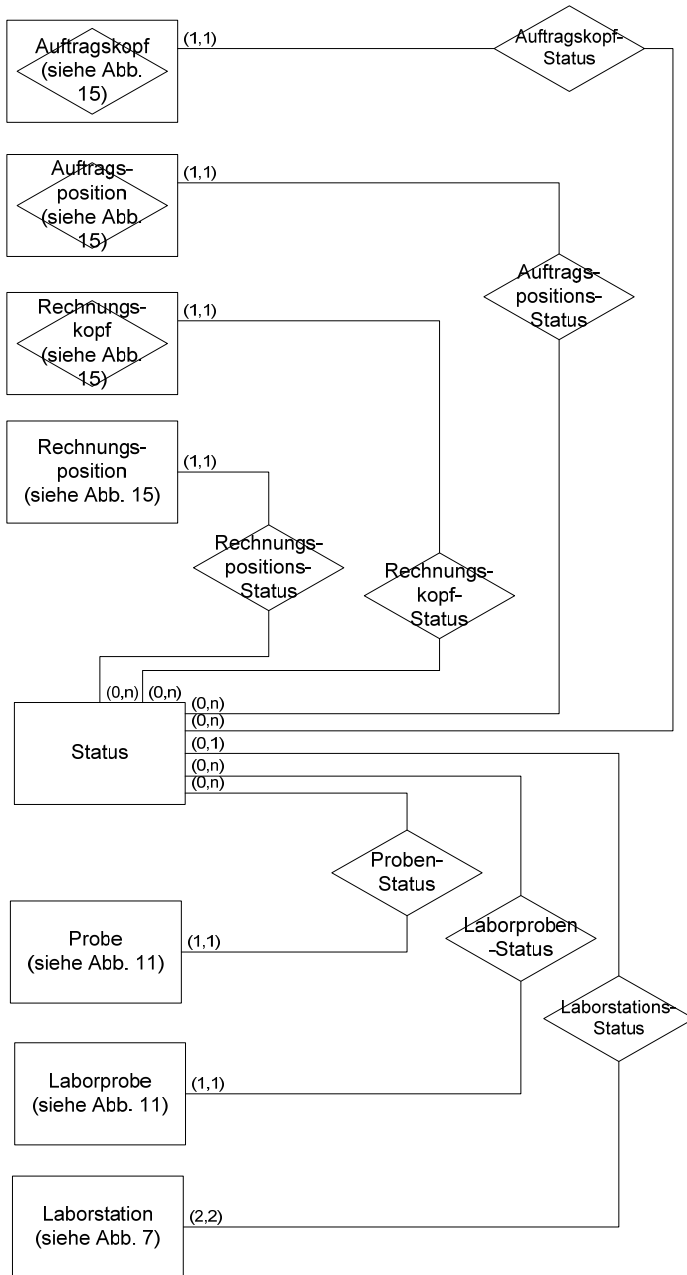


Abb. 17: Status geben im ReSeqLIMS Auskunft über den Fortschritt der Prozessierung von Aufträgen, Auftragspositionen, Proben, Laborproben und Rechnungen.

3.1.4.2 Konfiguration, Fehlermeldungen, Systembarcode und Datenbanksperr

Während die bisher beschriebenen ERM's sich mindestens in einem Entity überschneiden und sich damit in ein einziges ERM zusammenfassen lassen, gibt es einige Entities, die ohne Verbindung zu anderen Entities modelliert werden mussten. Diese

sind insbesondere für konfigurations- und aktionsbedingte Einstellungen der ReSeqLIMS Anwendung notwendig.

Das Entity Konfiguration beinhaltet allg. Konfigurationseinstellungen des ReSeqLIMS wie z.B. Ordnerpfade, Adressen und E-Mail-Adressen. Die Konfigurationseinstellungen sind über die ReSeqLIMS-Weboberfläche zu administrieren.

Die Nutzermeldungen werden zentral im Entity Nutzermeldungen modelliert und können z.B. bei fehlenden, falschen oder falsch formatierten Eingaben als Feedback und zur Benutzerführung angezeigt werden. Die Nutzermeldungen können über die ReSeqLIMS-Weboberfläche administriert werden.

Um den Mehrbenutzerbetrieb zu gewährleisten, werden Aktionen, in denen schreibend und lesend auf Datenbankdaten zugegriffen wird, exklusiv für einen Benutzer gesperrt. Um möglichst geringe Nutzungsbeschränkungen zu erhalten, werden spezifische Sperrbereiche definiert. Der sperrende Nutzer, der Zeitpunkt der Sperre und der Sperrbereich werden im Entity Arbeitsmodus modelliert. Gesperrte Bereiche werden bei Inaktivität des sperrenden Nutzers nach einer einzustellenden Zeitspanne automatisch wieder freigegeben.

Der Barcode des ReSeqLIMS kann in einen alphanumerischen und einen numerischen Bestandteil differenziert werden. Der alphanumerische Teil eines Barcodes besteht aus einem Buchstaben des Alphabets oder eines Sonderzeichens wie z.B. das @-Zeichen, wobei Umlaute ausgeschlossen sind. Für systemintern generierte Barcodes (siehe Kap. 3.2.3.3) muss der zu verwendende Buchstabe sowie der numerische Teil definiert werden. Beide Teile werden über das Entity Systembarcode modelliert.

3.2 Das ReSeqLIMS

3.2.1 Die ReSeqLIMS-Webanwendung

Die ReSeqLIMS-Webanwendung ist unter der Internetadresse <http://www.pba-zoo.de> im WWW durch die Verwendung eines Webbrowsers erreichbar. Inkompatibilitäten, die Auswirkungen auf die Darstellung oder Funktionalität der ReSeqLIMS-Webanwendung haben, sind bei der Verwendung des Microsoft Internet Explorer, des Mozilla Firefox, des Apple Safari und des Google Chrome (zu den jeweiligen Versionen siehe Kap. 2.3) nicht aufgetreten. Die ReSeqLIMS-Webanwendung strukturiert das Browserfenster in drei Abschnitte (siehe Abb. 18).

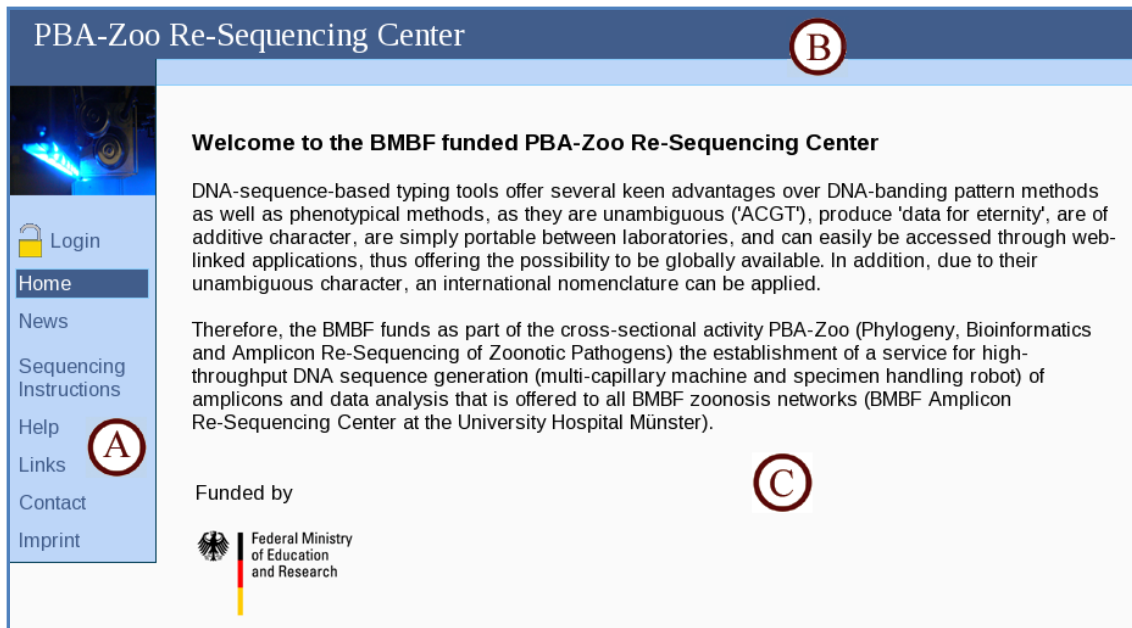


Abb. 18: GUI der ReSeqLIMS-Webanwendung. (A) zeigt die Navigationsleiste, (B) die Kopfzeile, der Inhalt und die Funktionalität befindet sich in (C).

Auf der linken Seite befindet sich die Navigationsleiste (A), die einen Zugriff auf die Hauptfunktionalitäten der Anwendung ermöglicht. Die Kopfzeile (B) enthält den Titel der ReSeqLIMS-Webanwendung. Der zentrale Bereich (C) ist der Content-Bereich, der zur Anzeige des Inhaltes und der Funktionalität genutzt wird. Als Sprache der ReSeqLIMS-Webanwendung wurde Englisch festgelegt. Die ReSeqLIMS-Webanwendung unterscheidet einen öffentlichen und einen nicht-öffentlichen Bereich. Der öffentliche Bereich umfasst Menüpunkte über die allgemeine Informationen z.B. über verwendete Sequenzierungsprotokolle, eine Hilfeseite und eine Kontaktmöglichkeit über ein Webformular zu erreichen sind. Auf der Hilfeseite sind grundlegende Informationen zur Verwendung der ReSeqLIMS-Webanwendung aufgeführt. Hier befinden sich neben zwei Tutorien, die in einer Flash-Animation den Registrierungs- und den Bestellprozeß erklären, auch die für einen Auftrag zu verwendenden Auftragstemplates (siehe Kap. 3.2.6.1) im Excel-Format, grundlegende Bedingungen an die einzusendenden Proben und die Anschrift, an die die Proben gesendet werden müssen. Für den nicht-öffentlichen Bereich ist eine Registrierung erforderlich.

Entsprechend des Datenmodells können im ReSeqLIMS ein Kunden-, ein Labor- und ein Gerätemodul unterschieden werden (siehe Kap. 3.1).

3.2.2 Das ReSeqLIMS-Kundenmodul

3.2.2.1 Die ReSeqLIMS-Registrierung

Der Zugang zum nicht-öffentlichen Bereich der ReSeqLIMS-Webanwendung ist nur nach vorheriger Registrierung möglich. Während der Registrierung werden die Personenstammdaten per Webformular erfasst (siehe Abb. 19).

The image contains two screenshots of the PBA-Zoo Re-Sequencing Center website. The left screenshot, labeled (A), shows the registration form titled 'Create a new user account (* mandatory field)'. It includes a sidebar with navigation links (Home, News, Sequencing Instructions, Help, Links, Contact, Imprint) and a 'Login' button. The form fields include: Title, Lastname*, Name*, PBA-Zoo username*, Gender* (with a dropdown menu set to 'Male'), Company/Institution*, Department/Institute, Street or P.O. Box*, Zip-Code*, City*, State, Country*, Telephone*, Fax, E-mail*, and Confirm E-mail*. There are also two additional e-mail fields and a comment box. At the bottom, there is a checkbox for 'Add a separate billing address' and buttons for 'Back', 'Reset', and 'Continue Registration'. The right screenshot, labeled (B), shows the confirmation message: 'Your account has been created successfully! It will be activated by our laboratory staff as soon as possible. An e-mail will inform you when your account is available for log in. Important: Your customer ID is 55. Click [here](#) to print a registration overview.' The sidebar and navigation links are identical to the first screenshot.

Abb. 19: Während der Registrierung an der ReSeqLIMS-Webanwendung werden in einem Webformular die Personenstammdaten erfasst (A). Nach der erfolgreichen Registrierung wird die Kundennummer generiert (B).

Durch das Abschließen der Registrierung wird ein neues Benutzerkonto angelegt, indem die Personendaten mit einer eindeutigen ID (Kundennummer) in der Datenbank abgespeichert werden. Im Gegensatz zu anderen Webanwendungen, bei denen nach erfolgreicher Registrierung das Anmelden an der Webanwendung und damit der Zugriff auf den nicht-öffentlichen Bereich direkt möglich ist, muss ein Benutzerkonto der ReSeqLIMS-Webanwendung zunächst durch einen Mitarbeiter aktiviert werden. Die manuelle Aktivierung eines Benutzerkontos ist erforderlich, da einer Person eine Nutzergruppe und eine Organisation zugeordnet werden muss (siehe Kap. 3.1.2.1). Während die Nutzergruppe die Funktionalität definiert, die einer Person nach einem Login zur Verfügung steht, ist die Zuordnung einer Organisation erforderlich, um die korrekte

Zuteilung von Konditionen auf Organisationsebene zu gewährleisten. Zusätzlich können bei der manuellen Benutzerkontenaktivierung die für Aufträge benötigten Barcodes bzw. Barcodeaufkleber einer Person zugeordnet und der Versand veranlasst werden. Als Nebeneffekt der manuellen Benutzerkontenaktivierung ist zudem beabsichtigt und zu erwarten, dass die Zahl der willkürlichen Registrierungen ohne reales Nutzungsinteresse stark eingeschränkt bzw. verhindert werden kann.

Der nach einem Login geladene Startbildschirm und die Anzahl der Menüpunkte der Navigationsleiste sind abhängig von der Nutzergruppe der angemeldeten Person (siehe Abb. 20).

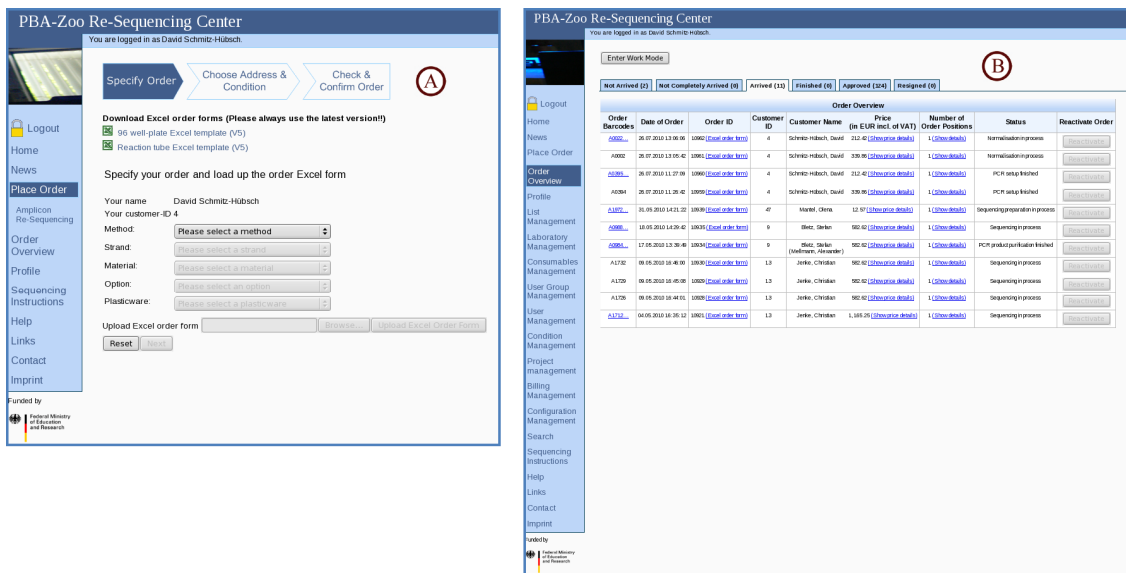


Abb. 20: Beispiele für Startbildschirme nach erfolgreichem Login an der ReSeqLIMS-Webanwendung. Kunden sehen den „Place Order“- (A), Mitarbeiter den „Order Overview“-Bildschirm (B). Die Menüpunkte der Navigationsleisten in (A) und (B) richten sich nach den Nutzergruppen, z.B. für den Kunden (A) und den Administrator (B).

Während der Administrator (siehe Abb. 20 (B)) auf alle Funktionen zugreifen kann, werden die Nutzungsrechte für z.B. Labormitarbeiter und Kunden (siehe Abb. 20 (A)) sukzessiv eingeschränkt. Eine Nutzergruppe im ReSeqLIMS besteht datenbankseitig aus einer Folge von booleschen Variablen, die jeweils eine Funktionalität repräsentieren. Durch das Setzen der booleschen Variablen wird die Funktionalität in der entsprechenden Nutzergruppe freigegeben. Auf der Ebene des GUIs werden die booleschen Variablen dazu benutzt, die Funktionalitäten nutzergruppenspezifisch ein- und

auszublenden. Die Profildaten der registrierten Anwender können über die ReSeqLIMS-Weboberfläche verwaltet werden. Von Änderungen ausgeschlossen sind ausschließlich der Benutzername und die Kundennummer, die für die eindeutige Identifikation einer Person verwendet werden.

3.2.2.2 Die ReSeqLIMS-Konditionsverwaltung

Die ReSeqLIMS-Konditionen können auf den drei Ebenen Organisation, Projekt und Person zugeordnet werden (siehe Kap. 3.1.2.2). Das Konstrukt des Projektes ermöglicht es neben organisations- und personenspezifischen auch organisationsübergreifende Konditionen an eine Gruppe von Personen zu vergeben. Generell können bei den Konditionen des ReSeqLIMS zeit- und mengenbezogene Konditionen unterschieden werden. Die zeitbezogenen Konditionen definieren Konditionsbedingungen für einen bestimmten, befristeten Zeitraum ohne Beschränkung der Auftragsmenge, während die mengenbezogenen Konditionen Konditionsbedingungen definieren, die sowohl für einen Zeitraum als auch für eine spezifische Auftragsmenge gelten. Durch die Zuordnung von Methoden-Material-Option-Plastikware Kombinationen zu Workflow-Preiskategorien, die Bestandteil der Organisations-, Projekt- oder Personenkondition sind, sind sowohl zeit- als auch mengenbezogene Konditionen immer auf eine oder mehrere Methoden-Material-Option-Plastikware Kombinationen bezogen (siehe Kap. 3.1.2.2). Für diese Kombinationen gelten die Workflow-Step-Preise, die der entsprechenden Workflow-Preiskategorie zugeordnet sind. Über das ReSeqLIMS können für Kunden und Mitarbeiter zeitbezogene Standardkonditionen definiert werden, die verwendet werden, wenn keine weitere Zuordnung von Konditionen zu Organisationen der Organisationshierarchie sowie Projekten der Person bzw. der Person selber vorgenommen wurde.

3.2.2.3 Der ReSeqLIMS-Auftrag

Nach der Registrierung an der ReSeqLIMS-Webanwendung ist der Auftrag Auslöser der Kunden-Labor Interaktion. Ein Auftrag ist definiert durch die Auftragsparameter, die mit einer Dropdown-Komponente festgelegt werden (siehe Abb. 20 (A)). Wie die überwiegende Mehrzahl der mit Werten vorbelegten JSF- oder RichFaces-Komponenten der ReSeqLIMS-Webanwendung, stammen die Auswahlelemente des Auftragsdropdowns aus der ReSeqLIMS-Datenbank. Die Besonderheit des Auftrags-

dropdowns ist die Verkettung mehrerer Dropdown-Listen. Durch die Verkettung wird erreicht, dass die Elemente einer Dropdown-Liste durch den selektierten Wert der vorangegangenen Dropdown-Liste bestimmt werden. Unzulässige Methoden-Material-Option-Plastikware Kombinationen können durch die verketteten Dropdown-Listen ausgeschlossen werden. Die logisch sinnvolle Auswahlreihenfolge ist durch das Sperren einer Dropdown-Liste, bis an der vorherigen Dropdown-Liste eine Auswahl erfolgte, für den Nutzer intuitiv festgelegt. Das Einlesen der Auftragsdatei (siehe Kap. 3.2.6.1), die die Probandaten des Auftrages enthält, ist erst nach der Auswahl der Auftragsparameter möglich. Während des Einlesens wird zur Vermeidung von nicht korrekten Eingaben überprüft, ob Material und Plastikware der Auftragsdatei mit den Auftragsparametern übereinstimmt. Nur bei einer Übereinstimmung werden die eingelesenen Proben zur Kontrolle der Probenbarcodes und -namen tabellarisch angezeigt. Für die Proben eines Auftrages verwendet das ReSeqLIMS standardmäßig die günstigste, zeitbezogene Kondition, die für die Methoden-Material-Option-Plastikware Kombination des Auftrages verfügbar ist, da für diese keine mengenmäßige Beschränkung besteht. Sind zusätzlich zu den zeitbezogenen auch quantitative Konditionen verfügbar, werden diese aufgrund der Mengenbeschränkung für eine aktive Auswahl durch den Auftraggeber tabellarisch aufgelistet. Aus den maximal zwei Adressen einer Person kann während des Bestellprozesses die als Rechnungsadresse zu verwendende Adresse ausgewählt werden. Eine Übersichtseite fasst die Auftragsinformationen übersichtlich zusammen, das Bestätigen der AGBs und das Absenden des Auftrages lässt einen rechtlich abgesicherten Auftrag zustande kommen. Der dreistufige Bestellprozeß für Kunden, der sich aufteilt in die Spezifikation der Auftragsparameter, der anschließenden Auswahl von Kondition und Rechnungsadresse und der abschließenden Auftragsbestätigung, ist für Mitarbeiter zeitsparend auf einen Schritt verkürzt. Alle Einstellungen können direkt auf einer HTML-Seite vorgenommen werden. Definieren Mitarbeiter eine Person für die sie einen Auftrag anlegen wollen, orientieren sich die auswählbaren Konditionen und Adressen nach der spezifizierten und nicht nach der angemeldeten Person.

Die Zuordnung der Barcodes bzw. Barcodeaufkleber, die in einem Auftrag zur Auszeichnung der Plastikware verwendet werden, erfolgt systemintern zu der Organisation, die in der Organisationshierarchie der Organisation einer Person am höchsten steht. Personen, die einer Organisation in der gleichen Organisationshierarchie zugeordnet sind, können so auf die identische Menge von Barcodeaufklebern zurückgreifen.

Dadurch können Personen, die einer Arbeitsgruppe angehören, eine gemeinsame Menge von Barcodeaufklebern verwenden. Die direkte Zuordnung einer Person zu einer Menge von Barcodeaufklebern definiert gleichzeitig einen eindeutigen Ansprechpartner, an den die Barcodeaufkleber versendet wurden. Die für einen Auftrag verwendeten Barcodeaufkleber werden systemseitig aus der Menge verfügbarer Barcodeaufkleber entfernt. Über das ReSeqLIMS können Mitarbeiter mit den entsprechenden Nutzungsrechten personenspezifische Barcodeaufkleber-Thresholds definieren. Wird die Anzahl der für eine Person verfügbaren Barcodeaufkleber unterschritten, wird systemseitig eine E-Mail generiert, die die Labormitarbeiter über das Unterschreiten des Grenzwertes informiert (siehe Kap. 3.2.2.5). Anschließend kann das Versenden neuer Barcodeaufkleber veranlasst werden.

Die Benennung der Proben eines Auftrages obliegt vollständig dem Auftraggeber. Dies ist insbesondere unter dem Aspekt automatisierter Downstream-Analysen generierter Chromatogramme von Bedeutung. Für eine geeignete Probenbenennung ist zu beachten, dass das ReSeqLIMS den Laborprobenamen systemseitig, im Amplikon-Resequenzierung Workflow z.B. nach Beendigung der Laborprobenbearbeitung an den Laborstationen PCR- und Sequenzierungssetup, um den Gen- und den Primernamen erweitert. Als Trennzeichen der Namensbestandteile wird jeweils der Unterstrich verwendet. Die sprechenden Dateinamen können von Software-Produkten wie z.B. Applied Maths BioNumerics [13] und Ridom SeqSphere [51] verwendet werden, um die für eine automatisierte Downstream-Analyse erforderliche Sortierung der Chromatogramme automatisiert durchzuführen. Die Sortierung erfolgt zunächst anhand des ersten Teiles des Probenamens, der vom Auftraggeber vorgegeben wird. Eine sinnvolle Benennung der Proben durch den Kunden entscheidet daher über die sinnvolle Kategorisierung der Chromatogramme in einer an die Resequenzierung anschließenden automatisierten Downstream-Analyse. Die kategorisierten Chromatogramme werden daraufhin anhand des Gennamens in Allelen angeordnet, für die automatisch eine Konsensussequenz berechnet wird. Der automatisierte Abgleich der Konsensussequenz mit einer Referenzdatenbank liefert die Allelnummer des identifizierten Allels zurück. Die Allelnummer Kombination einer Auftragsprobe definiert das Allelprofil und damit den resultierenden MLST-Sequenztyp [22].

3.2.2.4 Die ReSeqLIMS-Auftragsübersicht

Der Status eines Auftrages kann unter dem Menüpunkt „Order Overview“ abgefragt und verfolgt werden (siehe Abb. 21 X). Die Reiter zeigen die unterschiedlichen Auftragskategorien. Die Einteilung der Aufträge in die jeweiligen Kategorien erfolgt anhand des Auftragsstatus. Aufträge, die sich im Laborworkflow befinden, werden z.B. unter der Kategorie „Arrived“ angezeigt, bereits freigegebene Aufträge unter der Kategorie „Approved“. Die Darstellung der Aufträge erfolgt tabellarisch, wobei zu jedem Auftrag hierarchisch angeordnete Detailebenen abrufbar sind (siehe Abb. 21 (A) - (D)).

The screenshot displays the PBA-Zoo Re-Sequencing Center interface. At the top, it shows the user is logged in as David Schmitz-Hübisch. Below this, there are navigation tabs for order status: Not Arrived (0), Not Completely Arrived (0), Arrived (12), Finished (0), Approved (204), and Resigned (0). The main content area is divided into several hierarchical levels:

- Order Overview (A):** A table listing orders with columns for Order Barcodes, Date of Order, Order ID, Customer ID, Customer Name, Chromatogram Zip File, Price (in EUR incl. of VAT), Number of Order Positions, Status (marked with an 'X'), and a Reactivate Order button.
- Order Position(s) (B):** A table providing details for a specific order position, including Order Position Barcodes, Position, Plate Barcode, Material, Option, Plasticware, Method, Strand, Start Laboratory Station, Price Category, Price (in EUR excl. of VAT), Number of Samples, Status (marked with an 'X'), and a Reactivate Order Position button.
- Sample Detail(s) (C):** A table showing sample information with columns for Barcode, Position Plate, Sample Name, Sample Comment, Number of Repetition, Status (marked with an 'X'), Edit Sample, and Reactivate Sample.
- Chromatogram Detail(s) (D):** A table for chromatogram data with columns for Extended sample name, Chromatogram filename, Chromatogram date, MBQ, Q20 ratio, Sequence, and Status (marked with an 'X').

Additional features include a 'Download Chromatogram' link (E) and a 'Show Further Details' button (D) for the chromatogram data.

Abb. 21: In der Auftragsübersicht kann der Status (X) eines Auftrages eingesehen, Detailinformationen angezeigt ((A) - (D)) und z.B. die Chromatogramme in einer gezippten Ergebnisdatei heruntergeladen werden (E).

Auf der ersten Ebene (A) stehen die allgemeinen Auftragsinformationen, die den Attributen des Auftragskopfes und zusammenfassbaren Informationen der Auftragspositionen entsprechen. Die zweite Ebene (B) beinhaltet Informationen der einzelnen Auftragspositionen inkl. zusammenfassbarer Informationen der Proben, während die Probeninformationen in der Detailansicht auf der dritten Ebene (C) angezeigt werden. Das Laden und Anzeigen der Tabellen erfolgt erst bei einem Aufruf der jeweiligen Detail Ebenen. Dadurch sind die Ladezeiten auf das Notwendigste reduziert. Aus Gründen der Übersichtlichkeit ist immer nur die Detailebene bzw. Detailebenenhierarchie eines Auf-

trages sichtbar. Bei den freigegebenen Aufträgen kann zusätzlich eine Detailebene (D) nach der Probenebene (C) aufgerufen werden, die Informationen der sequenzierten Laborproben inklusive der verlinkten Ergebnisse, z.B. der Chromatogramme, anzeigt. Die Ergebnisse eines Auftrages stehen auf Auftragsebene als gezippte Ergebnisdatei zum Download zur Verfügung (siehe Abb. 21 (E)). Die Struktur der gezippten Ergebnisdatei ist angelehnt an den Auftrag und dessen beinhaltender Proben (siehe Anh. C, Abb. 34).

Das Rechnungswesen schließt einen Auftrag bzw. mehrere Aufträge mit der Erstellung einer Rechnung bzw. Sammelrechnung ab (siehe Kap. 3.2.6.3). Die für eine Auftragsposition verwendete Kondition wird mit der Rechnungsposition, die der Auftragsposition direkt zugeordnet ist, in der Datenbank abgespeichert und gewährleistet die korrekte Abrechnung im Rechnungswesen. Auch das Rechnungswesen arbeitet mit Status. Neben noch nicht verrechneten Aufträgen („Unbilled“) gibt es offene („Open“) und bezahlte („Paid“) Rechnungen (siehe Anh. B). Die Statusveränderungen der Rechnungen sind ausschließlich für den Mitarbeiter einsehbar. Nach der Rechnungserstellung durch einen Mitarbeiter wird die generierte PDF-Rechnungsdatei automatisch an die für die Rechnungen verantwortliche Person gesendet und von dieser nach einer Überprüfung dem Kunden per Post zugestellt. Die Rechnungsparameter, wie die für Rechnungen und Rückfragen verantwortliche Person sowie die Zahlungsfrist, sind über die ReSeqLIMS-Weboberfläche konfigurierbar.

3.2.2.5 Die ReSeqLIMS-E-Mail-Kommunikation

Die Kunden-Labor Interaktion wird unterstützt durch eine automatisierte E-Mail-Kommunikation (siehe Abb. 22). Ziel der systemgestützten E-Mail-Kommunikation ist die unmittelbare Benachrichtigung der Kunden, z.B. bei eingegangenen Aufträgen oder bereitgestellten Auftragsergebnissen, ohne dass von Mitarbeitern manuell eingegriffen werden muss. Nur die Beantwortung der Kontaktformular E-Mails erfolgt weiterhin manuell durch einen Mitarbeiter.

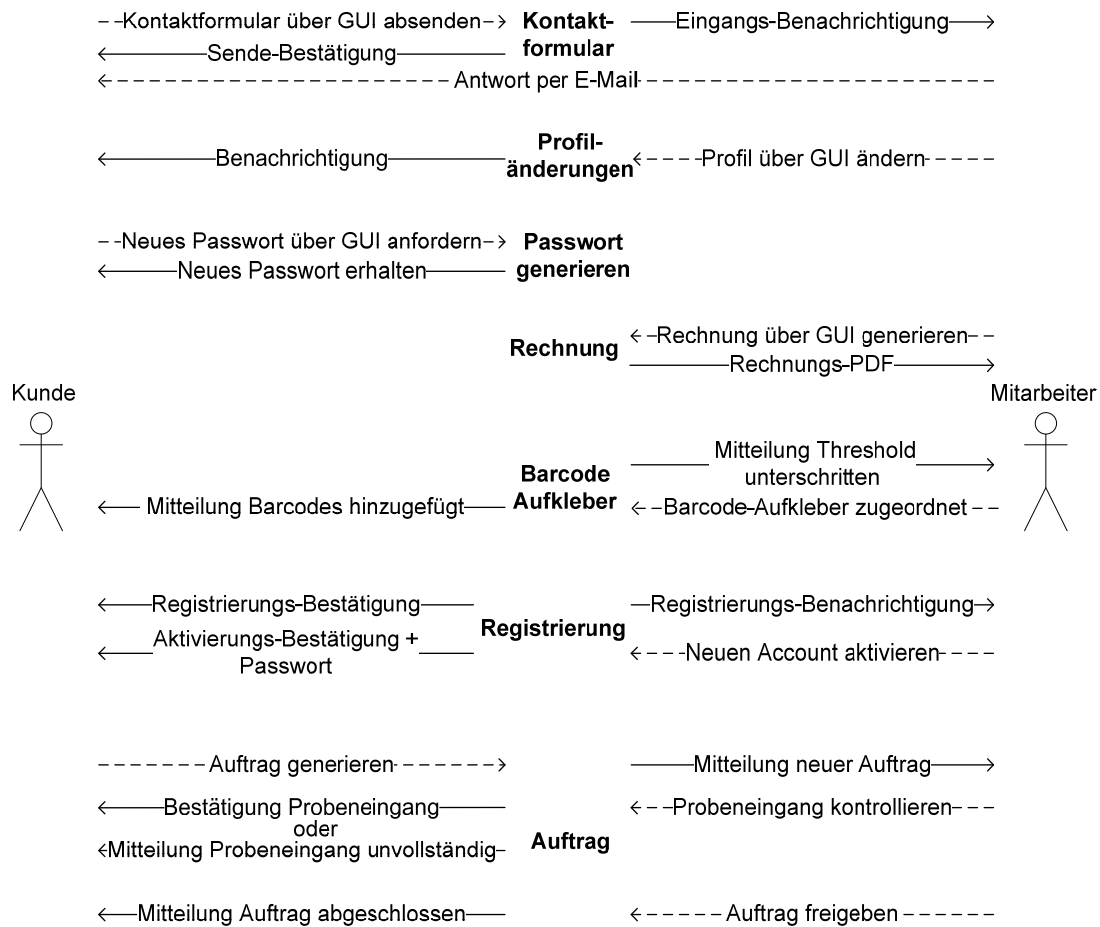


Abb. 22: Darstellung der automatisierten E-Mail-Kommunikation des ReSeqLIMS. Durchgezogene Linien repräsentieren systemgenerierte E-Mails, gestrichelte Linien repräsentieren manuelle Aktionen durch den Mitarbeiter oder den Kunden.

3.2.3 Das ReSeqLIMS-Labormodul

3.2.3.1 Allgemeines zum ReSeqLIMS-Labormodul

Das ReSeqLIMS-Labormodul ist mit seinem Listenmanagement („List Management“) der Arbeits-/Sequenzier- und Ausgabelisten das umfangreichste Modul des ReSeqLIMS. Ziel des ReSeqLIMS-Labormoduls ist die lückenlose Verfolgung der Proben bzw. Laborproben und Dokumentation aller Arbeitsschritte eines Workflows, von den Proben eines Auftrages bis zu den Ergebnissen, um eine optimale Qualität und Prozesssicherheit zu garantieren. Die Arbeitsschritte eines Workflows sind durch die Integration der Geräte eines Labores sowie die Übernahme unterstützender Aufgaben wie z.B. das Erstellen der Arbeitslisten bestmöglich zu unterstützen, um den Hochdurchsatz von Laborproben zu ermöglichen. Die Erfassung und Verfolgung der Proben von ihrem Ein-

treffen im Labor bis zur Auftragsfreigabe nach der Qualitätsprüfung der Resequenzierungsergebnisse basiert auf der Verwendung von Barcodes. Bereits beim Eintreffen der Proben eines Auftrages im Labor werden die Probenbarcodes erfasst und mit den Probenbarcodes der Auftragsdatei systemseitig verglichen. Eine Freigabe der Proben für das Labor ist erst nach der vollständigen und korrekten Erfassung der Auftragsbarcodes möglich. Nach der Freigabe für das Labor ist jede Laborprobe durch ihren Barcode (bei Verwendung von Reaktionsgefäßen) bzw. zusätzlich der Plattenposition (bei Verwendung von Mikrotiterplatten) und ihrem Workflow-Step eindeutig identifizierbar. Der Workflow-Step ist zur eindeutigen Identifizierung erforderlich, wenn Laborproben an einer Laborstation nicht auf eine neue Plastikware pipettiert werden und dadurch keinen neuen Barcode erhalten, wie es z.B. bei der PCR-Produktaufreinigung vorkommen kann. Hier liegen anschließend systemtechnisch zwei Laborproben vor, eine unaufgecleanigte und eine gereinigte Laborprobe, die anhand ihres Barcodes und ihrer Plattenposition nicht mehr unterschieden werden können. Wird zusätzlich der Workflow-Step zur Identifikation verwendet, liegen auch systemseitig zwei differenzierbare Laborproben vor. Für eine lückenlose Verfolgung der Laborproben von der Ausgangsprobe bis zu den Ergebnissen in einem mehrstufigem Workflow (im Amplikon-Resequenzierung Workflow z.B. den Chromatogrammen), ist die Verkettung der Laborproben erforderlich (siehe Abb. 23).

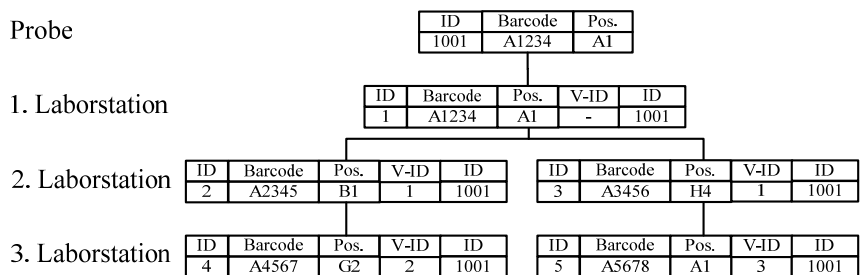


Abb. 23: Verkettung der Laborproben über mehrere Laborstationen. An der ersten Laborstation findet eine Laborprobenvervielfältigung statt (Pos.: Platten-Position, V-ID: Vorgänger-Proben-ID).

Werden die Laborproben an einer Laborstation bearbeitet, wird die ID der ursprünglichen Laborprobe bei der Speicherung der neuen Laborprobe(n) vermerkt. Durch die Vervielfältigung von Laborproben, z.B. beim PCR-Setup, ergibt sich eine Baum-

struktur, in der zu jeder Zeit alle Laborproben einer Probe in der Workflow-spezifischen Reihenfolge der Laborstationen nachverfolgt und anhand ihres Barcodes, evtl. zusätzlich der Plattenposition, und dem Workflow-Step eindeutig identifiziert werden können. Die eindeutige Identifizierung der Laborproben über alle Workflow-Steps eines Workflow ermöglicht, durch die Zuordnung von verwendeten Geräten, Verbrauchsmaterialien und Reagenzien, eine umfassende Dokumentation zu Zwecken der Qualitätssicherung der Laborprozesse.

3.2.3.2 Der ReSeqLIMS-„List Management“-Bildschirm

Der „List Management“-Bildschirm kann in vier Fensterbereiche unterteilt werden (siehe Abb. 24 (A) - (D)), wobei die Anzeige einiger Fensterbereiche optional ist. Das Arbeitslisten-Fenster (A) wird Workflow-übergreifend angezeigt, wenn eine Arbeitsliste für den Pipettierroboter zur Bearbeitung ausgewählt wurde. Wird mehr als ein Pipettierroboter verwendet, müssen alle in Bearbeitung befindlichen Arbeitslisten angezeigt werden. Im Workflow-Fenster (B) werden die Laborstationen des ausgewählten Workflows in der Workflow-spezifischen Reihenfolge angezeigt. Die Zahlen in den Klammern hinter jeder Laborstation geben Auskunft über wartende, aktuell in Bearbeitung, verworfene und zu wiederholende Laborproben. Falls vorhanden, werden im nachfolgenden Plastikwaren-Fenster (C) für den Amplikon-Resequenzierung Workflow die Plastikwaren angezeigt, die noch sequenziert werden müssen, oder bei denen nach erfolgter Sequenzierung die Chromatogramme den sequenzierten Laborproben noch zuzuordnen sind. Das Auftrags-Fenster (D) zeigt Workflow-übergreifend die Aufträge, die bereits abgeschlossen, aber noch nicht freigegeben wurden.

PBA-Zoo Re-Sequencing Center
You are logged in as David Schmitz-Hübsch.

Logout

Home

News

Place Order

Order Overview

Profile

List Management

Laboratory Management

Consumables Management

User Group Management

User Management

Condition Management

Project management

Billing Management

Configuration Management

Search

Sequencing Instructions

Help

Links

Contact

Imprint

Funded by
Federal Ministry of Education and Research

WorkList Summary						
Barcode	Generated on	Chosen Target Plasticware	Generated by	Worklist Laboratory Station	Show Worklist Details	Download
A0002 (Well Plate (96)) A0022 (Reaction Tube) A0023 (Reaction Tube) A0024 (Reaction Tube) A0025 (Reaction Tube) A0026 (Reaction Tube)	02.08.2010 11:23:47	Well Plate (96) (13)	David Schmitz-Hübsch	DNA Quantification	Open Popup	Download

Read In OutputList
Read In Manual OutputList
Push to Hamilton Directory
Refresh
Work Mode

Read in Output List
Read in manual output list
Unpush Worklist
Refresh
Enter Work Mode

Select workflow: Amplicon Re-Sequencing

Amplicon Re-Sequencing List Management	
Laboratory Station (pending/processing/resign/repeat)	Lab Station Menu
DNA Quantification (5/13/0/0)	Open
DNA Normalisation (0/0/0/0)	Open
PCR Setup (0/9/0/0)	Open
PCR Product Purification (48/672/0/0)	Open
Sequencing Setup (96/576/0/0)	Open
Sequencing Product Purification (0/1152/0/0)	Open
Sequencing Preparation (14/968/0/0)	Open
Sequencing (0/1183/0/0)	Open

Waiting for Sequencing / Approvement					
Plasticware	Barcode	Counts (Selected)	Select Sample	Repeat Type	Detail View
Well Plate (96)	A0994	96(0)	<input type="checkbox"/>	Plate	Show details
Well Plate (96)	A1745	96(0)	<input type="checkbox"/>	Plate	Show details
Well Plate (96)	A1746	96(0)	<input type="checkbox"/>	Plate	Show details

Read In Chromatograms
Repeat Samples
Work Mode

Start
Repeat
Enter Work Mode

Finished Laboratory Stations		
Workflow	Laboratory Station (pending/resign/repeat)	Lab Station Menu
Amplicon Re-Sequencing	Finished (288/0/0)	Open
Fragment Analysis	Finished (0/0/0)	Open

Abb. 24: Der ReSeqLIMS-Bildschirm des „List Management“ ist unterteilt in das Arbeitslisten-Fenster (A), das Workflow-Fenster (B), das Plastikwaren-Fenster (C) und das Auftrags-Fenster (D).

Nach der Auswahl einer Laborstation im Workflow-Fenster, werden an den methodenunabhängigen Laborstationen wie z.B. der DNA-Normalisierung, die noch nicht bearbeiteten („Pending samples“), in Bearbeitung befindlichen („Samples in process“), verworfenen („Resigned“) und wiederholten Laborproben („Repeated“) durch Reiter

getrennt voneinander tabellarisch angezeigt. An den methodenabhängigen Laborstationen, z.B. des PCR- und Sequenzierungssetups, können nur Laborproben, die mit der gleichen Methode bearbeitet werden sollen zusammen verarbeitet werden. An diesen Laborstationen erfolgt nach der Auswahl einer Laborstation daher zunächst die Auswahl der Methode, gefolgt von der methodenspezifischen Anzeige der Laborproben in den bereits genannten Kategorien. Während im „Order Overview“ eine auftragsorientierte Darstellung der Proben erfolgt, werden die Laborproben im „List Management“ barcode- bzw. plastikwarenorientiert angezeigt (siehe Abb. 25 (A) - (B)).

The screenshot shows the 'PBA-Zoo Re-Sequencing Center' web interface. The user is logged in as David Schmitz-Hübisch. The interface displays a 'Laboratory Station: PCR Product Purification' view. A table shows the details for a specific method: 'C. perfringens partial phospholipase C sequencing'. Below this, there are controls for scanning barcodes and selecting samples. A table titled 'Plate/Reaction Tube Detail(s)' lists individual samples with columns for Order ID, Date of Order, Option, Method, Gene, Primer, Position Plate, Sample Name, Error Code, Sample Comment, Repeat Counter, Status, Select Sample, and Edit Sample. Callout (A) points to the 'Plasticware' column in the top table, and callout (B) points to the 'Order ID' column in the bottom table.

Method	Gene	Primer	Plasticware	Barcode	Number Carrier Reaction Tube	Position on Carrier Reaction Tube	Counts (Selected)	Select Sample	Detail View
C. perfringens partial phospholipase C sequencing	-	-	Well Plate (96)	A1324	0	0	48(0)	<input type="checkbox"/>	hide_details

Order ID	Date of Orde	Option	Method	Gene	Primer	Position Plate	Sample Nam	Error Code	Sample Comment	Repeat Counter	Status	Select Sample	Edit Sample
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	A1	10/vak#241#1cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	B1	10/vak#239#3cpaf	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	C1	10/vak#198#8cpar	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	D1	10/vak#176#1cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	E1	10/vak#175#1cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	F1	10/vak#175#4cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	G1	10/vak#175#5cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	H1	10/vak#155#1cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	A2	10/vak#151#1cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>
10945	11.06.2010 10:29	Not applicable	C. perfringens pa	-	-	B2	10/vak#149#2cpa	0		0	Arrived	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 25: Die Reiter einer Laborstationsansicht im „List Management“ zeigen die Laborproben Status-spezifisch an. Abschnitt (A) zeigt die Plastikwaren, (B) die darauf enthaltenen Laborproben.

Durch das Pipettieren von Laborproben aus Reaktionsgefäßen bzw. Mikrotiterplatten auf eine oder mehrere Mikrotiterplatten, verteilen sich die Laborproben eines Auftrages beim Durchlaufen des Laborworkflows auf mehrere Mikrotiterplatten, eine Mikrotiterplatte kann Laborproben eines oder mehrerer Aufträge beinhalten. Zur Reduzierung der Ladezeiten und zur Erhöhung der Übersichtlichkeit wurde eine zweistufige Darstellung der Plastikwaren definiert (siehe Abb. 25 (A) - (B)). Auf der ersten Ebene (A) werden allgemeine Informationen zu den Laborproben der Plastikware angezeigt, in

der Ansicht der Detailebene (B) werden die Daten der auf einer Plastikware enthaltenen Laborproben aufgelistet.

Für die Bearbeitung von Laborproben ist das Generieren einer Arbeitsliste (siehe Kap. 3.2.6.2) obligat. In Abhängigkeit von der Laborstation, der gewählten Zielplastikware und dem verwendeten Gerät ist die Anzahl der Ausgangsmikrotiterplatten/-reaktionsgefäße und Zielmikrotiterplatten beschränkt (siehe Kap. 3.1.3). Nach Auswahl der Laborproben für eine Bearbeitung wird die Anzahl der ausgewählten Ausgangs- und benötigten Zielplastikwaren überprüft. Nur bei einer erfolgreichen Plausibilitätsprüfung wird die Arbeits- oder Sequenzierliste erstellt. Kurzinformationen zu den erstellten und noch nicht bearbeiteten Arbeitslisten werden tabellarisch unter der Kategorie „Samples in process“ angezeigt. Zusätzlich können Detailinformationen zu jeder Arbeitsliste per Pop-up abgerufen und ausgedruckt werden (siehe Abb. 26 A-D).

PBA-Zoo Re-Sequencing Center

A Worklist name WL3_PCRSetup_20100212_082237.xls
 Worklist generated on 12.02.2010 07:22:37
 Worklist generated by Ramminger, Isabell
 Chosen target plate type Well Plate (96) (91)
 Number of target plate(s) 1

B

Worklist Info Positions		
Chosen Source Plasticware	Chosen Source Barcode	Number samples of chosen Barcode
Well Plate (96)	A0775	13

C

Consumables Usages Parameters				
PCR Mastermix Source Position	Genename	Needed PCR mastermix volume (µl)	Needed Reaction Tube(s) of PCR Mastermix	
1	adkEc	167.0	1	
2	fumCEc	167.0	1	
3	gyrBEc	167.0	1	
4	icdEc	167.0	1	
5	mdhEc	167.0	1	
6	purAEc	167.0	1	
7	recAEc	167.0	1	

D

Selected Plate Positions		
Source Barcode	Position Plate	Sample Name
A0775	F2	64#1
A0775	G2	67#1
A0775	H2	80#1
A0775	A3	80#2
A0775	B3	80#6
A0775	C3	88#1
A0775	D3	93#1
A0775	E3	93#5
A0775	F3	100#1
A0775	G3	100#2
A0775	H3	104#3
A0775	A4	104#4
A0775	B4	106#1

Print

Abb. 26: Das Arbeitslistenpopup an der Laborstation PCR-Setup enthält detaillierte Informationen u.a. bzgl. der Ausgangs- und Zielplastikwaren sowie der beinhaltenden Laborproben.

Das Arbeitslistenpopup informiert den Labormitarbeiter über die Anzahl und den Typ der Zielplastikware (A), die Ausgangsplastikwaren in einer Kurzzusammenfassung (B), optional evtl. verwendete Reagenzien mit der Angabe der benötigten Mengen und Positionen am Pipettierroboter (C) sowie einer tabellarischen Auflistung der Laborproben einer Arbeitsliste (D). Die Arbeitslisten werden am Pipettierroboter manuell ausgewählt. Um Verwechslungen zu vermeiden, muss die aktuelle Arbeitsliste aktiv im ReSeqLIMS ausgewählt werden, die anschließend in einen Arbeitsordner des Pipettierroboters kopiert wird, der sich in der ReSeqLIMS-Ordnerstruktur befindet (siehe Anh. A, Abb. 33). Die Auswahl der Arbeitsliste am Pipettierroboter erfolgt durch den direkten Zugriff auf den Arbeitsordner, da die ReSeqLIMS-Ordnerstruktur über ein Netzlaufwerk lokal eingebunden wurde. Das Auswählen und Kopieren weiterer Arbeitslisten über die ReSeqLIMS-Weboberfläche wird daraufhin bis zur Freigabe der aktuellen Arbeitsliste bzw. bis die vom Pipettierroboter generierte Ausgabeliste eingelesen wurde blockiert. Die Prozessnummer in der Arbeitsliste (siehe Kap. 3.2.6.2), die mit der Modulnummer des Pipettierroboters übereinstimmt, wird zusätzlich auf Äquivalenz mit dem ausgewählten Modul verglichen, um die korrekte Wahl des gewählten Moduls am Pipettierroboter zu gewährleisten. Die Ausgabeliste (siehe Kap. 3.2.6.2) des Pipettierroboters wird in einen Ausgabeordner gespeichert, der der Laborstation des ausgewählten Moduls zuzuordnen ist (siehe Anh. A, Abb. 33). Das Einlesen der Ausgabeliste des Pipettierroboters in das LIMS erfolgt über das Arbeitslisten-Fenster des „List-Managements“. Nur bei Vorliegen einer Ausgabeliste in einem Ausgabeordner, der der Laborstation der ausgewählten Arbeitsliste entspricht, ist das Einlesen einer Ausgabeliste des Pipettierroboters über die ReSeqLIMS-Weboberfläche freigeschaltet. Um die eindeutige Zuordnung von aktueller Arbeitsliste und einzulesender Ausgabeliste sicher zu stellen, muss das Einlesen der Ausgabelisten manuell angestoßen werden und ist bei mehr als einer Ausgabeliste im entsprechenden Ausgabeordner nicht möglich.

Eine Ausnahme bei den Laborstationen bildet die Laborstation DNA-Sequenzierung z.B. im Amplikon-Resequenzierung Workflow. Anstelle der Arbeitslisten für den Pipettierroboter im Excel-Format werden hier speziell formatierte Sequenzierlisten für den ABI-Sequenzierer im TXT-Format generiert (siehe Kap. 3.2.6.4). Für die DNA-Sequenzierung muss die Sequenzierliste nicht ausgewählt werden, da ansonsten der vom DNA-Sequenzierer unabhängig arbeitende Pipettierroboter blockiert werden würde. Der DNA-Sequenzierer generiert keine Ausgabeliste, die anschließend ein-

gelesen werden kann. Nach der DNA-Sequenzierung liegt eine Menge von Chromatogrammen vor, die den sequenzierten Laborproben zugeordnet werden müssen. Für jedes Chromatogramm generiert der DNA-Sequenzierer zudem eine Sequenzdatei mit der DNA-Sequenz der sequenzierten Laborprobe und, bei hinreichender Qualität der Ergebnisse, eine Protokolldatei mit Informationen des Sequenzierprozesses. Die Benennung der Dateien erfolgt durch den DNA-Sequenzierer auf Basis des Namens der Laborprobe, dem zusätzlich der Barcode und die Position der Laborprobe auf der Sequenzier-Mikrotiterplatte angehängt werden. Die Zuordnung der Chromatogramme und der Zusatzinformationen zu der sequenzierten Laborprobe erfolgt über das Plastikware-Fenster des „List Managements“. Die Dateien werden dazu in einen spezifischen Ordner der ReSeqLIMS-Ordnerstruktur kopiert. Eine manuell zu aktivierende Funktion des ReSeqLIMS extrahiert anschließend Barcode und Mikrotiterplattenposition aus den Dateinamen, anhand derer mit Hilfe der Datenbank eine eindeutige Zuordnung der Chromatogramme und der Zusatzinformationen zu der sequenzierten Laborprobe möglich ist. Für eine Methode mit abschließender Fragmentlängen-Bestimmung können wie bei der DNA-Sequenzierung die Ausgabe- bzw. Fragmentdateien des DNA-Sequenzierers über den Dateinamen entsprechend zugeordnet werden.

3.2.3.3 Das manuelle Bearbeiten von Laborproben

In einem Hochdurchsatz-Laborworkflow kann es sinnvoll sein, Arbeitsschritte manuell durchzuführen und nicht auf den Pipettierroboter zurückzugreifen. Zur systemseitigen, lückenlosen Verfolgung und Protokollierung der Laborproben müssen auch die manuell durchgeführten Arbeitsschritte systemtechnisch im LIMS erfasst werden. Das ReSeqLIMS bietet eine Funktion, mit der durch das manuelle Generieren von Arbeits- und Ausgabelisten die manuelle Bearbeitung der Proben in eine Verarbeitung mit dem Pipettierroboter transformiert wird (siehe Abb. 27 (A) - (B)). Damit werden die Laborproben nicht nur lückenlos verkettet, sondern ist auch die Protokollierung der Reagenzien und Verbrauchsmaterialien für jede einzelne Laborprobe möglich. Nach Auswahl der manuellen Listengenerierung im Arbeitslisten-Fenster des „List Managements“ ist die Ziellaborstation, an der die Laborproben als „Pending samples“-Laborproben vorliegen sollen, und die verwendete Zielplastikware zu wählen sowie evtl. benötigte Daten wie die Konzentration und verwendete Normalisierungspuffer Mengen anzugeben (siehe Abb. 27 (A)). Durch die Angabe der Ziellaborstation können auch mehrere Workflow-

Steps übersprungen werden. Für die Erstellung der zwischen der Ausgangs- und der Ziellaborstation liegenden Arbeitslisten wird die als Zielplastikware angegebene Plastikware mit systemseitig generierten Barcodes verwendet. Die Zielausgabeliste, die identisch ist mit der Ausgabeliste der Laborstation, die im Workflow vor der Ziellaborstation liegt, wird in tabellarischer Form angezeigt (siehe Abb. 27 (B)). Die Reiter präsentieren die Anzahl der benötigten Zielplastikwaren, zu denen jeweils der Zielbarcode anzugeben ist. Das Speichern der Zielausgabeliste mit den angegebenen Barcodes bewirkt, dass alle Zwischenlisten und Zwischenlaborproben mit einem systemgenerierten Barcode versehen, generiert und abgespeichert werden. Die Simulation von Arbeitslisten beschränkt sich auf die Laborstationen, die mit dem Photometer und dem Pipettierroboter arbeiten. Eine Simulation der DNA-Sequenzierung und damit von Chromatogrammen ist nicht sinnvoll.

PBA-Zoo Re-Sequencing Center
 You are logged in as David Schmitz-Hübisch

Select Device: Photometer (A)
 Select Device: Hamilton STARlet
 Select Target Plasticware: Well Plate (96)
 Worklist Laboratory Station: DNA Quantification
 Select Target Laboratory Station: Sequencing
 Concentration: 0.0
 End Concentration User: 0.0
 UsedBufferVolume: 0.0
 End Volume: 0.0
 Sample Comment:
 Sorting Mode(t/r):

1. Well Plate (96) | 2. Well Plate (96)

Plate barcode: (B)
 Row by Row

Target Plasticv	Target Barcod	Target Plate Position	Source Plasticware	Source Barcode	Source Plate Position	Number Carrier Reaction Tube	Position on Carrier Reaction Tube	Sample Name	Concentration	Sample Comment
96Plate	@10000284	A1	96Plate	@10000282	A1	0	0	sample10_adkEc	0.0	
96Plate	@10000284	B1	96Plate	@10000282	B1	0	0	sample10_adkEc	0.0	
96Plate	@10000284	C1	96Plate	@10000282	C1	0	0	sample10_tumCE	0.0	
96Plate	@10000284	D1	96Plate	@10000282	D1	0	0	sample10_tumCE	0.0	
96Plate	@10000284	E1	96Plate	@10000282	E1	0	0	sample10_gyrBE	0.0	
96Plate	@10000284	F1	96Plate	@10000282	F1	0	0	sample10_gyrBE	0.0	
96Plate	@10000284	G1	96Plate	@10000282	G1	0	0	sample10_icdEc	0.0	
96Plate	@10000284	H1	96Plate	@10000282	H1	0	0	sample10_icdEc	0.0	
96Plate	@10000284	A2	96Plate	@10000282	A2	0	0	sample10_mdhEc	0.0	
96Plate	@10000284	B2	96Plate	@10000282	B2	0	0	sample10_mdhEc	0.0	

Save Output List

Abb. 27: Für manuelle Arbeitsschritte im Labor ist das Erstellen der Zielausgabeliste über die ReSeqLIMS-Weboberfläche erforderlich. In Abschnitt (A) werden Ziellaborstation und -plastikware definiert, Abschnitt (B) zeigt plattenspezifisch die generierten Ausgabelisten, zu denen der Zielbarcode angegeben werden muss.

3.2.3.4 Das Wiederholen von Laborproben

In den bisherigen Beschreibungen eines Laborworkflows wird von einem stringenten Durchlaufen der Arbeitsschritte ausgegangen. Treten innerhalb des Laborworkflows Unregelmäßigkeiten auf oder erreichen die Chromatogramme nicht die erforderliche Qualität, ist das Wiederholen von Arbeitsschritten ausgewählter Laborproben notwendig. Die Wiederholung von Laborproben ist nur möglich, wenn die zu reaktivierende Laborprobe in einem Workflow-Step nicht vollständig verbraucht wurde. Laborproben können im ReSeqLIMS auf Laborproben-, Gen- und Primerebene wiederholt werden, in dem der Status der Laborprobe an der ausgewählten Wiederholerlaborstation auf „Pending samples“ zurückgesetzt wird. Die Laborproben, die zwischen der reaktivierten Laborprobe und der zu wiederholenden Laborprobe liegen, werden verworfen. Falls bereits Chromatogramme generiert wurden kann vom Benutzer festgelegt werden, ob vorliegende Chromatogramme gelöscht werden oder erhalten bleiben sollen. Die Auswahl der Wiederholerebene definiert wie viele Laborproben verworfen und reaktiviert werden. Wird z.B. für die MLST-Methode eine Laborprobe nach erfolgter DNA-Sequenzierung für eine Wiederholung auf Laborprobenebene ausgewählt, werden allein an der Laborstation DNA-Sequenzierung vierzehn Laborproben, optional zusätzlich der Chromatogramme, verworfen. Ist die Wiederholerlaborstation die DNA-Sequenzierung müssten dort dementsprechend vierzehn Laborproben, an der Laborstation der PCR-Produktaufreinigung sieben und an der Laborstation der DNA-Normalisierung nur eine Laborprobe reaktiviert werden. An den nicht methodenspezifischen Laborstationen unterscheidet sich die Auswahl und Bearbeitung der zu wiederholenden Laborproben nicht von den anderen Laborproben. An methodenspezifischen Laborstationen erfolgt die Verarbeitung der zu wiederholenden Laborproben dagegen separat von den anderen Laborproben, da z.B. beim PCR-Setup für die Wiederholerlaborproben nur spezifische Gene zur Wiederholung ausgewählt wurden und damit die Verbrauchsmengen und die Positionen der Reagenzien auf dem Pipettierroboter von der ursprünglichen Methode abweichen. Eine Ausnahme bilden hier die Wiederholerlaborproben auf Laborprobenebene. Da die gesamte Laborprobe und nicht nur spezifische Gene oder Primer zu wiederholen sind, unterscheidet sich die Behandlung der Laborproben nicht vom Ablauf der ursprünglichen Methode.

3.2.3.5 Die Verwendung von Verbrauchsmaterialien

Für die Verbrauchsmaterialien findet im Gegensatz zu den Reagenzien keine Berechnung der Verbrauchsmengen an den Laborstationen statt. Verbrauchsmaterialien werden den Laborstationen zugeordnet, an denen sie verwendet werden können (siehe Kap. 3.1.1.3). Zu den Verbrauchsmaterialien sind anschließend über die ReSeqLIMS-Weboberfläche Chargen anzulegen, die Verbrauchsreihenfolge der Chargen festzulegen und die aktuelle Charge zur Verwendung an den Laborstationen auszuwählen. An den Laborstationen erfolgt ausschließlich die Anzeige der aktuellen und der nachfolgenden Chargen. Die aktuelle Charge eines Verbrauchsmaterials wird automatisch bei der Generierung der Arbeitsliste für jede Laborprobe protokolliert. Ist eine aktuelle Charge aufgebraucht oder werden von einem Verbrauchsmaterial zwei Chargen bei der Bearbeitung der nächsten Arbeitsliste benötigt, ist dies aktiv vom Nutzer vor der Erstellung der Arbeitsliste anzugeben und die nachfolgende Charge zu aktivieren.

3.2.4 Das ReSeqLIMS-Gerätemodul

Im Gerätemodul des ReSeqLIMS werden die Eigenschaften der verwendeten Photometer, Pipettierroboter und DNA-Sequenzierer verwaltet und konfiguriert. Dazu gehört neben der Definition des Layouts von Arbeits- und Ausgabelisten für den verwendeten Pipettierroboter auch die Definition des Layouts der Sequenzierliste. Die gerätespezifischen Parameter beziehen sich u.a. auf die vom Sequenzierer verwendeten Protokolle und die Positionen der Reagenzien auf dem Pipettierroboter.

3.2.5 Die ReSeqLIMS-Statusberechnung

Der Status eines Auftrages, einer Auftragsposition, der Proben und der Laborproben ergibt sich aus dem Status der direkt untergeordneten Auftrags Ebene bzw. der direkt nachfolgenden Proben und Laborproben. Entscheidend ist jeweils der niedrigste Status, der sich aus der Workflow-spezifischen Anordnung der Laborstationen und deren Status für noch nicht bearbeitete und bearbeitete Laborproben ergibt (siehe Kap. 3.1.4.1). Ändert sich der Status einer Laborprobe, muss der Status der vorrangegangenen Laborprobe unter Berücksichtigung der dieser Laborprobe nachfolgenden Laborproben geprüft werden (siehe Abb. 28, 1. - 4.). Eine Überprüfung des Status auf einer nächsthöheren Ebene muss nur bei Statusänderungen auf der aktuellen Ebene durchgeführt werden. Für den Kunden werden nur die Status „Not arrived“, „Arrived“, „In Process“ und „Appro-

ved“ verwendet. Der Mitarbeiter sieht jeweils den exakten Auftrags- bzw. Probenstatus (siehe Anh. B). Durch das Wiederholen von Laborproben ist es möglich, dass der Auftragsstatus auf einen früheren Status zurückspringt. Dem Kunden Statussprünge zu kommunizieren ist aus Gründen der Übersichtlichkeit und Verständlichkeit nicht sinnvoll.

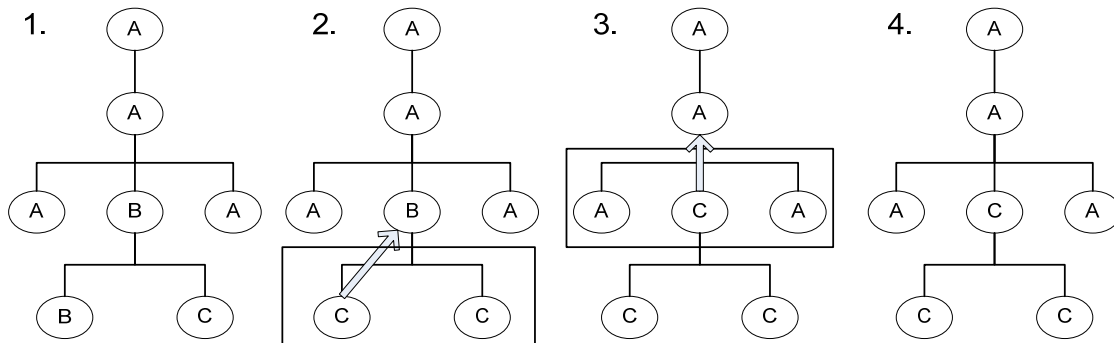


Abb. 28: Berechnung der Statusänderungen bei Laborproben, wenn die Statusreihenfolge A, B und C ist (von niedrig nach hoch). Der zu C veränderte Status in 2. führt auf der nächsthöheren Ebene in 3. ebenfalls zu einer Statusänderung. Diese Statusänderung führt auf der nächsthöheren Ebene in 4. nicht mehr zu einer weiteren Statusveränderung, da auf der darunterliegenden Ebene noch zwei Laborproben mit einem niedrigerer Status vorliegen, die damit den Status auf der nächsthöheren Ebene bestimmen.

3.2.6 Die ReSeqLIMS-Dokumente

3.2.6.1 Auftragstemplates im Excel-Format

Die ReSeqLIMS-Webanwendung stellt Auftragstemplates im Excel-Format zur Verfügung, um den Kunden eine einfache Möglichkeit der Auftragsgenerierung anbieten zu können. Da sich das Layout der Auftragstemplates an der Plastikware orientiert, gibt es Auftragstemplates für Reaktionsgefäße und jeweils für 96er- und 384er-Mikrotiterplatten (siehe Abb. 29 (A) - (C)), die zusätzlich in den drei Versionen Standard, Intern und Raw vorliegen, wovon ausschließlich das Standardformat externen Personen zum Download zur Verfügung steht. Dieses besteht aus zwei Arbeitsblättern. In das erste Arbeitsblatt werden die auftragsrelevanten Daten in Abschnitt (A) und die Probenamen in Abschnitt (B) eingetragen. Das Layout der Tabelle im ersten Arbeitsblatt der Auftragstemplates für die Mikrotiterplatten repräsentiert das originale Plattenlayout mit den entsprechenden Spalten- und Zeilenbeschriftungen. Desweiteren stehen

im unteren Bereich des Arbeitsblattes Funktionen zur Verfügung, die zum Eintragen der Probandaten benutzt werden können, wozu z.B. das Anhängen des Barcodes an eingetragene Probenamen gehört (siehe Abb. 29 (C)). Eine Besonderheit ist der Button „Paste Clipboard Content“. Diese Funktion erlaubt es dem Kunden Daten in einer anderen Excel-Datei zu kopieren und passend in das Plattenlayout einzufügen. Das zweite Arbeitsblatt ist ein nicht-editierbarer Kontrollbogen, in dem die Probandaten des ersten Arbeitsblattes tabellarisch zusammengefasst werden. Die Übertragung der Daten vom ersten in das zweite Arbeitsblatt erfolgt automatisch.

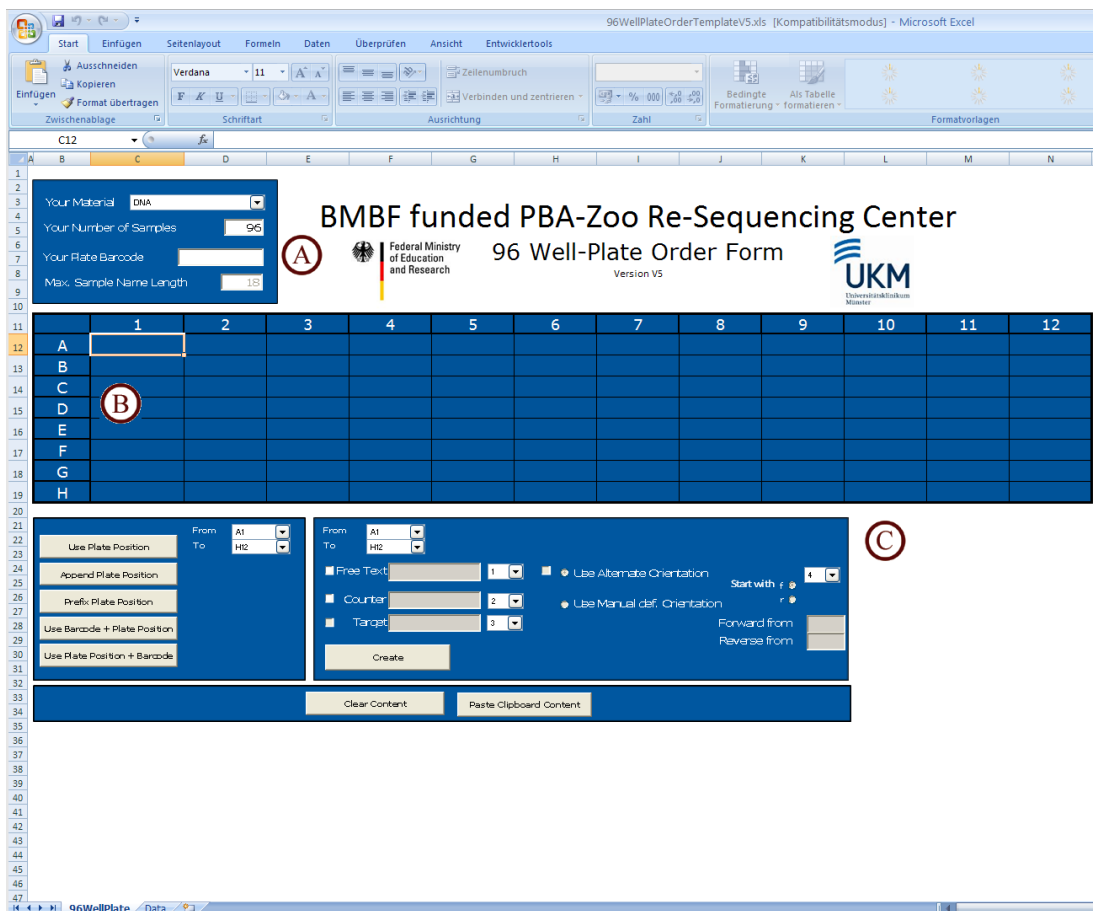


Abb. 29: Auftragstemplates für Probandaten einer 96er Mikrotiterplatte. Abschnitt (A) beinhaltet allgemeine Auftragsinformationen, Abschnitt (B) die Probenamen, die mit Hilfsfunktionen aus Abschnitt (C) erstellt werden können.

Das Interne-Format des Auftragstemplates beinhaltet im zweiten Arbeitsblatt eine zusätzliche, editierbare Spalte zum Eintragen von Probenkonzentrationen. Im Gegensatz zu dem Standard- und Internen-Format besteht das Raw-Format des Auftragstemplates ausschließlich aus einem Arbeitsblatt, vergleichbar mit den zweiten Arbeitsblättern der

anderen Formate, mit dem Unterschied, dass die Probanddaten durch Kopieren aus anderen Dateien direkt eingefügt werden können. Unterstützende Funktionen zum Generieren des Probennamens sind hier nicht vorhanden. Das Raw-Format ermöglicht das einfache Kopieren von bereits vorliegenden spaltenorientierten Probanddaten in einer anderen Excel-Datei. Durch Copy&Paste ist die Verwendung des Raw-Formats für Mitarbeiter häufig die schnellste Form, um ihre Probanddaten auftragsgerecht aufzubereiten. Die Programmierung der Funktionalität der Auftragstemplates erfolgte in Visual Basic for Applications (VBA). Zur Vermeidung von Fehleingaben und Veränderungen interner Formeln der Auftragstemplates sind die Zellen der Arbeitsblätter bis auf die Bereiche, in die Daten eingetragen werden müssen, gesperrt. Im Raw-Format wurde auf das Sperren der Zellen verzichtet, um damit das problemlose Copy&Paste zu ermöglichen. Zum Einlesen eines Auftrages in das ReSeqLIMS muss stets die aktuellste Version der Auftragstemplates verwendet werden, da die Versionsnummer aus dem ersten Arbeitsblatt beim Einlesen des Auftragstemplates systemseitig abgefragt wird. Neben der Version wird ein weiteres, verstecktes Arbeitsblatfeld abgefragt, das einen Schlüssel enthält, um die Echtheit des ReSeqLIMS-Auftragstemplates zu verifizieren.

3.2.6.2 Arbeits- und Ausgabelisten des Pipettierroboters

Die Arbeits- und Ausgabelisten des Pipettierroboters sind im Excel-Format definiert und enthalten alle notwendigen Informationen zur lückenlosen Protokollierung der Laborproben in einem Workflow (siehe Abb. 30 (A) - (B)). Das Arbeiten mit Arbeits- und Ausgabelisten ist notwendig, da die aufeinander folgenden Arbeitsschritte eines Workflows wie z.B. die DNA-Normalisierung und das PCR-Setup als eigenständige Module des Pipettierroboters implementiert wurden. Das hat den Vorteil, dass die Module in mehreren Workflows verwendet werden können. Der Nachteil unabhängiger Module ist der fehlende Informationsaustausch zwischen den aufeinander folgenden Arbeitsschritten. Die Reihenfolge der Arbeitsschritte und der Informationsfluss werden daher erst durch die Workflow-Steps und die Arbeits- und Ausgabelisten im ReSeqLIMS definiert, wodurch zusätzlich sichergestellt wird, dass die Arbeits- und Ausgabelisten nur in der Workflow-spezifischen Reihenfolge generiert und eingelesen werden können. Durch die offene Implementierung der Module des Pipettierroboters waren umfassende Anpassungen der Arbeits- und Ausgabelisten an die Erfordernisse der Labormitarbeiter, der Qualitätsstandards und Laborprozesse möglich. Die Zuordnung von

Informationen aus den Excel-Dateien erfolgt durch Parsen der Spaltenüberschriften. Während die Reihenfolge der Spalten für den Pipettierroboter nicht von Bedeutung ist und nur die Spaltenüberschriften exakt übereinstimmen müssen, sind sowohl die exakten Spaltenüberschriften als auch die Reihenfolge der Spalten für das ReSeqLIMS strikt einzuhalten. Zusätzlich gilt, dass für jede Arbeits- und Ausgabeliste die eindeutige Identifikation der Laborproben zu jeder Zeit gewährleistet sein muss.

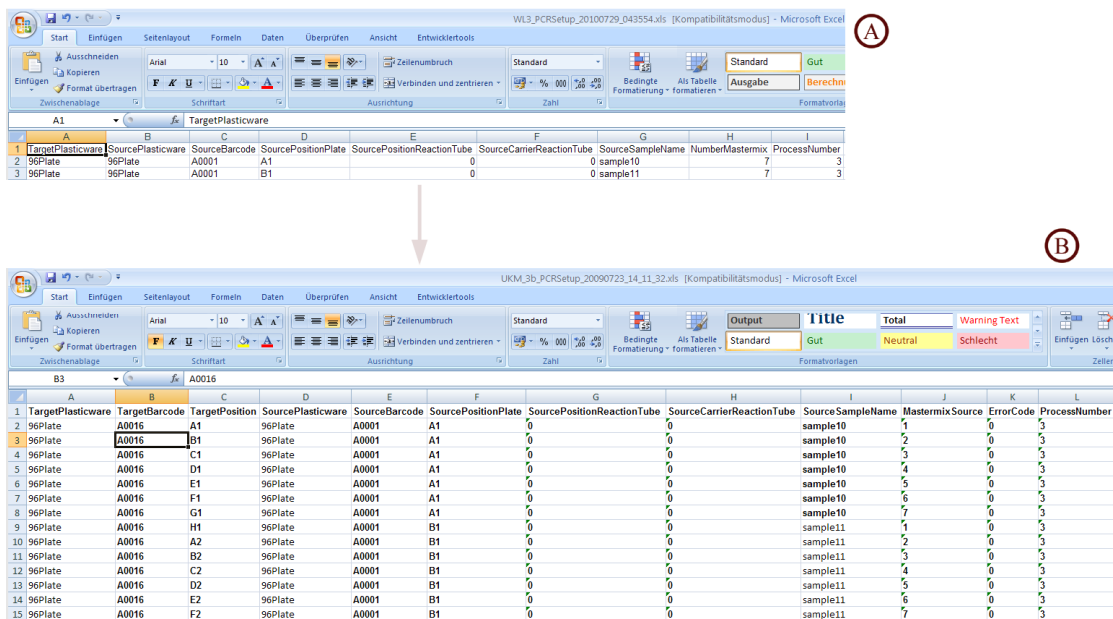


Abb. 30: Beispiel einer Arbeits- (A) und der zugehörigen Ausgabeliste (B) an der PCR-Setup Laborstation.


Wie in Kap. 3.2.3.1 beschrieben, wird dazu der in den Arbeits- und Ausgabelisten enthaltene Barcode, im Falle der Mikrotiterplatte zusätzlich die Plattenposition, und der Workflow-Step verwendet, der in der Arbeitsliste repräsentiert wird durch die ProcessNumber, die während des Hochladens der Arbeitsliste am Pipettierroboter mit dem ausgewählten Modul bzw. dessen Modulnummer abgeglichen wird. Stimmen die beiden Werte nicht überein, ist das Starten eines Moduls am Pipettierroboter nicht möglich. Die Arbeitsliste (siehe Abb. 30 (A)) wird während der Modulausführung um den Zielbarcode, die Zielplattenpositionen der Mikrotiterplatte und laborstationsspezifische Parameter für jede Laborprobe ergänzt und als Ausgabeliste (siehe Abb. 30 (B)) im Ausgabebereich der Laborstation des jeweiligen Moduls abgespeichert. Beim Einlesen der Ausgabeliste des Pipettierroboters in das ReSeqLIMS werden die Ausgangsbarcodes

und Ausgangsplattenpositionen der Mikrotiterplatte oder die Ausgangsbarcodes der Reaktionsgefäße dazu verwendet, die Ausgangs- und Zielproben miteinander zu verketten (siehe Kap. 3.2.3.1). Das Anfügen von Gen- oder Primernamen an die Laborprobenamen erfolgt durch das ReSeqLIMS erst beim Einlesen der entsprechenden Ausgabelisten von PCR- oder Sequenzierungssetup. Dazu wird auf Spalteninformationen der Ausgabelisten zurückgegriffen, denen entnommen werden kann, welcher Laborprobe welcher Mastermix oder welcher Primer hinzu pipettiert wurde.

3.2.6.3 ReSeqLIMS-generierte Dokumente im PDF-Format

Die Rechnung des ReSeqLIMS wird im PDF-Format erstellt. Im oberen Bereich der Rechnung können die Absenderadresse (rechts) und der Adressbereich des Empfängers (links) differenziert werden (siehe Abb. 31 (A)). Im Auftragsbereich einer Rechnung werden die Aufträge durch ihre Auftrags-ID eindeutig identifiziert (siehe Abb. 31 (C)). Der Auftrags-ID nachstehend befinden sich die zu einem Auftrag gehörenden Auftragspositionen mit den Auftragspositionsparametern (siehe Abb. 31 (D)). Die tabellarische Abrechnung der Auftragspositionen erfolgt Workflow-spezifisch für jede in Anspruch genommene Laborstation (siehe Abb. 31 (E)). Handelt es sich um eine Sammelrechnung, befinden sich im Auftragsbereich mehrere Auftrags-IDs inklusive der dazugehörigen Auftragspositionen und der tabellarischen Abrechnung blockweise untereinander, die tabellarische Abrechnung allerdings ohne eine Addition der Mehrwertsteuer. In Sammelrechnungen werden die Nettobeträge der beinhaltenden, einzelnen Aufträge bzw. Auftragspositionen nach der blockweisen Auflistung der selbigen tabellarisch aufsummiert und anschließend die Mehrwertsteuer zu der Gesamt-Nettosumme hinzuge-rechnet. Die Rechnung wird abgeschlossen durch die Fußzeile mit der Bankverbindung und dem Zahlungsziel (siehe Abb. 31 (F)). Neben der Rechnung wird das Anschreiben für den Kunden zur Versendung von Barcodeaufkleber ebenfalls im PDF-Format erzeugt. Während der Adressbereich des Barcodeaufkleber-Anschreibens mit dem der ReSeqLIMS-Rechnung vergleichbar ist, mit dem Unterschied, dass keine Rechnungsspezifische Daten aufgeführt werden, ist der Auftrags- und Fußbereich der Rechnung durch ein kurzes Anschreiben ausgetauscht (siehe Abb. 31 (G)).

(A)



Universitätsklinikum Münster, Institut für Hygiene
48149 Münster

Universitätshilfsklinik Münster
Institut für Hygiene
Robert-Koch-Straße 41
48149 Münster

Institut für Hygiene
Ina Fichtner
Robert-Koch-Straße 41
D 48149 Münster
Tel.: +49 (0)251 - 83-55052
Fax: +49 (0)251 - 83-55044
Id. Vernetzung: 49 701251 83 0
ina.fichtner@ukmuenster.de
www.klinikum.uni-muenster.de
Münster, 05.05.2010
UKM-Steuer Nummer: 336/5228/0014
Int.-Id.-Nr.: DE212/277/72

Rechnung der Servicestelle Hygiene:
(Projekt PBA-Zoo)

Rechnungsnummer: PBA-2010-0003

Für Frau Olivia Marital
Universitätsklinikum Münster, Institut für Hygiene, Robert-Koch-Straße 41, 48149 Münster
o.o.rules@uni-muenster.de, Tel.: 00492518355052
wurden folgende Leistungen erbracht:

Zu Auftrag Nr. 10908 vom 27.04.2010

Option: in solution	Methoden:	Kondition: BMBF Amplicon Re-Sequencing
1. Auftragsposition: 1	Anzahl Sample(s): 96	Material: Purified sequencing product
		Plastikware: Well Plate (96)

Rechnungsposten	Anzahl	Einzelpreis(in EUR)	Gesamtpreis(in EUR)
Sequencing	96	1,37	132,72
		Summe Netto	132,72
		zzgl. MwSt (19,00%)	31,68
		Endsumme	164,40


- Bei Rückfragen wenden Sie sich bitte an: 025183-55052 -

Wir bitten um die Überweisung des Betrages bis zum 20.05.2010 auf das UKM-Konto bei der Westfälischen Landesbank

Westdeutsche Landesbank
Kontonummer: 072527
BIC: 2410 2510 00
Verwendungszweck: PBA-2010-0003

(C) **(D)** **(E)** **(F)**

(B)



Universitätsklinikum Münster, Institut für Hygiene
48149 Münster

Institut für Hygiene
Robert-Koch-Straße 41
D-48149 Münster
sequenci@ukmuenster.de
www.klinikum.uni-muenster.de
Münster, 19.05.2010
UKM-Steuer Nummer: 336/5228/0014
Ust.-Id.-Nr.: DE212/277/72

Sehr geehrter Herr Ginzol,

anbei finden Sie neue Barcode Aufkleber, die bei Bestellungen am PBA-Zoo Re-Sequencing Center verwendet werden können.

Wir bedanken uns, dass Sie sich für die Nutzung des PBA-Zoo Re-Sequencing Center entschieden haben!

Mit freundlichen Grüßen,
The PBA Zoo Re-Sequencing Center Team

(G)

Abb. 31: PDF-Dokumente des ReSeqLIMS. (A), (C) - (F) zeigt eine Rechnung, (B) und (G) das Barcodeaufkleber-Anschreiben.

3.2.6.4 Die Sequenzierliste

An der Laborstation DNA-Sequenzierung wird keine Arbeitsliste, sondern eine Sequenzierliste erzeugt. Während die Arbeits- und Ausgabelisten im Excel-Format generiert werden, wird die Sequenzierliste für den bzw. die DNA-Sequenzierer als TXT-Datei abgespeichert. Das Format der TXT-Datei orientiert sich jeweils an den spezifischen Anforderungen der im Labor eingesetzten DNA-Sequenzierer. Nach der Auswahl der Sequenzier-Mikrotiterplatte und des zu verwendenden DNA-Sequenzierers werden zunächst die Sequenzierprioritäten und das zu verwendende Sequenzer-spezifische Sequenzierprotokoll der einzelnen Laborproben festgelegt. Die Spalten der anschließend generierten Sequenzierliste sind z.B. für den zurzeit im Labor verwendeten DNA-Sequenzierer durch das TAB-Zeichen getrennt (siehe Abb. 32). Der Container-Name entspricht dem Barcode der zu sequenzierenden Mikrotiterplatte, der Container-Type dem Plattentyp (z.B. 96er oder 384er Mikrotiterplatte).

Container Name	Description	ContainerType	AppType	Owner	Operator	Well	Sample Name	Comment	Priority	Results	Group 1	Instrument	Protocol 1	Analysis Protocol 1
A0477	96-Well	Regular	IFH	IFH										
AppServer AppInstance														
SequencingAnalysis														
A1	E07_207_EC0002_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
B1	E07_207_EC0076_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
C1	E07_207_EC0093_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
D1	E07_207_EC0100_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
E1	E07_207_EC0204_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
F1	E07_207_EC0268_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
G1	E07_207_EC0406_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
H1	E07_207_EC0444_f	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
A2	E07_207_EC0002_r	100	IFH	Seq_700bp_Z	Seq_700bp_Z									
B2	E07_207_EC0076_r	100	IFH	Seq_700bp_Z	Seq_700bp_Z									

Abb. 32: Beispiel einer ReSeqLIMS-Sequenzierliste. Das Format ist spezifisch an den vom Labor verwendeten DNA-Sequenzierer angepasst.

4 Diskussion

Im Rahmen des PBA-Zoo Projektes wurde mit dem ReSeqLIMS ein LIMS speziell für die Verwendung in Amplikon-Resequenzierungsprojekten entwickelt, um deren spezifischen Anforderungen zu entsprechen. Nach einer detaillierten Analyse des Anwendungsbereiches wurde zunächst das ReSeqLIMS-ERM erstellt, das als Grundlage für die Struktur der SQL-Datenbank verwendet wurde. Basierend auf der ReSeqLIMS-Datenbank wurde anschließend das ReSeqLIMS als Webanwendung bestehend aus Kunden-, Labor- und Gerätemodul implementiert. Das ReSeqLIMS integriert die am Laborworkflow beteiligten Personen, Laborprozesse und verwendeten Geräte optimal in einen qualitätsüberwachten Workflow. Dadurch ist es möglich, die Proben- und Qualitätsdaten effektiv und effizient während der Laborprozesse im Hochdurchsatz zu steuern und zu verwalten. Die webbasierte Client-Server-Architektur macht das ReSeqLIMS durch die Verwendung eines Standard-Webrowsers plattformunabhängig einsetzbar. Die Weboberfläche bietet zudem ein intuitives Benutzerinterface für Kunden und Mitarbeiter, um zu jedem Zeitpunkt Zugriff auf die aktuellen Auftrags- und Laborprozessinformationen zu erhalten. Die Resequenzierungsergebnisse sind durch die vom ReSeqLIMS während des Laborworkflows vorgenommene Erweiterung der kundendefinierten Probenamen um Gen - und Primernamen sehr gut für nachfolgende, automatisierte Downstream-Analysen geeignet.

Mit dem ERM von Chen wurde für das ReSeqLIMS-Datenmodell (siehe Kap. 3.1) eine Entwurfsmethode für den fachkonzeptionellen Entwurf gewählt, die den Datenbankentwurf sehr gut unterstützt, und die Datenintegrität und -konsistenz durch die Transformationsregeln der Überführung des ERMs in das DV-Konzept auch im relationalen Datenbankmodell und der Implementierung sicherstellt. Die Datenintegrität betrifft die logische Korrektheit, Gültigkeit und Genauigkeit der Daten in Datenbanken [8]. Nur die für eine Anwendung relevanten Daten sind in das Datenmodell zu über-

nehmen, die Beziehungen und die definierten Wertebereiche der Daten müssen korrekt sein. Das redundante Speichern von Daten in der Datenbank ist zu vermeiden. Die Datenkonsistenz wird auch als Transaktionskonsistenz bezeichnet. Konsistente Daten sind in sich stimmig und widerspruchsfrei [8]. Veränderungen an den Daten dürfen nicht zu ungültigen Datenbankzuständen führen. Im ERM definieren Entities die anwendungsrelevanten Daten und Relationships die Beziehungen der Daten. Durch die Vergabe von Primärschlüsseln ist die eindeutige Identifizierung der Entities und damit die Entityintegrität sichergestellt. Die referentielle Integrität im Datenbankmodell wird durch die Transformationsregeln des Fach- in das DV-Konzept (siehe Kap. 2.2.2, Abb.5) eingehalten, da nur Primärschlüssel von Tabellen zu Fremdschlüsseln in anderen Tabellen werden können. Das ReSeqLIMS-ERM ist ein grundlegendes Dokument zur Dokumentation der Datenbasis, die unmittelbar mit der Leistungsfähigkeit und der Implementierung des ReSeqLIMS in Beziehung zu setzen ist. Sind im ReSeqLIMS-ERM einer Person maximal zwei Adressen zugeordnet (siehe Kap. 3.1.2.1), ist eine Anpassung des ERMs notwendig, wenn diese Beschränkung für das ReSeqLIMS geändert werden soll. Während ein ERM durch die leichte Lesbarkeit gut anzupassen und zu erweitern ist, wirken sich Änderungen des ERMs immer auf die Datenbank und auf die Implementierung des ReSeqLIMS in z.T. nicht unerheblicher Art und Weise aus.

Im Rahmen des PBA-Zoo Projektes musste zunächst festgestellt werden, ob evtl. eine existierende LIMS-Lösung zur Verfügung steht, die den spezifischen Anforderungen des PBA-Zoo Projektes entspricht. Dazu wurden die kommerziellen LIMS-Softwareprodukte GeneSifter von Geospiza [26], Geneus von GenoLogics [25], Simatic IT Unilab von Siemens [60], Nautilus und SampleManager von Thermo Fischer Scientific [65] und STARLIMS von STARLIMS [62] und als Vertreter von Open-Source-Softwareprodukten die LIMS-Lösungen T.I.M.S. [42], PACLIMS [19], SLIMS [67], 2D-LIMS [44], MMP-LIMS [54], die Variant Identification Pipeline [18] und je eine LIMS-Lösung für das Mutationsscreening von Kandidatengenomen [69] und für die Genotypisierung von Getreide [34] analysiert (siehe Anh. D, Tab. 1). Voraussetzung zur Aufnahme in die Analyse war eine vergleichbare Anwendungsumgebung oder ein vergleichbares Softwarekonzept, das den Anforderungen an das ReSeqLIMS entsprach. Für eine mögliche LIMS-Eigenentwicklung wurden bei der Analyse der genannten Produkte der Aufbau, die Funktionalität, die Zielgruppen und die Implementierungsstrategie genauer betrachtet (siehe Anh. D, Tab. 2). Die kommerzielle Software ist gekenn-

zeichnet durch einen modularen Aufbau und eine hohe Konfigurierbarkeit. Dadurch überdecken die Softwareprodukte ein breites Anwendungsspektrum. Während STARLIMS mit der Probenverwaltung nur einen geringen Teil der notwendigen Funktionalität abdeckt, entsprechen die anderen kommerziellen Softwareprodukte hinsichtlich der Funktionalität den Anforderungen an eine Hochdurchsatz-Resequenzierungsumgebung, inkl. einer Workflow-orientierter Prozessabbildung, einer Probenverfolgung, einer Qualitätssicherung, der verwendeten Client-Server-Architektur und der Implementierung einer webbasierter Benutzerschnittstelle. Die Funktionalitäten übersteigen die tatsächlichen Anforderungen des ReSeqLIMS sogar bei weitem. Alle kommerziellen Softwareprodukte umfassen z.B. umfangreiche Analyse- und Berichtsmodule, die im ReSeqLIMS vorerst nicht vorgesehen sind. Die Geräte und Plastikwaren der großen Anbieter sind bereits in großer Anzahl vordefiniert. Neben den initialen Anschaffungskosten und den Kosten für Customizing und Restrukturierungsmaßnahmen, müssen im Rahmen der Einführung von kommerziellen LIMS-Produkten neben den laufenden Kosten in Form von Lizenzierungskosten oder Kosten für Wartungsverträge auch Kosten für Mitarbeiterschulungen berücksichtigt werden. Nachträgliche Änderungen an der LIMS-Software zur Verbesserung der Benutzerführung oder Anpassungen an sich verändernde Laborprozesse, bedürfen der kostenpflichtigen Entwicklungsarbeit seitens der Hersteller, falls diese angeboten wird. Bezüglich der Nutzung, Weiterentwicklung und Anpassung kommerzieller LIMS besteht eine Abhängigkeit vom Hersteller. Der Vorteil von kommerziellen Softwareprodukten ist die relativ kurzfristige Verfügbarkeit und Nutzbarkeit der Software, wenn nach einem komplexen Auswahlverfahren eine geeignete LIMS-Software aus der Vielzahl der Anbieter bestimmt wurde. Die betrachtete Open-Source-Software ist von Laboren für ein spezifisches Projekt entwickelt worden, weil die verfügbare kommerzielle Software überdimensioniert ist, zu teuer ist oder den spezifischen Anforderungen nicht ausreichend angepasst werden kann [4,44,67,69]. Sind für einen Anwendungsbereich keine kommerziellen LIMS-Softwareprodukte vorhanden, ist die Eigenentwicklung häufig die einzige Möglichkeit einer adäquaten LIMS-Lösung [4]. Eine Eigenentwicklung kann durch die Herstellerunabhängigkeit einfacher und schneller an sich ändernde Prozesse und Benutzerwünsche angepasst werden, wenn die dazu benötigte Arbeitskraft aufgebracht und investiert wird. Die Kosten der Entwicklung von Individualsoftware übersteigen jedoch nicht selten die Kosten von kommerziellen Softwareprodukten, insbesondere auch unter Berücksichti-

gung der nach der Einführung der Eigenentwicklung entstehenden Kosten, da z.B. für die Fehlerbeseitigung keine Updates bezogen werden können, sondern interne Arbeitskraft gebunden wird. Durch die gezielte Entwicklung für spezifische Projekte sind die maßgeschneiderten LIMS-Lösungen nur eingeschränkt konfigurierbar und umfassen ausschließlich Module, die exakt an die im jeweiligen Anwendungsfall gegebene Laborinfrastruktur angepasst wurden. Erweiterungen, Anpassungen und der Einsatz in anderen Anwendungsumgebungen sind mit Implementierungsaufwand verbunden, der durch die teilweise angewendete Modularisierung der Software erleichtert werden soll.

Die Analyse der kommerziellen und Open-Source-Software ergab, dass kein Softwareprodukt für die spezifischen Anforderungen und den kalkuliertem Kostenrahmen verfügbar war. So unterschied sich beispielweise die für den Hochdurchsatz geeignete Open-Source-Software z.T. erheblich in ihren Laborworkflows von dem Laborworkflow des Instituts für Hygiene des UKMs. Insbesondere die Arbeitslistengenerierung des Pipettierroboters hätte in allen Fällen eine nachträgliche Implementierung und Integration erforderlich gemacht. Eine kommerzielle Software wurde aufgrund der erforderlichen Anschaffungs-, der Anpassungs- und der laufenden Kosten ausgeschlossen. Nach der Analyse der ausgewählten kommerziellen und Open-Source-Software, wurde das ReSeqLIMS daher als Eigenentwicklung realisiert, um es exakt an die optimierten und standardisierten Prozesse und an die gegebene Infrastruktur des Instituts für Hygiene in Münster anpassen zu können.

Für das ReSeqLIMS stand die Implementierung als Web- oder als Desktopanwendung zur Auswahl. Da Webanwendungen von den Clients plattformunabhängig mit einem Webbrowser als Front-End verwendet werden können, wurde dieser Weg für die Implementierung des ReSeqLIMS gewählt, da dadurch eine lokale Installationen von Software, für die evtl. administrative Rechte erforderlich wäre, entfallen würde. Außerdem ist damit ist die Nutzung der gleichen und aktuellsten Version der Webanwendung durch alle Clients garantiert sowie eine zentral oder dezentral organisierte Datenhaltung durch die Webanwendung möglich, die den zu jeder Zeit realisierbaren Zugriff auf die aktuellste, konsistente Datenbasis garantiert. Die Performance und Erreichbarkeit der Webanwendung ist durch die Nutzung des Internets stark abhängig von der Bandbreite und Ausfallwahrscheinlichkeit, die dem Client von seinem Netzanbieter zur Verfügung gestellt wird, was jedoch heutzutage aufgrund der hohen Internetbandbreiten kein großes Problem mehr darstellt.

Das ReSeqLIMS-Kundenmodul integriert die Personen bzw. Anwender in das ReSeqLIMS und automatisiert und optimiert Prozesse wie die Registrierung neuer Anwender, den Zugriff auf den Status und die Ergebnisse von Aufträgen sowie das Rechnungswesen. Da die Nutzung der ReSeqLIMS-Webanwendung für Kunden und Mitarbeiter über eine Weboberfläche mit einem Standard-Webbrowser möglich ist, ist höchste Flexibilität bzgl. der Zugriffsmöglichkeiten unabhängig von Betriebssystem und Webbrowser gegeben. Anbieter von Sequenzierungsleistungen wie Eurofins [23], LGC Genomics [39], GATC [24] und Beckman Coulter Genomics [9] sind ebenfalls im Internet mit einem Webauftritt vertreten. Vergleichbar mit den genannten Anbietern von Sequenzierungsleistungen differenziert die ReSeqLIMS-Webanwendung einen öffentlichen Bereich, in dem Informationen zu angebotenen Serviceleistungen, Verfahrensanleitungen für Bestellungen, Anforderungen an die einzusendenden Proben und evtl. vorhandene Auftragstemplates, zur Verfügung gestellt werden, und einen nicht-öffentlichen Bereich. Für die Generierung von Aufträgen, den Zugriff auf den Auftragsstatus und die Ergebnisse ist bei allen Anbietern eine Registrierung erforderlich. Im Gegensatz zu anderen Anbietern erfolgt im ReSeqLIMS eine Benutzerkontenaktivierung durch Mitarbeiter des Amplikon Re-Sequencing Centers. Die manuelle Benutzerkontenaktivierung ermöglicht sowohl eine bessere Kontrolle der Registrierungen als auch die erforderliche Zuordnung einer Organisation und einer Nutzergruppe zu den neu registrierten Anwendern. Je nach zugeordneter Nutzergruppe erhält der Anwender nur Zugriff auf die Daten und Funktionen, die zur Bearbeitung seiner Aufgaben erforderlich sind. Zur Standardisierung des Bestellprozesses ist die Generierung von Aufträgen ausschließlich mit den auf der ReSeqLIMS-Weboberfläche bereitgestellten Auftragstemplates möglich. Auf Webformulare zur Eingabe von Probanddaten, wie sie von einigen der genannten Anbieter zur Verfügung gestellt werden, wurde verzichtet, da die erforderliche Online-Auftragsgenerierung mit den Auftragstemplates der komfortabelste Weg ist auch große Datenmengen einer z.B. voll belegten 384er Mikrotiterplatte in das ReSeqLIMS zu laden. Die Verantwortung der Korrektheit der Auftragsdaten liegt beim Kunden. Die Rechnungsgenerierung für freigegebene Aufträge kann mit dem ReSeqLIMS zeitnah durchgeführt und das Rechnungswesen durch die Verfolgung des Rechnungsstatus erleichtert und beschleunigt werden. Systemgenerierte E-Mails geben den Kunden Feedback bzgl. Registrierung, Auftragsstatus und Ergebnisverfügbarkeit. Einige Anbieter verwenden zur Anzeige der Aufträge bzw. der Probanddaten anstelle einer tabellarischen

Anzeige, wie sie zurzeit im ReSeqLIMS verwendet wird, eine layoutbezogene grafische Anzeige der Plastikware (z.B. [25,26]). Hierdurch ist der Inhalt der Mikrotiterplatte intuitiv abrufbar. Die Anzeige der Plastikwaren-bezogenen Proben in einer an der Plastikware orientierten, grafischen Anzeige ist auch für das ReSeqLIMS zu erwägen und in folgenden Versionen des ReSeqLIMS evtl. zu übernehmen. Während die Aufträge zurzeit nach der Auftragsfreigabe weiterhin im ReSeqLIMS in der Kategorie der freigegebenen Aufträge über den Menüpunkt „Order Overview“ einsehbar sind (siehe Abb. 21), ist in den nächsten Versionen des ReSeqLIMS geplant, freigegebene Aufträge nach einem bestimmten Zeitraum aus der Kategorie der freigegebenen Aufträge durch das Setzen des Auftrages, der Auftragspositionen sowie der Proben und Laborproben auf den neuen Status „Archived“ zu entfernen. Die Aufträge würden dadurch zur Dokumentation weiter erhalten bleiben, gleichzeitig die Ansicht der freigegebenen Aufträge aber deutlich übersichtlicher gestaltet werden. Über eine Suchfunktion wären die Aufträge weiterhin über die ReSeqLIMS-Suchfunktion für Mitarbeiter abrufbar. Für die Chromatogramme der freigegebenen Aufträge gilt dann, dass der Download-Link (siehe Abb. 21 (E)) der freigegebenen Aufträge für drei Monate zur Verfügung steht. Über das ReSeqLIMS können die nach den drei Monaten nicht mehr verlinkten gezippten Ergebnisdateien mit den Chromatogrammen in Backupordnern organisiert werden. Die Anzahl der Chromatogramme in einem Backupordner wird sich nach einer über die ReSeqLIMS-Weboberfläche definierbaren Megabyte (MB) oder Gigabyte (GB)-Größe richten, so dass die Backupordner anschließend auf CD oder DVD archiviert werden können. Die Backupordner bzw. CDs oder DVDs erhalten einen eindeutigen Schlüssel, der den Chromatogrammen der archivierten und gezippten Ergebnisdateien zugeordnet wird. Damit lässt sich die CD/DVD auf der sich spezifische Chromatogramme befinden über das ReSeqLIMS abfragen.

Aufgabe des Labormoduls ist die optimale und effiziente Unterstützung der Prozesse in einem Labor sowie die Integration der Laborprozesse, der beteiligten Personen und der verwendeten Geräte in einen qualitätsüberwachten Laborworkflow. Das ReSeqLIMS-Labormodul definiert durch das Konstrukt des Workflows die Abfolge der Laborstationen zur Bearbeitung von Methoden. Photometer, Pipettierroboter und DNA-Sequenzierer sind durch die Generierung und das Einlesen der Arbeits-/Sequenzier- und Ausgabelisten in einen Laborworkflow integriert. Mitarbeiter haben zu jeder Zeit Zugriff auf die aktuellsten Labordaten, das Generieren der Arbeits- und Ausgabelisten ist

zeitsparend und fehlerfrei möglich. Die gezielte Auswahl von Arbeitslisten zur weiteren Verarbeitung sorgt für Prozesssicherheit bei der Verbindung des ReSeqLIMS und des Pipettierroboters. Eingebaute Sicherheitsmechanismen wie z.B. die ProcessNumber in den Arbeitslisten verhindern die unsachgemäße Bedienung des Pipettierroboters. Plausibilitätsprüfungen, u.a. bei der Auswahl der Plastikware an einer Laborstation, unterstützen die Mitarbeiter bei der Bearbeitung von Aufträgen. Zusätzlich ermöglicht die eindeutige Identifizierung der Laborproben durch Verwendung von Barcodes, Plattenposition und Workflow-Step die für die Amplikon-Resequenzierung zwingend erforderliche Probenverfolgung auf jeder Stufe über den gesamten Laborworkflow. Das für Resequenzierungsprojekte maßgeschneiderte ReSeqLIMS integriert Laborworkflow, Geräte und Mitarbeiter, minimiert Fehlerquellen, wie sie z.B. bei dem manuellen Erstellen von Arbeitslisten auftreten könnten, erzeugt Zeitvorteile und steuert den Proben- und Datenfluss während der verschiedenen Laborprozesse im Hochdurchsatz. In den folgenden ReSeqLIMS-Versionen wird das Labormodul um eine Auftrags- bzw. laborproben-spezifische Laborprobenübersicht ergänzt. In dieser ist es nach der Selektion einer Laborprobe möglich, sich den Verlauf sämtlicher Vorgänger- und Nachfolgerlaborproben mit den Verbrauchs- und Qualitätsinformationen anzeigen zu lassen. Während die bisherige Implementierung nur die Laborstations-spezifische Ansicht der Laborproben über das Workflow-Fenster des „List Managements“ ermöglicht, ist damit die Ansicht der Proben bzw. Laborproben eines Auftrages über alle Workflow-Steps mit den relevanten Informationen in einer Ansicht möglich. Das automatisierte Starten von Modulen des Pipettierroboters durch das Hochladen der Arbeitsliste in den Arbeitsordner des Pipettierroboters und das automatisierte Einlesen der Ausgabeliste in das ReSeqLIMS, sobald der Pipettierroboter die Ausgabedatei in den entsprechenden Laborstation-spezifischen Ordner gespeichert hat, ist zunächst nicht vorgesehen. Unter dem Aspekt der weiteren Automatisierung des Laborworkflows wäre die direkte Kommunikation zwischen LIMS und Pipettierroboter, soweit dies möglich ist, zu implementieren. Die in Kap. 3.2.3.5 beschriebene Vorgehensweise bei der Verwendung der Verbrauchsmaterialien ist z.Z. noch nicht vollständig implementiert.

Das Gerätemodul ist verantwortlich für die Integration der Geräte eines Labores in das ReSeqLIMS. Es werden gerätespezifische Eigenschaften wie z.B. das Format der Arbeits- und Ausgabeliste des Pipettierroboters, die Anzahl der Kapillaren, das Format der Sequenzierliste und die verfügbaren Sequenzierprotokolle eines DNA-Sequenzierers

verwaltet. In kommerziellen Softwareprodukten steht eine Vielzahl von Geräten als „Default“-Geräte zur Verfügung. Eine derartige Implementierung einer Vielzahl von Geräten ist für das ReSeqLIMS nicht sinnvoll. Vielmehr muss der Schwerpunkt des ReSeqLIMS darauf liegen, das Hinzufügen von Geräten über die ReSeqLIMS-Weboberfläche zu unterstützen. Während das Anlegen neuer Geräte über die ReSeqLIMS-Weboberfläche mit der Definition gerätespezifischer Attribute möglich ist, ist das Anlegen bzw. Definieren von Formaten für die Arbeits- und Ausgabelisten zurzeit noch nicht möglich. Das Format von Arbeits- und Ausgabelisten ist fest im Quellcode definiert. Das Hinzufügen neuer Geräte mit anderen Arbeits- und Ausgabelistenformaten ist nur durch Implementierung der neuen Formate im Quellcode möglich. Eine Funktion zur einfachen Definition des Formates von Arbeits- und Ausgabelisten über die ReSeqLIMS-Weboberfläche oder die Verwendung von XML-Dateien, die das Format der Arbeits- und Ausgabelisten definieren, sollte in den folgenden ReSeqLIMS-Versionen angestrebt werden. Das gleiche gilt ebenso für die Sequenzierliste des DNA-Sequenzierers.

Im ReSeqLIMS ist die Protokollierung und laborprobenspezifische Zuordnung von Probenwerten sowie der in einem Arbeitsschritt verwendeten Verbrauchsmaterialien und Geräte durch die eindeutige Identifikation einer Laborprobe über sämtliche Workflow-Steps einer Methode möglich. In diesem qualitätsüberwachten und integrierten Workflow werden auftretende Qualitätsmängel schnell erkannt und können gezielt zurückverfolgt werden. Im Falle auftretender Qualitätsbeeinträchtigungen oder Verfahrensfehler bietet die Wiederholerfunktion des ReSeqLIMS die Möglichkeit Laborproben evtl. auch mehrmals zu wiederholen (siehe Kap. 3.2.3.4). Das Ziel des ReSeqLIMS ist es, durch eine effektive Protokollierung von Qualitätsdaten dem Anspruch der höchsten Ergebnisqualität gerecht zu werden.

Die Erweiterbarkeit und Konfigurierbarkeit ist ein Schwerpunkt bei der Implementierung des ReSeqLIMS. Neue Methoden, Materialien, Optionen, aber auch neue Laborstationen können über die ReSeqLIMS-Weboberfläche angelegt werden. Durch die Kombination von Laborstationspaaren können beliebige Laborworkflows definiert werden. Die bereits implementierten Laborworkflows sind stark orientiert an der modularen Arbeitsweise des Pipettierroboters und der Generierung von Arbeits- bzw. Sequenzierlisten sowie dem Einlesen von Ausgabelisten bzw. Sequenzierungsergebnissen an den einzelnen Laborstationen. Neue Methoden oder Geräte, die von dem bisherigen

Laborworkflow in größerem Maße abweichen, sind nur durch entsprechenden Programmieraufwand in das ReSeqLIMS zu integrieren. Die geplante Integration eines NextGeneration-Sequenziergerätes in das ReSeqLIMS würde der im ReSeqLIMS verwendeten Laborstruktur entsprechen. Auch der NextGeneration-Sequenzierer benötigt eine Arbeitsliste, um Proben zu verarbeiten. Die vom NextGeneration-Sequenzierer generierten Ergebnisse müssten den Ausgangsproben wiederum zugeordnet werden. Für die Integration des NextGeneration-Sequenzierers steht die Definition der neuen Arbeits- und Ausgabenlistenformate im Vordergrund. Hier ist entscheidend, in wie weit der NextGeneration-Sequenzierer mit den vom ReSeqLIMS definierten Datenstrukturen und umgekehrt arbeiten kann, da sich der Laborworkflow des NextGeneration-Sequenzierers erheblich vom bisherigen auf der DNA-Sequenzierung nach Sanger basierendem Laborworkflow unterscheidet. Eine separate Betrachtung von z.B. PCR-Setup und Sequenzierungssetup ist nun nicht mehr möglich. Die Zuordnung von Verbrauchsmaterialien ist nur noch zu dem gesamten Prozess und nicht mehr zu Laborproben der einzelnen Workflow-Steps möglich. Für eine detaillierte Statusanzeige des Fortschrittes einzelner Laborproben wäre eine direkte Kommunikation zwischen LIMS und NextGeneration-Sequenzierer erforderlich.

Das ReSeqLIMS wird im Labor des Instituts für Hygiene des UKMs bereits im Routineeinsatz und für den Resequenzierungsservice, der als Dienstleistung für die Netzwerkpartner angeboten wird, eingesetzt. Dadurch konnte der Bedarf der Optimierung bestehender und weiterer Funktionalitäten identifiziert werden, deren Implementierung aktuell und in den folgenden Versionen des ReSeqLIMS erfolgt. Die Weboberfläche und die Benutzerführung werden u.a. durch Sortierungsfunktionen in den Tabellen, dem möglichen Ein- und Ausblenden von Tabellenspalten, der evtl. sinnvollen grafischen Darstellung bisher tabellarisch angezeigter Daten und der farblichen Unterstützung von Selektionen, Auswahlmöglichkeiten und Qualitätsmerkmalen optimiert und den Benutzerwünschen weiter angepasst. Die Endkontrolle der Resequenzierungsergebnisse ließe sich z.B. durch ein intuitives Ampelsystem unterstützen. Chromatogramme, die eine sehr gute Qualität aufweisen, würden in einer Tabelle grün markiert, eine notwendige manuelle Überprüfung von Chromatogrammen mit einer Qualität im Grenzbereich mit Gelb gekennzeichnet, eine nicht ausreichende Qualität würde durch eine rote Markierung ausgewiesen. Ziel ist es, die Benutzerfreundlichkeit und intuitive Verwendung der ReSeqLIMS-Weboberfläche weiter zu optimieren. Um das direkte Arbeiten an

der Datenbank mittels SQL-Befehlen vollständig umgehen zu können, sind zudem noch einige administrative Funktionalitäten für das ReSeqLIMS zu implementieren. Gleichzeitig gilt, wie bereits für die Arbeits-/Sequenzier- und Ausgabelisten beschrieben, dass durch die Auslagerung von Quellcode in einfacher zu konfigurierende XML-Dateien oder durch Funktionen, die über die ReSeqLIMS-Weboberfläche verwendet werden können, auch für die systemseitig generierten PDF-Dokumente (siehe Kap. 3.2.6.3) durch eine weitere Auslagerung der Textbausteine, wie z.B. dem Text des Anschreibens für das Versenden von Barcodeaufklebern, aus dem Quellcode eine weitere Flexibilisierung in den folgenden ReSeqLIMS-Versionen erfolgen muss. Hier gilt, dass sich bei der Veränderung der Textbausteine in einem PDF-Dokument insbesondere Probleme bei deren Anordnung und dem Layout der neu generierten Dokumentenversion ergeben, da Veränderungen zu unerwünschten Seiten- und Zeilenumbrüchen führen können. Daher besteht die Herausforderung der Implementierung darin, auch nach vorgenommenen Änderungen an den Dokumenten über die ReSeqLIMS-Weboberfläche eine weiterhin korrekte Anordnung der Textbausteine in den PDF-Dokumenten zu garantieren.

Während das Management der Listen und Dokumenten einer weiteren Flexibilisierung bedarf, ist für die Fortschrittsbeschreibung von Aufträgen, Auftragspositionen, Proben, Laborproben und Rechnungen, die im ReSeqLIMS über die Vergabe von kontextsensitiven Status erfolgt, eine Optimierung erforderlich. Die Statusberechnungen erfolgen rekursiv basierend auf einer Baumstruktur, die sich durch die Verknüpfung von Aufträgen, Auftragspositionen, Proben und den Laborproben ergibt (siehe Kap. 3.2.5). Zurzeit zeigt sich, dass die rekursive Berechnung des neuen Status insbesondere in den späten Workflow-Steps des Amplikon-Resequenzierung Workflows wie z.B. der Sequenzierungsproduktaufreinigung bei stark gefüllten 96er Mikrotiterplatten einige Zeit in Anspruch nimmt. Die benötigte Zeit dürfte sich bei der Verwendung von 384er Mikrotiterplatten noch einmal steigern. Daher ist in den folgenden ReSeqLIMS-Versionen die algorithmische Lösung der iterativen Statusberechnung basierend auf der Baumstruktur zu optimieren und ggf. durch eine andere algorithmische bzw. im Bedarfsfall durch eine konzeptionell andere Lösung zu ersetzen.

Mit dem ReSeqLIMS ist ein LIMS entwickelt worden, das speziell für die Anwendung im Rahmen von Hochdurchsatz-Resequenzierungen konzipiert wurde. Durch die Integration der Laborprozesse inkl. der Proben- und Qualitätsdaten, beteiligter Personen und Geräte in einen qualitätsüberwachten Workflow, wird die Laboreffizienz

gesteigert und die Protokollierung von Proben- und Qualitätsdaten ermöglicht. Zusätzlich generiert die Automatisierung der Kundenkommunikation, die Erstellung von gerätespezifischen Listen und die Verwendung von Auftragstemplates sowie allgemein der ReSeqLIMS-Weboberfläche Zeitvorteile und verhindert Fehleingaben durch manuelle Arbeiten. Die Implementierung als Webanwendung macht das ReSeqLIMS zudem universell verfügbar und ermöglicht den leichten Zugriff auf aktuelle Auftrags- und Laborinformationen durch den Kunden und die Mitarbeiter.

Der nächste wichtige Entwicklungsschritt des ReSeqLIMS wird die Integration eines NextGeneration-Sequenzierers. Dieser ist in den bestehenden Laborworkflow zu integrieren, wobei die Herausforderung darin bestehen wird, die neuen Informationsstrukturen und den im Vergleich zu dem auf der DNA-Sequenzierung nach Sanger basierenden veränderten Laborworkflow des NextGeneration-Sequenzierers, in das ReSeqLIMS zu integrieren. Insbesondere im Zusammenhang mit dem NextGeneration-Sequenzierer treten zudem Fragestellungen des Datenmanagements in den Vordergrund, da die pro Sequenzierungslauf generierten extrem hohen Datenmengen effizient und effektiv verwaltet werden müssen. Die Performance des ReSeqLIMS ist aktuell bzgl. der Statusberechnung noch zu verbessern und darf sich trotz des Managements auch großer Datenmengen nicht verringern. Mit der zunehmenden Anzahl an generierten Aufträgen werden Suchfunktionen bzw. Suchmasken zum Auffinden von Aufträgen, Rechnungen, Mikrotiterplatten und Proben zunehmend wichtiger, die im ReSeqLIMS in adäquater Weise angepasst und zur Verfügung gestellt werden müssen. Neben den Methoden der Genotypisierung, die auf einer Resequenzierung von DNA-Fragmenten basieren, sollen in das ReSeqLIMS auch andere Methoden der Genotypisierung wie z.B. MLVA integriert werden. Die MLVA-Methode wird zurzeit im ReSeqLIMS angelegt, um damit auch das Angebot an Methoden weiter zu erhöhen.

5 Literaturverzeichnis

1. Achilles B, Semmler I, Benninger G, Wiethölter A (2010) Zoonosen auf dem Vormarsch - Forscher bündeln ihre Kräfte. DZKF 1-2: 51-54
2. Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62: 53-70
3. Allix-Béguec C, Harmsen D, Weniger T, Supply P, Niemann S (2008) Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol 46: 2692-2699
4. Anderson NR, Lee ES, Brockenbrough JS, Minie ME, Fuller S, Brinkley J, Tarczy-Hornoch P (2007) Issues in biomedical research data management and analysis: needs and barriers. J Am Med Inform Assoc 14: 478-488
5. Apache POI. <http://poi.apache.org/>. Zugegriffen am 6. September 2010
6. Apache Tomcat. <http://tomcat.apache.org/>. Zugegriffen am 6. September 2010
7. Balzert H (2009) Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements Engineering, Spektrum Akademischer Verlag, Heidelberg, 3. Aufl.
8. Becker J, Schütte R (1996) Handelsinformationssysteme, verlag moderne industrie, Landsberg/Lech
9. Beckman Coulter Genomics. <http://www.beckmangenomics.com/>. Zugegriffen am 16. September 2010
10. van Belkum A (2007) Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). FEMS Immunol Med Microbiol 49: 22-27

11. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DMD, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E, Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang G, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53-59
12. Berners-Lee T, Cailliau R, Luotonen A, Nielsen HF, Secret A (1994) The World-Wide Web. *Commun ACM* 37: 76-82
13. BioNumerics. <http://www.applied-maths.com/bionumerics/bionumerics.htm>. Zugegriffen am 9. September 2010
14. Carrilho E, Ruiz-Martinez MC, Berka J, Smirnov I, Goetzinger W, Miller AW, Brady D, Karger BL (1996) Rapid DNA sequencing of more than 1000 bases per run by capillary electrophoresis using replaceable linear polyacrylamide solutions. *Anal Chem* 68: 3305-3313
15. Caugant DA, Feavers IM, Achtman M, Spratt BG, Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95: 3140-3145

16. Chen PP (1976) The entity-relationship model - toward a unified view of data. ACM Trans Database Syst 1: 9-36
17. Cooper JE, Feil EJ (2004) Multilocus sequence typing--what is resolved? Trends Microbiol 12: 373-377
18. De Schrijver JM, De Leeneer K, Lefever S, Sabbe N, Pattyn F, Van Nieuwerburgh F, Coucke P, Deforce D, Vandesompele J, Bekaert S, Hellemans J, Van Criekinge W (2010) Analysing 454 amplicon resequencing experiments using the modular and database oriented Variant Identification Pipeline. BMC Bioinformatics 11: 269
19. Donofrio N, Rajagopalon R, Brown D, Diener S, Windham D, Nolin S, Floyd A, Mitchell T, Galadima N, Tucker S, Orbach MJ, Patel G, Farman M, Pampanwar V, Soderlund C, Lee Y, Dean RA (2005) 'PACLIMS': a component LIM system for high-throughput functional genomic analysis. BMC Bioinformatics 6: 94
20. Eclipse. <http://www.eclipse.org/>. Zugegriffen am 6. September 2010
21. Eclipse Subversive - SVN Team Provider Project. <http://www.eclipse.org/subversive/>. Zugegriffen am 6. September 2010
22. Enright MC, Spratt BG (1999) Multilocus sequence typing. Trends Microbiol 7: 482-487
23. Eurofins. <http://www.eurofins.de/>. Zugegriffen am 16 September 2010
24. GATC Biotech. <http://www.gatc-biotech.com/de/index.html>. Zugegriffen am 16 September 2010
25. Genomics Lab and Data Management System Geneus. <http://www.genologics.com/solutions/research-informatics/genomics>. Zugegriffen am 15 September 2010
26. Geospiza GeneSifter Lab Edition. <http://www.geospiza.com/Products/LabEdition.shtml>. Zugegriffen am 15 September 2010
27. Grundmann H, Aanensen DM, van den Wijngaard CC, Spratt BG, Harmsen D, Friedrich AW, the European Staphylococcal Reference Laboratory Working Group (2010) Geographic distribution of *Staphylococcus aureus* causing invasive infections in europe: a molecular-epidemiological analysis. PLoS Med 7: e1000215
28. Hibernate. <http://www.hibernate.org/>. Zugegriffen am 6 September 2010
29. Hof H, Dörris R (2009) Medizinische Mikrobiologie: Immunologie, Virologie, Bakteriologie, Mykologie, Parasitologie, Klinische Infektologie, Hygiene, Thieme, Stuttgart, 4. Aufl.

30. HTML 4.01 Specification. <http://www.w3.org/TR/1999/REC-html401-19991224/>. Zugegriffen am 6. September 2010
31. iText. <http://itextpdf.com/>. Zugegriffen am 6. September 2010
32. Java Technology. <http://www.sun.com/java/>. Zugegriffen am 6. September 2010
33. JavaServer Facelets. <https://facelets.dev.java.net/>. Zugegriffen am 6. September 2010
34. Jayashree B, Reddy PT, Leeladevi Y, Crouch JH, Mahalakshmi V, Buhariwalla HK, Eshwar KE, Mace E, Folksterma R, Senthilvel S, Varshney RK, Seetha K, Rajalakshmi R, Prasanth VP, Chandra S, Swarupa L, Srikalyani P, Hoisington DA (2006) Laboratory Information Management Software for genotyping workflows: applications in high throughput crop genotyping. *BMC Bioinformatics* 7: 383
35. Kappel G, Pröll B, Reich S, Retschitzegger W (2003) Web Engineering. Systematische Entwicklung von Webanwendungen, Dpunkt Verlag, Heidelberg, 1. Aufl.
36. Kemper A, Eickler A (2004) Datenbanksysteme, Oldenburg Verlag, München, Wien, 5. Aufl.
37. Kircher M, Kelso J (2010) High-throughput DNA sequencing--concepts and limitations. *Bioessays* 32: 524-536
38. Köck R, Harlizius J, Bressan N, Laerberg R, Wieler LH, Witte W, Deurenberg RH, Voss A, Becker K, Friedrich AW (2009) Prevalence and molecular characteristics of methicillin-resistant *Staphylococcus aureus* (MRSA) among pigs on German farms and import of livestock-related MRSA into hospitals. *Eur J Clin Microbiol Infect Dis* 28: 1375-1382
39. LGC Genomics. http://www.lgc.co.uk/divisions/lgc_genomics.aspx. Zugegriffen am 16. September 2010
40. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380
41. Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R (2008) Microbiology in the post-genomic era. *Nat Rev Microbiol* 6: 419-430

42. Monnier S, Cox DG, Albion T, Canzian F (2005) T.I.M.S: TaqMan Information Management System, tools to organize data flow in a genotyping laboratory. *BMC Bioinformatics* 6: 246
43. Moorhead SM, Dykes GA, Cursons RT (2003) An SNP-based PCR assay to differentiate between *Listeria monocytogenes* lineages derived from phylogenetic analysis of the sigB gene. *J Microbiol Methods* 55: 425-432
44. Morisawa H, Hirota M, Toda T (2006) Development of an open source laboratory information management system for 2-D gel electrophoresis-based proteomics workflow. *BMC Bioinformatics* 7: 430
45. Murchan S, Kaufmann ME, Deplano A, de Ryck R, Struelens M, Zinn CE, Fussing V, Salmenlinna S, Vuopio-Varkila J, El Solh N, Cuny C, Witte W, Tassios PT, Legakis N, van Leeuwen W, van Belkum A, Vindel A, Laconcha I, Garaizar J, Haeggman S, Olsson-Liljequist B, Ransjo U, Coombes G, Cookson B (2003) Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J Clin Microbiol* 41: 1574-1585
46. MyFaces. <http://myfaces.apache.org/>. Zugegriffen am 6. September 2010
47. PostgreSQL. <http://www.postgresql.de/>. Zugegriffen am 6. September 2010
48. Reese G (1997) Database Programming with JDBC and Java. Java 1.1, O'Reilly Associates, 1. Aufl.
49. REFA Verband für Arbeitsstudien und Betriebsorganisation e. V. (1993) REFA Methodenlehre der Betriebsorganisation, Lexikon der Betriebsorganisation, Fachbuchverlag Leipzig
50. RichFaces. <http://www.jboss.org/richfaces>. Zugegriffen am 6. September 2010
51. Ridom SeqSphere. <http://www.ridom.de/seqsphere/>. Zugegriffen am 9. September 2010
52. Robertson GA, Thiruvenkataswamy V, Shilling H, Price EP, Huygens F, Henskens FA, Giffard PM (2004) Identification and interrogation of highly informative single nucleotide polymorphism sets defined by bacterial multilocus sequence typing databases. *J Med Microbiol* 53: 35-45
53. Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11: 3-11
54. Sanchez-Villeda H, Schroeder S, Polacco M, McMullen M, Havermann S, Davis G, Vroh-Bi I, Cone K, Sharopova N, Yim Y, Schultz L, Duru N, Musket T, Houchins K, Fang Z, Gardiner J, Coe E (2003) Development of an integrated laboratory information management system for the maize mapping project. *Bioinformatics* 19: 2022-2030

55. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74: 5463-5467
56. Scheer A (1997) Wirtschaftsinformatik. Referenzmodelle für industrielle Geschäftsprozesse. Studienausgabe, Springer-Verlag, Berlin, 2. Aufl.
57. Schwartz DC, Cantor CR (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37: 67-75
58. Sendmail. <http://www.sendmail.org/>. Zugegriffen am 6. September 2010
59. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309: 1728-1732
60. Siemens Simatic IT Unilab.
<http://www.automation.siemens.com/mcms/mes/en/mescomponents/simaticitlims/Pages/Default.aspx>. Zugegriffen am 15. September 2010
61. Smith JM, Smith DCP (1977) Database abstractions: aggregation and generalization. ACM Trans Database Syst 2: 105-133
62. STARLIMS. <http://www.starlims.com/de/>. Zugegriffen am 15. September 2010
63. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsche-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Loch C, van Soolingen D (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 44: 4498-4510
64. The Apache HTTP Server Project. <http://httpd.apache.org/>. Zugegriffen am 6. September 2010
65. Thermo Fischer LIMS.
<http://www.thermoscientific.com/wps/portal/ts/products/catalog?navigationId=L10789&categoryId=81847&pageNum=2>. Zugegriffen am 15. September 2010
66. Urwin R, Maiden MCJ (2003) Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol 11: 479-487
67. Van Rossum T, Tripp B, Daley D (2010) SLIMS--a user-friendly sample operations and inventory management system for genotyping labs. Bioinformatics 26: 1808-1810
68. ViewletBuilder. <http://www.qarbon.com/presentation-software/viewletbuilder/?os=win>. Zugegriffen am 6 September 2010

69. Voegelé C, Tavtigian SV, de Silva D, Cuber S, Thomas A, Le Calvez-Kelm F (2007) A Laboratory Information Management System (LIMS) for a high throughput genetic platform aimed at candidate gene mutation screening. *Bioinformatics* 23: 2504-2506
70. Vossen G (2008) Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme, Oldenburg, München, 5. Aufl.
71. Web Tools Platform (WTP) Project. <http://www.eclipse.org/webtools/>. Zugegriffen am 6. September 2010
72. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Res* 38 Suppl: W326-331
73. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, Achtman M (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 60: 1136-1151

6 Abbildungsverzeichnis

Abb. 1: Im Client-Server-Modell werden weltweit verteilte Clients und Server über das Internet miteinander verbunden.	9
Abb. 2: Die URL wird zur eindeutigen Identifikation von Rechnern, Ressourcen oder Dokumenten im Internet verwendet.	10
Abb. 3: Die Drei-Schichten-Architektur strukturiert ein Softwaresystem in eine Präsentations-, Logik- und Datenschicht.	12
Abb. 4: Beispiel und ERM-Notation [8,56].	14
Abb. 5: Transformation eines ERM in ein relationales Datenbankmodell am Beispiel einer (0,n):(0,m)- und einer (1,1):(0,m)-Beziehung.	15
Abb. 6: Eine Methode im Bereich der DNA-Sequenzierung benötigt mindestens einen Mastermix für die PCR und mindestens einen Primer für die Sequenzierung.	23
Abb. 7: Methoden bestehen aus einer definierten Abfolge von Arbeitsschritten, die jeweils an Laborstationen abgearbeitet werden.	24
Abb. 8: Der Amplikon-Resequenzierung Workflow kann durch das paarweise Anordnen von Laborstationspaaren zusammengesetzt werden.	25
Abb. 9: Methoden lassen sich auf Material-Option-Plastikware Kombinationen und optional auf eine Auswahl zu sequenzierender DNA-Stränge einschränken.	26
Abb. 10: An einer Laborstation werden Verbrauchsmaterialien für die Bearbeitung der Laborproben verwendet. Für die Qualitätskontrolle werden die Verbrauchsmaterialien in Chargen eingeteilt, die den Laborproben verbrauchsbezogen zugeordnet werden.	27
Abb. 11: Eine Probe besitzt immer einen startenden Workflow-Step, an dem die Bearbeitung im Labor beginnt. Die Laborproben durchlaufen Workflow-Steps mit einem definierten Ziel, z.B. der DNA-Sequenzierung.	28
Abb. 12: Zur Bearbeitung von Workflow-Step-spezifischen Laborproben müssen Arbeitslisten generiert werden.	31
Abb. 13: Die Person ist ein zentrales Objekt des ReSeqLIM.	33
Abb. 14: Konditionen können auf den Ebenen Organisation, Projekt und Person definiert werden und beziehen sich i.d.R. auf spezifische Methoden-Material-Option-Plastikware Kombinationen.	34

Abb. 15: Der ReSeqLIMS-Auftrag ist der Auslöser der Kunden-Labor Interaktion.	36
Abb. 16: Das ReSeqLIMS-Gerätemodul verwaltet gerätespezifische Eigenschaften.....	38
Abb. 17: Status geben im ReSeqLIMS Auskunft über den Fortschritt der Prozessierung von Aufträgen, Auftragspositionen, Proben, Laborproben und Rechnungen.	39
Abb. 18: GUI der ReSeqLIMS-Webanwendung. (A) zeigt die Navigationsleiste, (B) die Kopfzeile, der Inhalt und die Funktionalität befindet sich in (C)...	41
Abb. 19: Während der Registrierung an der ReSeqLIMS-Webanwendung werden in einem Webformular die Personenstammdaten erfasst (A). Nach der erfolgreichen Registrierung wird die Kundennummer generiert (B).	42
Abb. 20: Beispiele für Startbildschirme nach erfolgreichem Login an der ReSeqLIMS- Webanwendung. Kunden sehen den „Place Order“- (A), Mitarbeiter den „Order Overview“-Bildschirm (B). Die Menüpunkte der Navigationsleisten in (A) und (B) richten sich nach den Nutzergruppen, z.B. für den Kunden (A) und den Administrator (B).	43
Abb. 21: In der Auftragsübersicht kann der Status (X) eines Auftrages eingesehen, Detailinformationen angezeigt ((A) - (D)) und z.B. die Chromatogramme in einer gepipten Ergebnisdatei heruntergeladen werden (E).	47
Abb. 22: Darstellung der automatisierten E-Mail-Kommunikation des ReSeqLIMS. Durchgezogene Linien repräsentieren systemgenerierte E-Mails, gestrichelte Linien repräsentieren manuelle Aktionen durch den Mitarbeiter oder den Kunden.....	49
Abb. 23: Verkettung der Laborproben über mehrere Laborstationen. An der ersten Laborstation findet eine Laborprobenvervielfältigung statt (Pos.: Platten- Position, V-ID: Vorgänger-Proben-ID).....	50
Abb. 24: Der ReSeqLIMS-Bildschirm des „List Management“ ist unterteilt in das Arbeitslisten-Fenster (A), das Workflow-Fenster (B), das Plastikwaren- Fenster (C) und das Auftrags-Fenster (D).	52
Abb. 25: Die Reiter einer Laborstationsansicht im „List Management“ zeigen die Laborproben Status-spezifisch an. Abschnitt (A) zeigt die Plastikwaren, (B) die darauf enthaltenen Laborproben.....	53
Abb. 26: Das Arbeitslistenpopup an der Laborstation PCR-Setup enthält detaillierte Informationen u.a. bzgl. der Ausgangs- und Zielplastikwaren sowie der beinhaltenden Laborproben.	54
Abb. 27: Für manuelle Arbeitsschritte im Labor ist das Erstellen der Zielausgabeliste über die ReSeqLIMS-Weboberfläche erforderlich. In Abschnitt (A) werden Ziellaborstation und -plastikware definiert, Abschnitt (B) zeigt plattenspezifisch die generierten Ausgabelisten, zu denen der Zielbarcode angegeben werden muss.	57

Abb. 28: Berechnung der Statusänderungen bei Laborproben, wenn die Statusreihenfolge A, B und C ist (von niedrig nach hoch). Der zu C veränderte Status in 2. führt auf der nächsthöheren Ebene in 3. ebenfalls zu einer Statusänderung. Diese Statusänderung führt auf der nächsthöheren Ebene in 4. nicht mehr zu einer weiteren Statusveränderung, da auf der darunterliegenden Ebene noch zwei Laborproben mit einem niedrigerer Status vorliegen, die damit den Status auf der nächsthöheren Ebene bestimmen.....	60
Abb. 29: Auftragstemplates für Probanden einer 96er Mikrotiterplatte. Abschnitt (A) beinhaltet allgemeine Auftragsinformationen, Abschnitt (B) die Probenamen, die mit Hilfsfunktionen aus Abschnitt (C) erstellt werden können.	61
Abb. 30: Beispiel einer Arbeits- (A) und der zugehörigen Ausgabeliste (B) an der PCR-Setup Laborstation.	63
Abb. 31: PDF-Dokumente des ReSeqLIMS. (A), (C) - (F) zeigt eine Rechnung, (B) und (G) das Barcodeaufkleber-Anschreiben.	65
Abb. 32: Beispiel einer ReSeqLIMS-Sequenzierliste. Das Format ist spezifisch an den vom Labor verwendeten DNA-Sequenzierer angepasst.....	66
Abb. 33: Die ReSeqLIMS-Ordnerstruktur.	I
Abb. 34: Die Struktur der gezippten Ergebnisdatei ist definiert durch den Auftrag und die (Auftrags-) Proben. Die Probenamen-Bestandteile sample10 und sample11 sind vom Kunden vorgegeben. Genname, Primernamen, Barcode und Plattenposition der Sequenzierplatte werden systemseitig an den Probenamen angehängt.....	IV

7 Tabellenverzeichnis

Tab. 1: Kurze Beschreibung der ausgewählten kommerziellen und Open-Source-Softwareprodukte, die auf einen Einsatz im Rahmen des PBA-Zoo Projektes analysiert wurden.	V
Tab. 2: Ausgewählte Funktionalitäten der kommerziellen und Open-Source-Software aus Tab. 1. Während das „X“ eine Funktion nachweist, bedeutet der „-“ nur, dass in der jeweiligen Publikation diese Funktionalität nicht erwähnt wurde und daher im Rahmen dieser Analyse davon ausgegangen werden muss, dass diese Funktion nicht vorhanden ist.....	VI

8 Abkürzungsverzeichnis

A	Adenin
Abb.	Abbildung
Anh.	Anhang
AJAX	Asynchronous JavaScript and XML
allg.	allgemein
API	Application Programming Interface
BMBF	Bundesministerium für Bildung und Forschung
bp	Basenpaare
bzgl.	bezüglich
bzw.	beziehungsweise
C	Cytosin
cDNA	complementary deoxyribonucleic acid
DACH	Deutsche Akkreditierungsstelle Chemie
ddNTP	dideoxy nucleoside triphosphates
DDL	Data Definition Language
DML	Data Manipulation Language
DNA	Deoxyribonucleic acid
DV	Datenverarbeitung
E-Mail	Electronic Mail
ERM	Entity Relationship Model
evtl.	eventuell
FTP	File Transfer Protocol
G	Guanin
Gb	Gigabasenpaare
GUI	Graphical User Interface
HGP	Human Genom Projekt
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
i.d.R.	in der Regel
IDE	Integrated Development Environment
InDel	Insertion und Deletion
inkl.	inklusive
IP	Internet Protocol
JEE	Java Platform Enterprise Edition
JRE	Java Runtime Environment
JSF	JavaServer Faces

JSP	JavaServer Pages
JVM	Java Virtual Machine
kb	Kilobasenpaare
LIMS	Laborinformations- und Managementsystem
MLST	Multilocus Sequence Typing
MLVA	Multi Loci Variable Number Tandem Repeat
MS	Microsoft
PBA-Zoo	Phylogenie, Bio-Informatik und Amplikon-Resequenzierung von Zoonose-Erregern
PCR	Polymerase Chain Reaction
PDF	Portable Document Format
PFGE	Pulsed Field Gelelectrophoresis
PHP	Hypertext Preprocessor
POI	Poor Obfuscation Implementation
RIA	Rich-Interface Applications
rRNA	ribosomal ribonucleic acid
SNP	Single Nucleotide Polymorphism
SQL	Structured Query Language
T	Thymin
Tab.	Tabelle
TCP	Transmission Control Protocol
u.a.	unter anderem
UKM	Universitätsklinikum Münster
URL	Uniform Resource Locator
v.a.	vor allem
VBA	Visual Basic for Applications
W3C	World Wide Web Consortium
WHO	World Health Organization
WTP	Web-Tools-Plattform
WWW	World Wide Web
XHTML	Extensible HTML
XML	Extensible Markup Language
z.B.	zum Beispiel
z.T.	zum Teil
z.Z.	zurzeit

- Lebenslauf -

- Danksagung -

Anhang

A Die Ordnerstruktur des ReSeqLIMS

Die Ordnerstruktur des ReSeqLIMS befindet sich zentral auf einem Webserver und ist damit jederzeit für die Webanwendung zu erreichen. Durch das Einbinden der Ordnerstruktur als Netzlaufwerk in die Ordnerstruktur eines beliebigen Arbeitsplatzrechners, kann auch lokal bequem auf alle Ordner der ReSeqLIMS-Ordnerstruktur auch lokal zugegriffen werden. Im Ordner „orderFiles“ werden die Aufträge der Kunden gesichert (siehe Abb. 33).

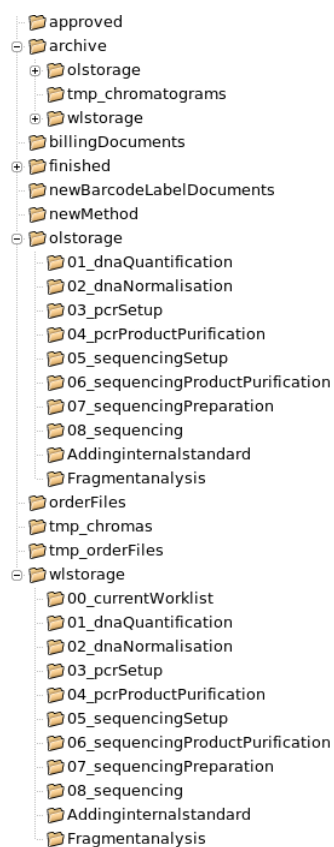


Abb. 33: Die ReSeqLIMS-Ordnerstruktur.

Die Ordner „wlstorage“ und „olstorage“ beinhalten die vom ReSeqLIMS bzw. vom Pipettierroboter generierten Arbeits- und Ausgabelisten. Die Laborstation-spezifischen Unterordner ermöglichen, dass Arbeits- und Ausgabelisten in den Ordnern der Laborstation abgespeichert werden, an der die Arbeits- und Ausgabelisten generiert worden sind. Während die abgearbeiteten Aufträge vor ihrer Endkontrolle im Ordner „finished“ abgelegt werden (z.B. im Amplikon-Resequenzierung Workflow in Form einer gezippten Ergebnisdatei, die die generierten Chromatogramme enthält), werden

die nach einer finalen Qualitätskontrolle freigegebenen Aufträge im Ordner „approved“ abgelegt. Die Ordner „newBarcodeLabelDocuments“ und „billingDocuments“ enthalten PDF-Dateien, die zum Versenden der Rechnungen oder neuer Barcodeaufkleber erstellt werden. Die mit tmp_ bezeichneten Ordner „tmp_orderFiles“ und „tmp_chromas“ beinhalten Auftragsdateien und Chromatogramme, die noch einer weiteren Bearbeitung bzw. Überprüfung unterzogen werden, bis sie in ihren Zielordner weiterkopiert werden können. Das Anlegen neuer Methoden kann durch den Import einer Excel-Datei erfolgen. Zu Dokumentationszwecken wird die Excel-Datei im Ordner „newMethod“ abgespeichert.

B Status des ReSeqLIMS

Status für Aufträge, Auftragspositionen, Proben und Laborproben

Not arrived
Arrived
Quantification in process
Quantification finished
Normalization in process
Normalization finished
PCR setup in process
PCR setup finished
PCR product purification in process
PCR product purification finished
Sequencing setup in process
Sequencing setup finished
Sequencing product purification in process
Sequencing product purification finished
Sequencing preparation in process
Sequencing preparation finished
Sequencing in process
Sequencing finished
Adding internal standard in process
Adding internal standard finished
Fragment analysis in process
Fragment analysis finished
Order approved (nur für Aufträge)
Order position approved (nur für Auftragspositionen)
Sample approved (nur für Proben)
Ok (nur für Laborproben)
Resigned

Status für Rechnungen

Unbilled
Open
Paid

C Die Struktur der gezippten Ergebnisdatei des Amplikon-Resequenzierung Workflow

Die Struktur der gezippten Ergebnisdatei des Amplikon-Resequenzierung Workflows orientiert sich an der Struktur des Auftrages und der (Auftrags-) Proben (siehe Abb. 34).

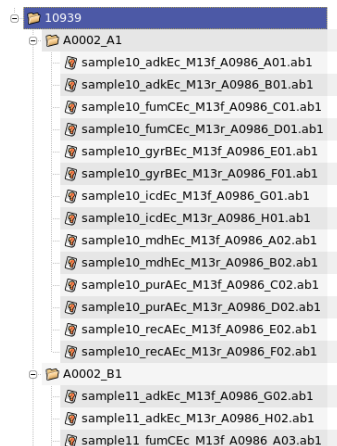


Abb. 34: Die Struktur der gezippten Ergebnisdatei ist definiert durch den Auftrag und die (Auftrags-) Proben. Die Probenamen-Bestandteile sample10 und sample11 sind vom Kunden vorgegeben. Genname, Primernamen, Barcode und Plattenposition der Sequenzierplatte werden systemseitig an den Probenamen angehängt.

Auf der obersten Ebene befindet sich ein Ordner, der entsprechend der Auftrags-ID benannt wird. Die nächste Ebene umfasst Ordner, die entsprechend der Barcodes der Auftragspositionen benannt sind. Handelt es sich bei einer Auftragsposition um eine Mikrotiterplatte, sind die Barcode-Ordnernamen zusätzlich mit der jeweiligen Plattenposition versehen. In den Barcode/Position-Ordner sind ausschließlich die Chromatogramme zu finden, die der entsprechenden Probe eindeutig zuzuordnen sind. Ausschließlich zur Zuordnung der Chromatogramme zu den sequenzierten Laborproben sowie zu Zwecken der Dokumentation und Qualitätssicherung enthält der Chromatogrammname zusätzlich den Barcode und die Plattenposition der Sequenzierplatte.

D Analysierte kommerzielle und Open-Source-Softwareprodukte

Im Rahmen des PBA-Zoo Projektes wurden kommerzielle und Open-Source-Softwareprodukte auf eine Eignung für den Einsatz in Hochdurchsatz-Resequenzierungen im Institut für Hygiene des Uniklinikum Münster analysiert. In Tab.

1 wird die Funktionalität der Softwareprodukte kurz beschrieben. Tabelle 2 zeigt detailliert die Funktionalität, die Softwarearchitektur und die verwendeten Programmiersprachen und Datenbanken der analysierten Softwareprodukte aus Tab. 1.

Kommerzielle Software	
Geospiza GeneSifter	Geospiza GeneSifter ist ein umfangreiches LIMS für den Einsatz in einem Hochdurchsatz-Labor. Es ermöglicht die Probenverfolgung und die Protokollierung von Qualitätsdaten über alle Schritte eines Workflows und ist sehr variabel konfigurierbar. Eine Vielzahl an Instrumenten, Plastikwaren etc. sind bereits vordefiniert.
Thermo Fischer Scientific Nautilus/Sample Manager	Thermo Fischer Scientific Nautilus und Sample Manager sind geeignet für Probenverfolgung und Protokollierung von Informationen in einem Workflow. Eingesetzt in Forschungs- und Entwicklungslaboren.
GenoLogics Geneus	Exklusiv für den Genomik Bereich entwickeltes LIMS. Ermöglicht Probenverfolgung, Integration der Laborinstrumente, Online-Auftragsgenerierung und Ergebnisverwaltung.
Siemens Simatic IT Unilab	LIMS Lösung mit Spitzenplatz in der Prozess-Industrie. Verwaltet und konfiguriert den vollständigen Workflow im Labor bzgl. Proben, Analysen, Instrumenten und Berichten. Fokus: Qualitätssicherung und Effizienzsteigerungen.
STARLIMS	LIMS zur Verwaltung, Archivierung und Rückverfolgung von Daten verschiedener Formate.
Open-Source-Software	
T.I.M.S.	TaqMan Information Management System (T.I.M.S) Proben-Management basiert auf dem Einlesen von TaqMan Arbeits- und Ausgabe Dateien.
PACLIMS	Komponenten LIMS für Hochdurchsatz funktionale Genomik-Analysen.
SLIMS	Das Sample-based LIMS (SLIMS), ermöglicht die Anzeige, die Bearbeitung und das Anlegen von Probeninformationen. Programm zum Plattenmanagement.
2D LIMS	LIMS für 2-D Gelelektrophorese basierenden Proteomik Workflow. Unterstützt Verarbeitung von Import- und Export-Daten von Massenspektroskopie und 2-D Gel-Image Analyse Software.
MMP-LIMS	Das Maize Mapping Projekt (MMP)-LIMS wird in einer Hochdurchsatz-Umgebung zur Erstellung einer Genkarte eingesetzt. Ziel: Datenmanagement, Datenqualität und -zugriff sicherstellen und den Datenaustausch zwischen Forschergruppen unterstützen.
Candidate gene mutation screening	LIMS in einer Hochdurchsatz-Umgebung, die striktes Prozessmanagement und sorgfältige Datenverfolgung bzw. -protokollierung erfordert.
Crop Genotyping	LIMS in einer Hochdurchsatz-Umgebung. Zur Handhabung der großen Datenmengen sind die Probenverfolgung und die Protokollierung von Daten auf jeder Stufe eines Workflow notwendig.
Variant Identification Pipeline	Pipeline für diagnostische Resequenzierungsprojekte. Automatische Analyse eines Single GS-FLX Amplikon-Resequenzierungslauf, von den Raw-Daten bis zu maßgeschneiderten Berichten.

Tab. 1: Kurze Beschreibung der ausgewählten kommerziellen und Open-Source-Softwareprodukte, die auf einen Einsatz im Rahmen des PBA-Zoo Projektes analysiert wurden.

	Client-Server	Web-anwendung	Labor-workflow-orientiert	Proben-/Datenverfolgung	Modularisierung	Registrierung	Qualitätsdaten-Aufzeichnung	Analysemodul	Nutzergruppen	Konfigurierbar	Programmiersprachen	Arbeits- und/oder Ausgabelisten	Barcodes	Online-Aufträge	Grafische Plattendarstellung	Berichte	Rechnungswesen	Datenbank
kommerzielle Software																		
Geospiza GeneSifter	X	X	X	X	X	X	X	-	X	X	-	X	X	X	X	X	X	-
Thermo Fischer Scientific Nautilus/Sample Manager	X	X	X	X	X	X	X	-	X	X	-	X	X	-	X	X	-	-
GenoLogics Geneus	X	X	X	X	-	X	X	X	X	X	-	X	X	X	-	X	X	-
Siemens Simatic IT Unilab	X	X	X	X	X	X	X	X	X	X	Oracle	X	X	-	-	X	-	-
StarLIMS	X	X	X	X	-	X	-	-	X	-	.NET	-	-	-	-	X	-	-
Open-Source-Software																		
T.I.M.S.	-	-	X	-	-	-	-	X	-	-	VisualBasic	X	-	-	-	-	-	lokale Datenbank
PACLIMS	X		X	X	X	-	-	X	-	-	Perl	-	X	-	-	X	-	PostgreSQL
SLIMS	X	X	-	X	-	X	-	-	X	-	Java/JSP	-	-	-	X	-	-	Hibernate DB2/MySQL
2D LIMS	X	X	X	-	-	X	-	-	X	-	PostgreSQL PHP	-	-	-	X	X	-	-
MMP-LIMS	X	X	-	X	X	-	-	X	X	-	VisualBasic JavaApplets Perl	X	-	-	-	-	-	Lokal: Access Web: Oracle
Candidate gene mutation screening	X	X	X	X	-	X	-	-	-	-	JSP	X	X	-	X	-	-	MySQL
Crop Genotyping	X	X	X	X	X	X	X	X	-	-	JavaStruts Framework	X	-	X	X	X	-	MS-SQL 2000/ PostgreSQL
Variant Identification Pipeline	-	-	-	-	X	-	X	X	-	-	Perl	X	-	-	-	X	-	MySQL

Tab. 2: Ausgewählte Funktionalitäten der kommerziellen und Open-Source-Software aus Tab. 1. Während das „X“ eine Funktion nachweist, bedeutet der „-“ nur, dass in der jeweiligen Publikation diese Funktionalität nicht erwähnt wurde und daher im Rahmen dieser Analyse davon ausgegangen werden muss, dass diese Funktion nicht vorhanden ist.