Wolf Jürgen Richter

Auf dem Wege zum elektronischen Laborjournal – ein eSciDoc-Projekt

Laborjournale sollen alle Daten zur Planung und Durchführung von Experimenten festhalten; dazu werden diese datiert und mit fortlaufenden Chiffren versehen. Eingesetzte chemische Verbindungen werden in Formelsprache oder graphisch festgehalten. Die Versuchsergebnisse und die Art der Charakterisierung durch physikalisch-chemische Methoden müssen im Laborjournal zeitnah dokumentiert werden. Bedingt durch die Aufgabenstellung liegt eine heterogene Datenstruktur vor, die zudem gegen nachträgliche Datenmanipulation gesichert sein muss. Zwar dient ein Laborjournal in erster Linie als Tätigkeitsnachweis, es ist aber auch ein Dokument, das bei einem Prioritätsstreit – etwa bei Patentansprüchen – als Beweismittel herangezogen wird. So spielten bei Erfindung des Polyethylens durch Ziegler Laborjournale eine zentrale Rolle.

Der Übergang von *handgeschriebenen* zu *elektronischen Laborjournalen* stellt durch die Zusammenführung von Textbausteinen, chemischen Formeln und analytischen Daten und deren dauerhafter Konservierung eine nicht-triviale Programmieraufgabe dar. Während es in der chemischen Industrie seit längerem Praxis ist, Laborjournale nur noch in elektronischer Form zu führen, und die Software in der Regel als Firmeneigentum unzugänglich ist, stehen Hochschul- oder Max-Planck-Institute bei einer Eigenentwicklung eines geeigneten Software-Paketes vor einer hohen Hürde, wenn Produkte kommerzieller Anbieter als unzureichend befunden werden.

1. Einleitung

Das Kaiser-Wilhelm-Institut für Kohlenforschung (seit 1949 Max-Planck-Institut für Kohlenforschung) wurde bereits 1912 gegründet ist damit das drittälteste Institut der Kaiser-Wilhelm- / Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. In der langen Institutsgeschichten standen überwiegend experimentelle chemische Arbeiten, insbesondere die Untersuchung katalytischer Prozesse, im Mittelpunkt. Zwei epochale Erfindungen sind aus dem Institut hervorgegangen: Die Verflüssigung von Kohle, als Fischer-Tropsch-Verfahren in die

Literatur eingegangen, und die katalytische Herstellung von Polyethylen und Polypropylen bei Normaldruck durch Karl Ziegler und Mitarbeiter, für die Ziegler 1963 den Nobelpreis erhielt.

Das Herzstück der experimentellen Arbeit war und ist die Dokumentation der Versuche in Laborjournalen, die von 1914 bis heute lückenlos im Institut in handschriftlicher Form vorliegen.

2. Das „klassische" Laborjournal

Da das Laborjournal die Grundlage für spätere Publikationen oder mögliche Patentanmeldungen darstellt, müssen alle enthaltenen Angaben so genau und vollständig sein, dass ein Kollege oder ein externer Fachmann den beschriebenen Versuch mit gleichem Ergebnis reproduzieren kann. Dazu werden in der Regel folgende Daten festgehalten:

- Name des Experimentators und der Arbeitsgruppe (als Kürzel),
- Datum des Versuchsbeginns,
- Titel des Experiments – als Text oder Formelbild,
- ggf. ein Literaturzitat,
- chemische Ausgangsmaterialien,
- Reaktionsbedingungen,
- Produkte und Ausbeute,
- Art der Analysen,
- Paraphe des Gruppenleiters oder Vorgesetzten mit Datum.

Der Fachmann erkennt aus dem Reaktionsprotokoll, um welche Reaktion es sich handelt, in welchem Maßstab sie durchgeführt wurde, welche Komponenten in welcher Reihenfolge und in welchem Lösungsmittel zu Reaktion gebracht wurden und wie hoch die isolierte Ausbeute war; die Analysenmethoden zur Charakterisierung der Produkte sind als Kürzel (GC, MS, NMR, IR) angegeben. Ein Eintrag ist gleichzeitig Arbeitsnachweis und dauerhaftes Dokument. Das manuelle Führen eines Laborjournals hat sich über viele Jahrzehnte bewährt, kann es doch direkt am Arbeitsplatz von den Mitarbeitern protokolliert werden; dadurch wird es generell bis heute auch als Arbeitsmittel akzeptiert. Zudem ist ein so geführtes Laborjournal rechtsbeständig. Als Besonderheit gilt noch im amerikanischen Patentrecht das Prinzip „first to invent", also das Festhalten im Laborjournal, während in anderen Ländern „first to file", also der Zeitpunkt der späteren Veröffentlichung, zählt. Das unterstreicht deren Wichtigkeit bei einer Auseinandersetzung.

Laborjournale waren im Prioritätsstreit bei der Erfindung des Polyethylens durch Karl Ziegler am Max-Planck-Institut für Kohlenforschung von großer Be-

deutung, wie in dem Buch des Ziegler-Mitarbeiters Heinz Martin ausführlich dokumentiert.[1]

Lange diente im Institut neben dem Laborjournal ein handschriftlich ausgefüllter Vordruck mit Sigel und Versuchsnummer des Auftraggebers (Auftragszettel) als „Schnittstelle“ zu den analytischen Abteilungen. Von dort kamen die Analyseergebnisse als Ausdruck zurück. Das Expertenwissen der Mitarbeiter in den analytischen Abteilungen findet seinen Niederschlag in der qualifizierten Datenanalyse, darüber hinaus in verstärktem Maße in Aufbau und Pflege von methodenspezifischen hauseigenen Datenbanken. Die analytischen Ergebnisse werden dem Auftraggeber in der Regel nicht als Primärdaten, sondern als bearbeitete Daten an die Hand gegeben.

Testweise wurden innerhalb des Max-Planck-Instituts für Kohlenforschung für die Mitarbeiter elektronische Auftragszettel eingeführt; das Versuchssigel (zum Beispiel ACK-AK-028-01 für den Mitarbeiter Lutz Ackermann) ist dabei der Verknüpfungspunkt zwischen Experimentator und den analytischen Abteilungen. Die Ergebnisse waren so erfolgversprechend, dass das Verfahren als verbindlich eingeführt wurde – keine Analyse mehr ohne elektronischen Auftragszettel!

Konsequenterweise wurde dann die Erfassung sämtlicher Labordaten in elektronischer Form und deren Verknüpfung mit der Analytik von der EDV-Abteilung des Max-Planck-Institut für Kohlenforschung als Großprojekt in Angriff genommen – also das elektronische Laborjournal.

Wieweit Nutzer von der Ebene des elektronischen Laborjournals auf die analytischen Primärdaten zugreifen dürfen oder sollen, wird sicher nicht nur von den Möglichkeiten der Programme und der Kompetenz der Mitarbeiter abhängen, sondern vor allem von der Struktur der Arbeitsgruppen oder Abteilungen und wird durch hierarchische Zugriffsrechte geregelt.

3. Das elektronische Laborjournal

Die Vorteile eines elektronischen Laborjournals im Vergleich zum handgeschriebenen liegen auf der Hand: Die Daten aus unterschiedlichen Feldern werden automatisch in einer relationalen Datenbank wie Oracle® gespeichert, sie lassen sich beliebig zusammenstellen und unter unterschiedlichsten Gesichtspunkten analysieren. Alle durchgeführten Versuche, auch die anderer Experimentatoren,

1 Martin, H., Polymere und Patente – Karl Ziegler, das Team, 1953–1998. Weinheim: Wiley-VCH 2001.

sind leicht zugänglich und erlaubten in der Zusammenschau neue Fragestellungen wie:

- Welche Komponenten wurden bereits eingesetzt und welche noch nicht?
- Unter welchen Reaktionsbedingungen wurden schon getestet?
- Welche Ausbeute und welche Analysen liegen vor?
- Sind biologische Tests in anderen Labors durchgeführt worden?

Mit diesem Werkzeug können Versuchswiederholungen vermieden, Optimierungen effektiv geplant und daraus neue Erkenntnisse gewonnen werden. Auch Literaturzitate und sogar Originalarbeiten lassen sich (bei entsprechender Lizenz) mit dem e-Journal verlinken.

Als Besonderheit stehen zum Zeichnen von chemischen Strukturformeln im elektronischen Laborjournal verschiedene Tools zur Verfügung (zum Beispiel ChemDraw®, ISIS Draw®), die dem Experimentator das bisweilen mühevolle manuelle Zeichnen von Molekülen durch vorgefertigte Templates ersetzen: Einmal ins e-Lab eingegeben vereinfacht sich das wiederholte Zeichnen von Molekülen außerordentlich. Auch eröffnen sich mit elektronisch gespeicherten Strukturformeln neue Such- und Verknüpfungsmöglichkeiten, denn nur mit geringem Zusatzaufwand kann aus den Formeln eine in-house-Datenbank generiert werden. Zudem ist durch Datentransfer Substruktur- und Ähnlichkeitssuchen in externen kommerziellen Datenbanken und Katalogen möglich (zum Beispiel im *Registry File* der Chemical Abstracts®, in der Beilstein-Datenbank, im Aldrich® Katalog). Neben diesen lizensierten Programmen hat die IUPAC (*International Union of Pure and Applied Chemistry*) mittlerweile (2005) IncChi® als ein Open-Source-Produkt zum Zeichnen und Speichern chemischer Strukturen entwickelt, das zum Beispiel vom *National Institute of Health* als PubChem-Datenbank mit vier Millionen Einträgen genutzt wird.[2]

Nach jahrelanger firmeneigener Softwareentwicklung und Erfahrung mit unterschiedlicher Protokollierung ist die Führung von elektronischen Laborjournalen mittlerweile bei großen chemischen Firmen wie BAYER oder Degussa obligatorisch.

Bei der Führung dieser e-Laborjournals sind bestimmte Regeln vorgeschrieben und Standards vorgeben; autorisierte Personen können je nach Aufgabenstellung Daten mitbenutzen und e-Labs anderer lesen, wobei sie aber einem besonderen Datenschutz unterliegen. Die Vorteile, Daten aus unterschiedlichen Sparten – etwa aus Biologie und Anwendungstechnik – zusammenführen zu können, vermeiden Doppelarbeit, und machen den relativ hohen technischen und finanziellen Aufwand eines e-Laborjournals bei Weitem wett. Der Wermutstropfen für

2 http://www.iupac.org/inchi

Externe: Software oder in-house Datenbanken sind in der Regel aus Firmeninteresse unzugänglich.

4. Das Projekt: Elektronisches Laborjournal am Max-Planck-Institut für Kohlenforschung

Im Jahre 2002 wurde das e-Lab-Projekt erneut im Max-Planck-Institut für Kohlenforschung angegangen; da aus angeführten Gründen bewährte Produkte aus der chemische Industrie ausschieden, wurden verschiedene kommerzielle Anbieter von elektronischen Laborjournalen zu Gesprächen und Präsentationen ins Institut eingeladen. Im Vordergrund stand die Prüfung auf Kompatibilität mit den vorhandenen Daten aus den analytischen Abteilungen und dem gesamten Arbeitsablauf; zudem sollten die experimentell arbeitenden Gruppen ihre Erwartungen an ein *e-Lab Journal* artikulieren können. Die Datensicherheit der Einträge spielte nach den Erfahrungen der Vergangenheit ebenfalls eine wichtige Rolle.

Die nachfolgenden Anbieter, die mit einer Repräsentanz in Deutschland vertreten waren, wurden zu Demonstrationen eingeladen: Das „Hybrid Electronic Notebook“ ELAN von MDL Information Systems®;[3] Adept Scientific® Notebook;[4] CambridgeSoft®[5] und das *Dokumenten Management System* der Firma WinDream GmbH.[6]

Als aktuelle Ergänzung sollen noch neuere Produkte wie die eLabs des französischen Software-Produzenten Klee[7] und das Data-Sharing-Projekt der Universität Sheffield „SmartTea“[8] erwähnt werden, die erst nach Beendigung der Testphase auf den Markt kamen und nicht untersucht wurden. Auch ein *Open Source Electronic Laboratory Notebook* steht inzwischen zur Verfügung (Stand 2006).[9]

Das e-Labourjournal ELAN von MDL beispielsweise weist eine Oberfläche wie ein vertrautes handgeschriebenes Laborjournal auf: Überschrift, chemische Formelbilder, Ansatzgröße und Analytik usw. Eine Umstellung würde den Mitarbeitern demnach leicht fallen, zumal unser Institut mit dem elektronischen Strukturerfassungs- und Suchprogramm REACCS von MDL langjährige Erfahrung hat – es dient unter anderen als Grundlage zum erfolgreichen Aufbau der

3 http://www.mdl.com/products/experiment/elan/index.jsp
4 http://www.adeptscience.co.uk/tcm/tcm22/elab.html
5 http://products.cambridgesoft.com/software/details/?ds=9
6 http://www.windream.com/cgi-bin/winmain.ais?PAGE=de110
7 http://www.kalabie.com
8 http://mytea.uk.org
9 http://collaboratory.emsl.pnl.gov

institutsinternen chemischen BEStell- und SIcherheitsdatenbank BESSI, die jeder Chemiker benutzen muss, wenn er Chemikalien aus dem Lager bestellen will.

Zudem setzen die meisten untersuchten *e-Lab Journals* auf dem weitverbreiteten MS Windows® auf, und erreichen damit eine schon vertraute Benutzerführung und größere Akzeptanz beim Benutzer.

Unterschiedliche Erwägungen spielten bei der Eignungsprüfung eine Rolle:

- Einbindung in die vorhandene Infrastruktur,
- Möglichkeiten der Eigen- oder Fremdprogrammierung,
- Berücksichtigung von Open-Source-Protokollen für Strukturformeln,
- Abhängigkeit vom Software-Hersteller,
- Erfahrungen mit chemischen Struktur-Editoren,
- Kosten per Arbeitsplatz,
- Datensicherheit.

Auf eine detaillierte Bewertung der getesteten *e-Lab Journals* soll hier nicht eingegangen werden, jedoch blieben bei den Workshop-Teilnehmern bei allen Produkten Fragen und Wünsche offen, so dass am Ende 2002 der Ruf nach Eigenentwicklung eines e-Labs im Institut verstärkt wurde. Zu diesem Zeitpunkt begann die Max-Planck-Gesellschaft im Rahmen des Heinz-Nixdorf-Zentrums für Informationsmanagement umfassende EDV-Aktivitäten zu entwickeln.

5. Bedarfanalyse durch das Heinz-Nixdorf-Zentrum für Informationsmanagement – ZIM

Nach Gründung der ZIM, des Heinz-Nixdorf-Zentrums für Informationsmanagement, unter dem Dach der Max-Planck-Gesellschaft wurden unterschiedliche Aspekte eines zentralen EDV-Supports auf potentiellen Bedarf und Anwendbarkeit innerhalb der Institute der Max-Planck-Gesellschaft untersucht. Nach angelsächsischem Vorbild wurde ein zwischen den Instituten nutzbarer „Scholarly Communication Service“ angeboten. Im vorbereitenden Schritt wurden Bedarfsanalysen an ausgewählten Instituten aus den drei Sektionen der Max-Planck-Gesellschaft durchgeführt.[10]

Im Rahmen eines solchen Workshops stand nach etlichen Vorbesprechungen das Projekt „Elektronisches Laborjournal – eLab“ als Teil des Gesamtprojekts für zwei Tage im Februar 2005 im Max-Planck-Instituts für Kohlenforschung im Mittelpunkt. Neben Mitarbeitern des ZIM nahmen auch Mitarbeiter des FIZ Karlsruhe als Projektbeteiligte teil.

10 http://www.zim.mpg.de

Folgende Anforderungen an ein e-Lab wurden im Protokoll festgehalten (hier leicht verkürzt):

Anforderungen an das Interface:

- chemischer Formeleditor, stöchiometrische Berechnungen
- Link zu den analytischen Daten
- Interface zum „Auftragszettel“
- Interface zu einem Word-Processor
- Möglichkeit zum Datenaustausch mit anderen Max-Planck-Instituten.

Eine Suche soll in folgenden Feldern möglich sein:

- Inhaltsverzeichnisse,
- Strukturen,
- Versuchsbeschreibung,
- Metadaten und Schlagworte.

Daraus entwickelte sich eine gemeinsame Vision für das Projekt: die e-Labs sind genauso einfach innerhalb des Arbeitsablaufs der Forschungsgruppen zu integrieren, wie es die statischen handgeschriebenen waren; die Benutzer identifizieren sich durch Smartcards; diese sind eng mit den analytischen Daten verknüpft. Das zukünftige Lab-Journal ist im Einklang mit den Erfordernissen des internationalen Patentrechts.

6. Kaufen oder selber machen?

Für kommerzielle Systeme gilt generell:

- teuer durch die Lizenzgebühren und notwendige Anpassungen (–)
- wenige Open-Source-Produkte (–)
- nicht flexibel für alle potenziellen chemischen Nutzer (–)
- gute Benutzeroberfläche, große Nutzererfahrung (+)
- Weiterentwicklung durch Firmen zu erwarten (+)
- hohe Servicekosten (–).

Eigene Programmentwicklung hat gleichermaßen Vor- und Nachteile:

- optimale Anpassung an die Bedürfnisse des Instituts (+)
- verallgemeinerbare Lösung für andere Max-Planck-Institute (+)
- Interfaces müssen entwickelt werden, Anleihen bei kommerziellen Anbietern denkbar
- hohe Entwicklungskosten (–)
- zeitaufwendig (–).

Das Projekt *e-Labs* wird somit als vierte Säule neben

- *Scholarly Workbench* – neue Art der wissenschaftlichen Kommunikation,

- *Publication Management* – integrierte Selbstarchivierung und Open-Access-Lizensierung und
- *eLib* – Aufbau einer lokalen *High-Quality*-Datensammlung

in das Gesamtkonzept von „Enhanced Science Documentation" (eSciDoc) integriert. eSciDoc als Nachfolgeprojekt für ZIM ist eine neuartige Plattform für vernetztes wissenschaftliches Arbeiten, das im Rahmen einer nationalen Initiative vom Bundesministerium für Bildung und Forschung über einen Zeitraum von fünf Jahren mit erheblichem Mitteleinsatz gefördert wird. Jede der vier Säulen wird von verschiedenen Max-Planck-Instituten als Piloten getestet, bevor eine Übertragung auf andere Institute oder die gesamte Max-Planck-Gesellschaft erfolgt.[11]

Die Verbindung von institutsinternem Know-how mit externen Programmierern, die durch das eSciDoc-Projekt finanziert werden, bietet sich als erstrebenswerte Lösung an, wobei ein fester Zeithorizont (Mitte 2009) ins Auge gefasst wird. Vom Ergebnis eines *e-Lab Journal* auf Open-Source-Basis könnten viele andere Max-Planck-Institute durch eine Public-Key-Infrastruktur profitieren.

Ein aktueller Beitrag zum Thema *eLab Notebook* findet sich in *Nature* unter dem Titel "A New Leaf".[12]

11 http://www.escidoc-project.de/fileadmin/be_user/downloads/prs_Beiratssitzung_130505_final_public.pdf

12 Butler, D., Electronic notebooks: A New Leaf. – In: Nature. 436(7 July 2005), S. 20 – 21.

Gesellschaft für
Wissenschaftsforschung

GEWIF
BERLIN

Frank Havemann
Heinrich Parthey
Walther Umstätter
(Hrsg.)

Integrität wissenschaftlicher Publikationen in der Digitalen Bibliothek

Wissenschaftsforschung
Jahrbuch 2007

Mit Beiträgen von:
Bettina Berendt • Stefan Gradmann
Frank Havemann • Andrea Kaufmann
Philipp Mayr • Heinrich Parthey
Wolf Jürgen Richter • Peter Schirmbacher
Uta Siebeky • Walther Umstätter
Rubina Vock

Wissenschaftsforschung
Jahrbuch **2007**

Integrität wissenschaftlicher Publikationen in der Digitalen Bibliothek: Wissenschaftsforschung Jahrbuch 2007 / Frank Havemann, Heinrich Parthey u. Walther Umstätter (Hrsg.). Mit Beiträgen von Bettina Berendt... – Berlin: Gesellschaft für Wissenschaftsforschung 2007.

Bibliographische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Dieses Buch ist unter einer Creativ-Commons-Lizenz lizenziert. Sie dürfen für nichtkommerzielle Zwecke das Werk und Teile davon vervielfältigen, verbreiten und öffentlich zugänglich machen, wenn Sie auf die Urheber (Autoren, Herausgeber) und den Verlag verweisen. Im Falle einer Verbreitung müssen Sie anderen die Lizenzbedingungen, unter welche dieses Werk fällt, mitteilen.

Das Werk ist in allen seinen Teilen urheberrechtlich geschützt.

Jede kommerzielle Verwertung ohne schriftliche Genehmigung des Verlages ist unzulässig. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in Systeme(n) der elektronischen Datenverarbeitung.

Gesellschaft für Wissenschaftsforschung
1. Auflage 2007

Verlag: Gesellschaft für Wissenschaftsforschung
c/o Institut für Bibliotheks- u. Informationswissenschaft
der Humboldt-Universität zu Berlin,
Unter den Linden 6, D–10099 Berlin
verlag@wissenschaftsforschung.de

Druck: BoD Norderstedt

ISBN 3-924682-43-x