

Serviceleistungen der Bibliothek des Deutschen Krebsforschungszentrums im Bereich Forschungsdatenmanagement

Dagmar Sitek

Deutsches Krebsforschungszentrum (DKFZ), Heidelberg

Abstract

Das Thema Forschungsdatenmanagement gewinnt zunehmend an Bedeutung. Die Entwicklung und Implementierung von disziplinspezifischen Lösungen für die Wissenschaft und Forschung sind essentiell, um die großen Herausforderungen, die sich in diesem Bereich ergeben, zu bewältigen. Innerhalb dieses Entwicklungsprozesses können auch Bibliotheken neue Aufgaben übernehmen, um sich als leistungsstarker Partner für die Wissenschaft zu positionieren. Die Bibliothek des Deutschen Krebsforschungszentrums hat in Kooperation mit den Wissenschaftlern „maßgeschneiderte“ Lösungen in diesem Kontext entwickelt, um adäquate Unterstützung anbieten zu können.

Forschungsdatenmanagement – ein Thema von zunehmender Relevanz

Der Umgang mit Forschungsdaten hat in den letzten Jahren zunehmend an Bedeutung gewonnen. Die Anforderungen, die von den verschiedenen Stakeholdern an die Forschenden in diesem Kontext gestellt werden, steigen zunehmend und werden immer komplexer. So fordern beispielsweise wissenschaftliche Zeitschriften, dass die Daten, die den eingereichten Publikationen zugrunde liegen, für den Review-Prozess zur Verfügung gestellt und mit dem Artikel veröffentlicht werden müssen. Dies erfolgt in der Regel in Form von „Supplementary Material“, welches in die Artikel integriert wird und damit die publizierten Ergebnisse überprüfbar und nachvollziehbar macht. Ebenso steigt bei den Zuwendungsgebern zunehmend das Bewusstsein, dass Daten, ebenso wie Publikationen, die aus öffentlichen Mitteln finanziert wurden, auch der Öffentlichkeit zur Verfügung stehen sollten, weil die dadurch ermöglichte Nachnutzung der Daten positive Auswirkungen auf die Forschungsprozesse hat. So hat unter anderem die Europäische Union in ihrem Rahmenprogramm für Forschung und

Innovation, Horizon 2020¹, einen „Open Research Data Pilot“ integriert, welcher beinhaltet, dass für bestimmte Projekte, die aus ausgewählten Programmen finanziert werden, Forschungsdatenmanagementpläne vorgelegt und die Daten Open Access zur Verfügung gestellt werden müssen.²

Auf der einen Seite gibt es also steigende Anforderungen an die Forschenden, ihre Daten effektiv zu managen und zur Verfügung zu stellen, auf der anderen Seite mangelt es aber zur Zeit noch häufig an geeigneten Systemen, Workflows und Mechanismen, um diese Aufgaben ohne allzu großen zusätzlichen Aufwand realisieren zu können.

In diesem Spannungsfeld bietet sich für Bibliotheken die Möglichkeit, neue Aufgaben zu erschließen und sich so als leistungsstarker Partner für die Wissenschaft zu positionieren.

Forschungsdaten im Deutschen Krebsforschungszentrum

Das Deutsche Krebsforschungszentrum (DKFZ)³ ist die größte biomedizinische Forschungseinrichtung in Deutschland und ist Mitglied in der Helmholtz-Gemeinschaft Deutscher Forschungszentren⁴. Seine Aufgabe besteht in der Erforschung von Krebsentstehung, Krebsrisikofaktoren und Krebsprävention sowie in der Entwicklung von neuen Behandlungsansätzen und Therapien.

Wie in den meisten Forschungseinrichtungen fallen auch im DKFZ sehr große Mengen von unterschiedlich strukturierten Forschungsdaten an. Die großen, sehr datenintensiven Projekte, wie z.B. bestimmte klinische Studien oder Genomsequenzierungsprojekte, verfügen in der Regel über Forschungsdatenmanagementpläne sowie geeignete Infrastrukturen und sind damit gut erschlossen. Ganz anders stellt sich die Situation bei den unzähligen einzelnen Datenelementen, wie z.B. den Excel-Tabellen, Abbildungen, Elektronenmikroskopieaufnahmen usw., die während eines Forschungsprozesses anfallen, dar. Diese vielen kleinen Datenmengen, die zusammengenommen durchaus den Umfang von Big-Data-Projekten erreichen, liegen häufig verteilt auf den Rechnern und Laptops der einzelnen

¹ <http://www.horizont2020.de/>

² <http://www.horizont2020.de/einstieg-open-access.htm>

³ <http://www.dkfz.de/de/index.html>

⁴ <http://www.helmholtz.de/>

Forschenden. Diese Situation in Verbindung mit der hohen Fluktuation, die in Forschungsinstitutionen herrscht, führt zu zahlreichen Problemen. Im günstigsten Fall erfolgt eine Übergabe dieser Forschungsdaten an den verantwortlichen Abteilungsleiter⁵, der sie speichern und entsprechend erschließen muss, um das Wiederauffinden zu gewährleisten. Es besteht aber die Gefahr, dass dies nicht passiert und auf diese Weise Daten verloren gehen, wenn zum Beispiel Forschende das DKFZ verlassen oder Abteilungen aufgelöst werden. Dies ist nicht nur problematisch, weil die Daten auf diese Weise für eine Nachnutzung verloren gehen, sondern weil so auch die Regeln der „Guten wissenschaftlichen Praxis“ nicht eingehalten werden. In ihrer Denkschrift „Sicherung guter wissenschaftlicher Praxis“, fordert die Deutsche Forschungsgemeinschaft aus Gründen der Nachprüfbarkeit: „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden“.⁶

Somit ist ein effektives Forschungsdatenmanagement der vielen kleinen Datenelementen nicht nur sinnvoll sondern auch zwingend erforderlich.

Anders, aber ebenfalls problematisch, stellt sich die Situation bei den Forschungsdaten dar, die in speziell entwickelten Datenbanken gespeichert sind. Es gibt im DKFZ zahlreiche Inhouse-Datenbanken, die von einzelnen Abteilungen im Haus für ganz spezifische Forschungsdaten entwickelt wurden. Als Beispiel ist hier eine Datenbank mit Ultraschallbildern oder eine weitere Datenbank mit Röntgenbildern zu nennen. Dort liegen die Daten zwar in der Regel gut archiviert und erschlossen, jedoch unverbunden vor. Im Bedarfsfall müssen sie gegebenenfalls mühsam recherchiert und zusammengestellt werden.

Ein sinnvoller Ansatz in diesem Kontext wäre es, alle Forschungsdaten, die in einem Projekt anfallen, an einer Stelle zusammenzuführen. Da das Ergebnis eines Forschungsprozesses, zumindest momentan noch, als Zeitschriftenartikel veröffentlicht wird, liegt es nahe, diese Verknüpfung in der Publikationsdatenbank zu realisieren.

⁵ Aufgrund der einfacheren Lesbarkeit wird im Folgenden nur die männliche Form verwendet, die weibliche Form ist dabei jeweils mit eingeschlossen.

⁶ http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf S 21

Die DKFZ-Publikationsdatenbank enthält alle Veröffentlichungen der Mitarbeiter des DKFZ. Da die Bibliothek für die Entwicklung, Organisation und Pflege der Datenbank verantwortlich ist, konnte in diesem Bereich ein Serviceangebot implementiert werden, welches den Forschenden Unterstützung im Forschungsdatenmanagement anbietet. Ziel ist es, alle Forschungsdaten, die zu einer Publikation geführt haben, in der Publikationsdatenbank bei der entsprechenden Publikation zusammenzuführen und nachzuweisen. Dies wird, je nach Art der Forschungsdaten, unterschiedlich realisiert und soll zukünftig noch stark ausgebaut werden.

Verbindung Mausdatenbank - Publikationsdatenbank

Die Abteilung Kryokonservierung des DKFZ hat eine Datenbank für transgene Mauslinien entwickelt. In ihr werden Mausmodelle erfasst und detailliert beschrieben. Unter anderem werden dort auch Artikel, in denen das jeweilige Mausmodell publiziert wurde, eingetragen.⁷ Viele dieser Artikel wurden von Forschenden des DKFZ publiziert, aus diesem Grund wurde bei dem Aufbau der Mausdatenbank unter anderem in der Publikationsdatenbank recherchiert. Da beide Datenbanken Inhouse-Lösungen waren und sich beide noch in der Entwicklungsphase befanden, bot sich eine sehr gute Gelegenheit, beide miteinander zu verbinden. Es wurde eine Schnittstelle programmiert, die dazu führt, dass der Eintrag einer Publikation in der Mausdatenbank automatisch eine Meldung an die Bibliothek generiert. Handelt es sich um einen Artikel, der in der Publikationsdatenbank enthalten ist, so wird von der Bibliothek eine Verknüpfung durch einen Link zwischen den beiden Einträgen in beiden Datenbanken geschaffen. In der Mausdatenbank wird dadurch der verlinkte Hinweis zu der Publikation in der Publikationsdatenbank angezeigt. (Abb. 1)

⁷ Die Datenbank steht nur DKFZ intern zur Verfügung. Nähere Informationen dazu sind unter anderem in folgendem Artikel zu finden: Staudt M, Trauth J, El Hindi I, Galuschka C, Sitek D, Schenkel J. Managing major data of genetically modified mice – from scientific demands to legal obligations. *Transgenic Res* 2012; 21 (5): 959-966.

Einleitung	Übersicht	Kontakt				
Beschreibung der Mauslinie						
Mauslinie#	ID: 483 Eingabedatum: 24.03.2009 Änderungsdatum: 26.11.2013					
Short name: *	HPV11 E2 x URR11					
Tierbase Nr.:	2390	Stock-Nr.:				
Beschreibung der Mutante:	Mutationsart	mutiertes Gen / Transgen	Tg Spender	reg. Element	reg. Elem. Spender	verwendeter Vektor
	Transgener Überexprimierer	HPV11 E2	HPV 11	Ubiquitin C promoter	Homo sapiens	pUC 18
	Transgener Überexprimierer	LacZ	E. coli	HPV11 URR	HPV 11	pBluescript KS
Risikogruppe: (GenTG) *	1					
Gene ID# (MGI, OMIM):						
Vater (Name der Mutante), Genotyp:	HPV11 E2 x URR11					
Mutter (Name der Mutante), Genotyp:	HPV11 E2 x URR11 or WT					
Genetischer Hintergrund:	B6D2Fn					
Rückkreuzungsgeneration:	unknown/mixed					
Full name:	B6;D2-Tg(HPV11-lacZ)1704Aal Tg(UBC-HPV11E2)613Josc					nützliche Links: MGI Jax ILAR Labcode OMIM
Synonyme:						
Fragestellung hinter der Entwicklung/ Kurzbeschreibung der Mutante: *	Mouse model for HPV-11 regulation					
Belastungseinstufung: (TierSchG) *	keine					
Beschreibung der Belastung inkl. der belasteten Genotypen: *	Reportermaus, keine Belastung beobachtet					
Welche Besonderheiten weisen die Mäuse auf:						
Homozygot Letal:	Nein					
Besondere Körpermerkmale:						
Besondere Verhaltensmerkmale:						
Besondere Reproduktionsmerkmale:						
Keywords:	Promoter, skin, transcription, hair follicle, animal model, E2 protein, Papillomavirus, transgenic mice Ubiquitin C promoter					
Publikation:	Schenkel J, Gaissert H, Protopapa EE, Weiher H, Gissmann L, Alonso A (1999): The human Papillomavirus type 11 upstream regulatory region triggers hair-follicle-specific gene expression in transgenic mice. J Invest Dermatol 112, 893-898. Leykauf,K., Kabsch,K., Gassler,N., Gissmann,L., Alonso,A., Schenkel,J.: Expression of the HPV11 E2 gene in transgenic mice does not result in alterations of the phenotypic pattern. Transgenic Research 17 (1), 1-8, 2008.					

Abbildung 1

In der Publikationsdatenbank wiederum sind bei dem Artikel die Links zu den Mäusen in der Mausdatenbank aufgeführt, die darin publiziert sind. (Abb.2)

Journal-Artikel

Kostenstelle/n	W430, F020, F050
Allianzen / Kooperationen	
Programmatische Zuordnung	Krebsforschung
Artikel	Originalartikel
Titel	Expression of the HPV11 E2 gene in transgenic mice does not result in alterations of the phenotypic pattern
Autoren	Leykauf,K., Kabsch,K., Gassler,N., Gissmann,L., Alonso,A., Schenkel,J.
Journalname	Transgenic Research
Volume	17
Heftnummer	1
Seitenzahl von	1
Seitenzahl bis	8
Erscheinungsjahr	2008
Online First / epub ahead of print	Nein
Erscheinungsjahr	
in press	Nein
supplementary material	
Web of Science	000251869000001
PubMed	17701441
Laborbuchnummer	
Mausdatenbank	HPV11 E2 613 HPV11 E2 623 HPV11 E2 871 HPV11 E2 871 HPV11 E2 x URR11

Abbildung 2

Durch diese Verknüpfung bekommen die Nutzer die Information zusammenhängend präsentiert und können komfortabel zwischen beiden Datenbanken hin und her wechseln, ohne eine neue Recherche starten zu müssen. Dieses Konzept ist auf großes Interesse gestoßen und es sollen zukünftig weitere Datenbanken integriert werden.

Verbindung Laborbücher – Publikationsdatenbank

Neben der Anbindung von Datenbanken ist auch die Anbindung von Laborbüchern in der Planung. Im DKFZ sind zurzeit sowohl gedruckte Laborbücher als auch ein elektronisches Laborbuchsystem im Einsatz. Die elektronischen Laborbücher sollen, analog zu den Datenbanken, per Link mit den Veröffentlichungen in der Publikationsdatenbank verbunden werden. Für die gedruckten Laborbücher ist bereits ein Feld für die Laborbuchnummer angelegt (Abb. 2), so dass diese dort eingetragen werden kann. Da die Ausgabe und die Verwaltung der gedruckten Laborbücher durch die Bibliothek erfolgt, ist nicht nur nachzuvollziehen, in welchem Laborbuch Forschungsdaten zu der Publikation enthalten sind, sondern auch wo es sich befindet.

Verbindung Supplementary Plattform - Publikationsdatenbank

Für die bereits erwähnten vielen einzelnen Datenelemente, die nicht strukturiert in einer Datenbank abgelegt sind (Abbildungen, Tabellen usw.) wurde von der Bibliothek zusammen mit der Abteilung für Informationstechnologie (ITCF) eine Plattform für „Supplementary Material“ implementiert, auf der alle Arten von Daten gespeichert werden können. Der Wunsch nach einer solchen Möglichkeit, seine Daten schnell und komfortabel, unabhängig von den Servern der Zeitschriftenverlage, ablegen zu können, kam von Wissenschaftlern aus dem Haus. Die Verbindung der dort gespeicherten Materialien mit der entsprechenden Publikation in der Publikationsdatenbank erfolgt ebenfalls mittels eines Links.

Die Plattform wurde mittlerweile um einige Features erweitert und kann nun auch von den Forschenden dazu genutzt werden, Reviewern für einen befristeten Zeitraum Supplementary Material zu eingereichten Publikationen zugänglich zu machen.

Wird der Artikel angenommen und die Veröffentlichung in die Publikationsdatenbank eingetragen, dann können die Materialien schnell und komfortabel damit verbunden werden, ohne sie erneut hochladen zu müssen

Vorteile dieses Lösungsansatzes

Die Vorteile eines solchen pragmatischen Ansatzes sind sehr vielfältig.

So lässt er sich schnell und unproblematisch mit allen Arten von Forschungsdaten, egal wie sie strukturiert sind und wo sie gespeichert sind, realisieren und damit erweitern und ausbauen. Dies beinhaltet zum Beispiel nicht nur Inhouse-Datenbanken sondern es kann auch auf externe Systeme verlinkt werden, wenn die Inhalte frei zugänglich sind. Aufgrund des Links in der Publikationsdatenbank auf die verschiedenen Systeme, die Forschungsdaten enthalten, bleiben deren „Besonderheiten“, wie zum Beispiel Zugangsbeschränkungen für bestimmte Nutzer, ohne zusätzlichen Aufwand erhalten. So kann sich ein externer Nutzer gegebenenfalls das Supplementary Material zu einer Publikation ansehen, weil es freigegeben ist, die damit ebenfalls verlinkten Daten aus einer Datenbank aber unter Umständen nicht, wenn diese nur DKFZ intern zur Verfügung steht. Für die Wissenschaftler bietet dieser Ansatz zwei große Vorteile. Erstens werden die Informationen zu allen Forschungsdaten, die einer Publikation zugrunde liegen, an einer Stelle, der Publikationsdatenbank, nachgewiesen. Zweitens sind die Forschungsdaten aufgrund der Verbindung zur Publikation durch diese gewissermaßen erschlossen. Die Aufbereitung von solchen Daten, die gewährleisten, dass sie schnell wiedergefunden werden können, ist sehr zeitaufwendig und wird von den Forschenden daher oft nicht praktiziert. Verlässt dann zum Beispiel ein Doktorand das DKFZ kommt es häufig vor, dass er eine große Menge von Dateien hinterlässt bei denen niemand auf den ersten Blick erkennen kann was sich genau dahinter verbirgt. Sind diese aber mit der Publikation verknüpft, so können sie den Projekten, in denen sie entstanden sind, zugeordnet werden.

Ergänzende und geplante Serviceangebote

Neben den bereits erläuterten Angeboten im Bereich Forschungsdatenmanagement hat die Bibliothek noch ergänzende Angebote etabliert. So ist zum Beispiel in den letzten Jahren ein zunehmender Beratungsbedarf in diesem Bereich festzustellen. Die Anzahl der teilweise sehr komplexen Fragestellungen an die Bibliothek zu diesem Themengebiet steigt stetig. Hier wird daher entsprechende Hilfestellung und Unterstützung angeboten. Weiterhin hat sich die Bibliothek für das DKFZ als Datenzentrum für Digital Object Identifiers (DOI) registrieren lassen. Falls gewünscht, können die Forschungsdaten mit DOIs versehen werden.

Ein Nachteil des bisherigen Systems besteht darin, dass die Forschungsdaten, erst nachdem es eine Publikation gibt, entsprechend eingetragen werden können. Das führt dazu, dass sie nach der Veröffentlichung des Artikels unter Umständen erst mühsam zusammengesucht und in der Publikationsdatenbank eingetragen werden müssen, was einen zusätzlichen Aufwand für die Nutzer bedeutet. Viel sinnvoller wäre es daher, ein System anzubieten, in dem die Daten schon im Vorfeld verzeichnet werden können, so dass sie nach der Veröffentlichung der Publikation schon vorhanden sind und nur noch mit ihr verbunden werden müssen. Die bisherige Publikationsdatenbank bietet eine solche Möglichkeit nicht. Die Bibliothek arbeitet zurzeit jedoch an der Implementierung einer neuen Publikationsdatenbank-Software. Sie ist Projekt-Partner im JOIN²-Projekt⁸ (Just anOther INvenio INstance) geworden. Das System basiert auf der INVENIO-Software, die vom CERN entwickelt wurde. Sie ist flexibel einsetzbar und kann an die unterschiedlichsten Anforderungen angepasst werden. Geplant ist, das System so auszubauen, dass das Management von Forschungsdaten auch unabhängig von der Veröffentlichung einer Publikation möglich sein soll. Eine solche Funktion wäre auch für zwei weitere Projekte, die bei der Bibliothek angefragt wurden, relevant. Zum einen wird im DKFZ momentan über die Gründung eines wissenschaftlichen Archivs nachgedacht. In ihm sollen relevante Forschungsdaten, die nicht oder noch nicht publiziert wurden, archiviert und erschlossen werden. Weiterhin soll eine Sammlung von Informationen zu Wissenschaftlern des DKFZ aufgebaut werden. Sie soll unter anderem deren wissenschaftlichen Lebenslauf abbilden und Publikationslisten, Informationen zu Preisen, Fotos usw. enthalten. Beide Projekte könnten sinnvoll in die neue Publikationsdatenbanksoftware integriert werden und sollen von der Bibliothek realisiert werden.

Dies macht deutlich, dass Bibliotheken schon jetzt und auch zukünftig wichtige Aufgaben zur Unterstützung der Wissenschaft im Kontext Forschungsdatenmanagement bieten können.

⁸ <https://join2-wiki.gsi.de/cgi-bin/view>

Referenz

Sicherung guter wissenschaftlicher Praxis: Denkschrift. – Ergänzte Auflage. – Bonn: Deutsche Forschungsgemeinschaft, 2013, S. 21
http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf