
Erfassung und Speicherung von Forschungsdaten im Fachbereich Chemie: Bereitstellung moderner Forschungs-Infrastrukturen durch ein elektronisches Laborjournal mit Repositorium-Anbindung

Nicole Jung¹, Pierre Tremouilhac², Claudia Kramer³, Jan Potthoff⁴

1 Institut für Organische Chemie, Institut für Toxikologie und Genetik, KIT

2 Institut für Toxikologie und Genetik, KIT

3 KIT-Bibliothek

4 Steinbuch Centre for Computing, KIT

Zusammenfassung. Eine moderne Infrastruktur zur Erfassung, Bearbeitung und Speicherung von Forschungsdaten bildet die Voraussetzung für eine strukturierte Dokumentation der Forschung zur effizienten Nachnutzung der Ergebnisse und der Daten. Für den Fachbereich Chemie wurden zwei Komponenten (elektronisches Laborjournal [ELN] und Repositorium) identifiziert, durch deren Entwicklung und Vernetzung Forschungsdaten über das bisher mögliche Verfahren hinaus bewahrt und nutzbar gemacht werden sollen. Beide Komponenten des Chemotion-Projektes werden am KIT programmiert und als Open Source Infrastruktur anderen Forschern zur Verfügung gestellt. Die Entwicklung des Laborjournals ist zunächst auf die Bedürfnisse organischer Chemiker ausgelegt und umfasst in der aktuellen Version umfassende Möglichkeiten für eine moderne Arbeitsweise im synthetischen Labor. Neben den Basisfunktionen wurden verschiedene Elemente eingebaut, die über die Dokumentation von Forschungsergebnissen hinaus nachhaltiges Arbeiten und eine eindeutige, korrekte Strukturierung wissenschaftlicher Arbeit unterstützen. Durch die Beteiligung des Rechenzentrums werden Strukturen geschaffen, die eine parallele, nachhaltige Sicherung der Rohdaten zu den jeweiligen Experimenten erlauben. Durch die Kombination des ELNs mit einem Repositorium, dessen Schnittstelle die Datenstruktur der ELN Inhalte aufgreift, wird es dem Forscher ermöglicht, Daten ohne weitere Vorbereitung auf direktem Weg anderen Forschern zur Verfügung zu stellen. Durch die parallele Entwicklung des ELN und des Repositoriums werden mehrere Aspekte des Forschungsdatenmanagements adressiert. Darüber hinaus ermöglicht dies eine effiziente Suche fachspezifischer Forschungsdaten. Das neue Modell aus Kombination von ELN und Repositorium sieht eine Kombination aus Erfassung und Dokumentation sowie Publikation und Speicherung von Datensätzen vor. Die Indexierung, welche aktuell erst nach der Publikation von Daten über Journale oder Patente stattfindet (z.B. über den Chemical Abstracts Service, CAS), soll schon direkt nach der Entstehung vorgenommen werden können. Damit werden bereits die Rohdaten mit semantischen Informationen sowie Metadaten versehen, die insbesondere bei der Verwendung des OAIS-Referenzmodells (Open Archival Information System) von entscheidender Bedeutung für eine nutzerorientierte Verwendung der archivierten Daten sind. So kann in allen Publikationsprozessen auf eine einheitliche, eindeutige Quelle verwiesen werden. Der freie Zugang und die kostenfreie Nutzung für akademische Forscher sollen in Zukunft die Grundlage für gemeinsames Arbeiten bilden, eine Bereitstellung der Forschungsdaten ermöglichen und den Austausch unter den Wissenschaftlern fördern.

Schlagwörter. elektronisches Laborbuch, Repositorium, Datenmanagement, Chemie, Synthese

Einleitung

Im Fachbereich Chemie besteht, wie in anderen naturwissenschaftlichen Disziplinen, bisher ein Mangel an frei nutzbarer, fachspezifischer Software für einen zeitgemäßen Umgang mit Forschungsdaten (Winkler-Nees 2013). Hierdurch wird besonders die Bereitstellung von Forschungsinformationen erschwert. In den letzten 15 Jahren konnte der Zugang zu publizierten Forschungsergebnissen stark verbessert werden. Dies liegt vor allen Dingen an der Verfügbarkeit von Publikationen in online-Versionen der wissenschaftlichen Journale und der web-basierten Bereitstellung von Informationen durch Datenbanken wie SciFinder oder Reaxys (Beilstein).^{1 2} Während diese Entwicklungen die Suche nach veröffentlichten Informationen erleichtert haben, fehlen Lösungen für eine umfassende digitale Verfügbarkeit anderer Forschungsdaten. Ein Grund dafür sind die anspruchsvollen Anforderungen an die Forschungsinfrastruktur und Software, da abhängig vom Forschungsgebiet mehrere Aspekte der Laborarbeit in den elektronischen Datenerfassungs- und Lagersystemen abgebildet werden müssen. Besondere Herausforderungen müssen im Forschungsbereich Chemie überwunden werden, da die Zeichnung und Verarbeitung von chemischen Strukturen ein entscheidender Schritt für die Korrelation von Forschungsdaten mit der entsprechenden chemischen Umsetzung oder Struktur ist (Coles et al. 2013). Während in den letzten Jahren mehrere ELN (wie SciNote, Biovia ELN, EMEN, Open BIS-ELN LIMS, LabFolder)^{3 4 5} (Rees et al. 2013; Barillari et al. 2016; Rubacha et al. 2011) entwickelt wurden, die intelligente Lösungen für die Dokumentation von Forschungsdaten bieten, sind im Bereich chemische Forschung nur wenige Produkte verfügbar. Beispiele für geeignete Systeme in der Chemie sind das PerkinElmer E-Notebook für Chemie, Indigo-ELN oder LabTrove - und Open Enventory, von denen nur die letzteren drei Produkte kostenlos zur Verfügung stehen (Day et al. 2015; Willoughby et al. 2014; Milsted et al. 2013; Rudolphi & Goosen 2012)^{6 7}. Für einen Einsatz in der Hochschule, der Flexibilität der Software-Struktur und möglichst kostenlose Verfügbarkeit erfordert, sollten entsprechende elektronische Journale vorzugsweise zusätzlich als Open Source zur Verfügung stehen. Die derzeit verfügbaren Systeme wurden von unserer Projektgruppe auf ihre Verwendung als modernes, flexibles Managementsystem für Chemieforschungsdaten untersucht. Die identifizierten Anforderungen an ein System, das den Herausforderungen des zukünftigen Datenmanagements genügt, wurden jedoch nicht erfüllt. Auch steht bisher kein Repositorium im Forschungsbereich Chemie zur Verfügung, welches als Infrastruktur die Speicherung und Bereitstellung von Daten auf einfachem Wege ermöglicht. Forschungsdaten, die für sich alleine genommen keine Publikation rechtfertigen, können bisher in Web-Portalen wie ChemSpider (Kelly & Kidd 2015) (Synthetic Pages) oder SDBS (AIST⁸) abgefragt werden. Allerdings bieten diese Portale meist keine Möglichkeit zur Bereitstellung eigener Daten (SDBS). Wenn dies der Fall ist, erschwert eine zeitaufwändige Eingabe von Forschungsdaten die Pflege der Portale mit aktuellen Daten und lässt eine kontinuierliche Beteiligung der Forscher unattraktiv werden. Weitere Schwierigkeiten bisheriger Repositorien sind die nur begrenzt mögliche Beschreibung von Reak-

1 <http://www.cas.org/products/scifinder>

2 <https://www.reaxys.com/reaxys/secured/search.do>

3 <https://www.labfolder.com/>

4 <http://accelrys.com/products/unified-lab-management/biovia-electronic-lab-notebooks/>

5 <https://github.com/biosistemika/scinote-web>

6 https://www.cambridgesoft.com/Ensemble_for_Chemistry/ENotebookforChemistry/

7 <https://github.com/ggasoftware/indigo>

8 <http://www.aist.go.jp>

tionen, die geringe Flexibilität der unterstützten Datenformate für experimentelle Daten (meist lediglich Bildformate) und das Fehlen von Schnittstellen zu elektronischen Laborjournalen. Da gerade letzterer Punkt zu einer nur sehr geringen Beteiligung bei der Eingabe von Daten in Repositorien führt, sollte eine moderne Infrastruktur zur Speicherung und Offenlegung von Forschungsinformationen den Transfer von Daten aus einem ELN hin zu einem offenen Repository, ebenso wie von einem Repository hin zu einem ELN ermöglichen. Die Nutzung einer Kombination aus elektronischem Laborjournal und Repository (dargestellt in Abbildung 1) wurde unseres Wissens in den chemischen Wissenschaften noch nicht bereitgestellt.

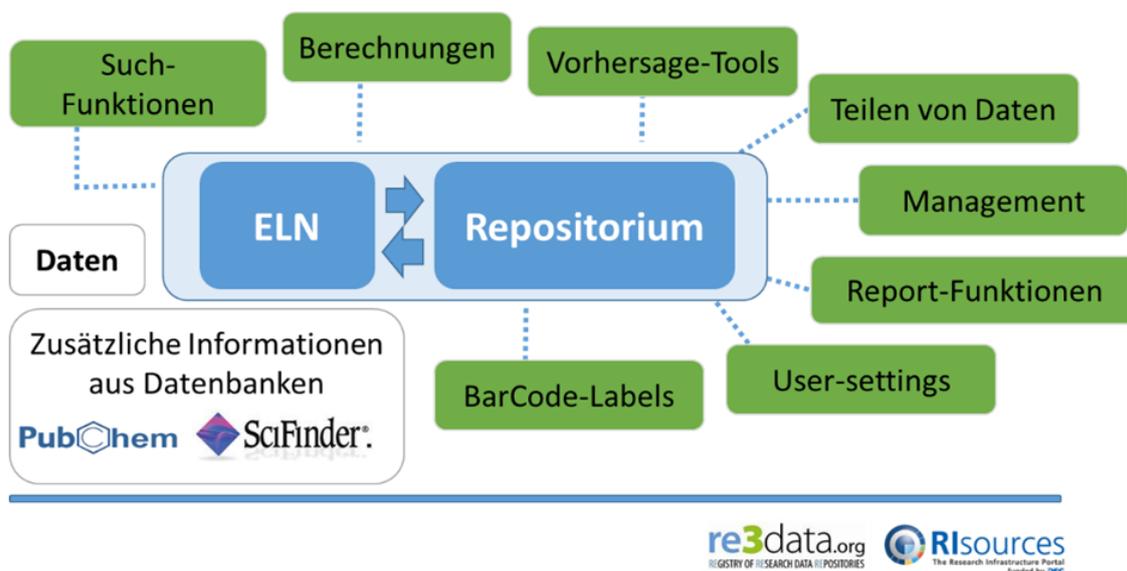


Abbildung 1. Gemeinsame Funktionen von ELN und Repository zum Austausch und zur Speicherung von Forschungsdaten.

Chemotion-ELN zur Speicherung digitaler Forschungsdaten

Generische Funktionen und Management

Das ELN der Chemotion-Projektgruppe bietet Grundfunktionen für die forschungsspezifische Arbeit und wird durch generische Werkzeuge für Projekt- und Datenmanagement ergänzt. Diese ermöglichen die Erstellung einer klaren Struktur zur Speicherung und Suche von Forschungsdaten. Das Management von Projekten und die Sortierung einzelner Elemente wird durch die Generierung von sogenannten Collections (oder Sammlungen) unterstützt, die durch einen separaten Organizer generiert, bearbeitet und gelöscht werden können. Die Zuordnung von ausgewählten Elementen zu den gewünschten Sammlungen oder Untersammlungen kann auf einfache und schnelle Weise per Drag & Drop durchgeführt werden. Einmal getroffene Entscheidungen über die Konfiguration der ELN-Struktur können jederzeit problemlos modifiziert werden, sodass das ELN flexibel auf Veränderungen der Forschungsprojekte ausgerichtet werden kann. Der Nutzer des ELNs kann durch das Management-System nicht nur die Daten verwalten, die in eigenen Sammlungen strukturiert sind, es können ebenfalls Projekte verwaltet werden, die durch Zusammenarbeit mit anderen Wissenschaftlern entstehen. Neben der Organisation der Projekte durch

Sammlungen, können einzelne Elemente des ELNs (z.B. Moleküle, Reaktionen, Wellplatten, Experimente oder Vorhaben) in getrennten Listen sortiert und verwaltet werden. Die Einträge innerhalb der Listen sind durch zusätzliche Informationen über die Präsenz der Elemente in den Sammlungen oder aber auch anderen Datenbanken versehen. Bei Auswahl des gewünschten Elements wird eine zweite Informationsebene einsehbar, die durch weitere Details die Informationen der Listen ergänzt. Eine schnelle Organisation der Forschungsdaten wird durch diverse Funktionen wie Drag & Drop, automatisierte Sortierung von Elementen, Farbkennzeichnung für Warn- und Erinnerungsfunktionen erleichtert. Alle Elemente wie Proben und Reaktionen, die bestimmten Sammlungen zugeordnet und in Listen zusammengefasst wurden, können nach ihrer Zuordnung organisiert werden, um eine flexible Gesamtstruktur des ELNs zu ermöglichen. Das ELN wurde als Webanwendung realisiert, so dass die Installation weiterer Software nicht notwendig ist. Zur Nutzung wird lediglich ein aktueller Webbrowser benötigt.

Besondere Qualitätssicherungsmodule

Das Chemotion ELN kann für eine exakte Nachverfolgung von einzelnen Elementen verwendet werden. Dies setzt die systematische und automatische Nummerierung aller Einträge einschließlich einer intuitiven Zuordnung zum jeweiligen Workflow voraus. Entsprechend dieser Bedingung tragen ELN-Einträge, die Teil eines Prozesses sind, Informationen über ihren Ursprung innerhalb des Namens. Einträge, die neu erstellt wurden oder die über eine Kopie erzeugt wurden, erhalten jeweils einen neuen Namen, der aus den Initialen des ELN-Benutzers und einer sequentiellen Nummernfolge besteht. Es ist möglich, Sub-Elemente aus Elementen herzustellen, welche als Child-Item betrachtet werden und durch die Anbringung einer speziellen Chargennummer vom Original unterscheidbar sind. Das ELN ist so konzipiert, dass eine flexible Dokumentation des Forschungsprozesses möglich wird. Gleichzeitig wird jedoch auch eine Manipulation von Werten verhindert. Während alle Parameter einer Reaktion protokolliert und entweder über vordefinierte Felder oder freie Textfelder eingegeben werden können, gibt es weitere Felder, in denen berechnete Daten nur sichtbar, aber nicht editierbar sind. Ein Beispiel für eine solche Begrenzung ist die Eingabe der Ausbeute für Reaktionen. Die Möglichkeit, einen Wert für die Ausbeute einer Reaktion hinzuzufügen, ist für alle Anwendungen deaktiviert, da die Ausbeute jeweils das Ergebnis aus der gewonnenen Menge ist und niemals als Zielwert eingegeben wird. Ein besonderes Merkmal des Chemotion ELN ist die Möglichkeit zur Aufzeichnung von realen Werten parallel zu jenen des ursprünglich geplanten Experiments. Dies ermöglicht die genaue Dokumentation des realen Experiments, während die Information der geplanten Prozedur zusätzlich gespeichert wird.

Einbindung von Informationen aus externen Datenbanken

Informationen aus externen Datenbanken ermöglichen eine objektive und schnelle Einschätzung der eigenen Forschung und wurden daher als wichtiges Element dem ELN beigelegt. In der aktuellen Version unterstützt das Chemotion-ELN die Abfrage von Daten aus der Datenbank SciFinder (kommerziell, Lizenz notwendig) und PubChem. Die Einbindung der SciFinder-Suche wurde durch ein PlugIn realisiert und unterstützt die Suche in der CAS SciFinder Datenbank nach drei verschiedenen Suchmodi. Die Anzahl der Suchergebnisse wird mit einem Link zur Antwort in der

SciFinder-Webanwendung abgerufen. Die direkte Sichtbarkeit der veröffentlichten Strukturen über die ELN ermöglicht einen schnellen Zugriff auf Informationen. Um einen umfassenden Überblick über die Neuheit der Forscherarbeit und die Verfügbarkeit von Forschungsdaten zu geben, wurde zusätzlich ein automatisiertes Verfahren zur Abfrage von gegebenenfalls vorhandenen PubChem-Einträgen implementiert. Wie für die eingebettete SciFinder-Funktion gegeben, sind die passenden Treffer über einen direkten Link zum PubChem Index des identifizierten Elementes zugänglich.

Fachspezifische Funktionen

Das ELN bietet neben der allgemeinen Ausstattung der Software alle notwendigen Funktionen für die Dokumentation und Bearbeitung von chemischen Projekten, einschließlich der Verarbeitung von Molekülen und Reaktionen. Hierzu wurde basierend auf dem als Open Source verfügbaren Struktur Editor Ketcher ein flexibel nutzbarer, fortgeschrittener Struktureditor erstellt. Eingebettet in das ELN und unterstützt durch eine Verwertung der generierten Information innerhalb des ELNs können zusätzliche Funktionen wie die Eingabe von Templates oder die automatische Berechnung von Strukturmerkmalen eingebunden werden. Die interne Struktur des ELNs folgt strengen Regeln bei der Schaffung neuer Elemente, die die Differenzierung zwischen Molekülen und Einzelproben bewirken. Während Moleküle jeweils mit den Merkmalen der chemischen Struktur und den hieraus errechenbaren Werten assoziiert werden (verfügbar durch z.B. OpenBabel), erlauben Einzelproben ebenfalls die Speicherung zusätzlicher Eigenschaften, die durch die Umgebung oder Anwendungsform (Reinheit, Zusammensetzung, Temperatur) bestimmt werden. Die Registrierung und konsequente Verwendung von Molekülen oder Einzelproben während der Arbeit mit dem ELN ist die Basis für eine gut organisierte und am Ende reproduzierbare synthetische Dokumentation. Während eine solche klare Differenzierung zwischen Molekülen und Proben in den meisten anderen Chemiejournalen nicht angeboten wird, stellt das Chemotion ELN dieses wichtige Merkmal in einer sehr einfachen Weise vor. Die Abbildung einer physikalisch genutzten Substanz oder ihrer Zubereitung umfasst die Zusammenfassung der verfügbaren Daten aus dem verwandten Molekül, die eine schnelle Verfügbarkeit aller Informationen ermöglichen, die für ein schnelles Management des Forschungsprojekts notwendig sind. Die automatisch bereitgestellten Daten sowie die vom Benutzer vorgegebene Eingabe sind in fünf sogenannten Sample-Tabs organisiert. Sie bestehen aus (1) Informationen für eine detaillierte Definition der Eigenschaften, (2) zusätzliche Daten, die an die hochgeladenen Dateien mit Forschungsdaten angehängt werden können, (3) Ergebnissen, die mit der Probe durch einen externen Prozess erhalten wurden, (4) einer automatisierten Online-Anfrage der SciFinder-Datenbank und einer direkten Verbindung zu den Suchergebnissen sowie (5) vorhergesagten NMR-Informationen.

Neben Einzelproben und Molekülen gehören Reaktionen zu den Hauptelementen, die durch das ELN erzeugt und verwaltet werden sollen. Reaktionen können sehr schnell erzeugt werden und durch Hinzufügen von Einzelproben in Funktion von Ausgangsmaterial, Reagenz oder Produkt definiert werden. Das Grundschema für Proben in Reaktionen erlaubt die Zugabe der Menge der Substanzen in verschiedenen Einheiten und die Definition der verwendeten Substanz in Äquivalenten. Aus den gegebenen Werten werden die fehlenden Informationen für eine Reaktion automatisch berechnet. Die Struktur der Reaktions-Benutzeroberfläche ist sehr flexibel, so dass eine Änderung der Zuordnung der Einzelelemente jederzeit möglich ist. Alle Abhängigkeiten werden

kontinuierlich an die Änderungen angepasst. Die Chemikalien, die der Reaktion zugeordnet sind, sind über eine direkte Verbindung zur Detailstufe der Einzelproben zugänglich und alle Daten und Änderungen, die den Proben (z.B. Dichte der Chemikalien) zugeordnet werden fließen sofort in die Berechnung der Reaktion ein. Während das Reaktionsschema und die Reaktionstabelle durch zusätzliche Informationen in Freitext-Form vervollständigt werden können, wird das Hinzufügen einer Reaktionsbeschreibung durch mehrere formatierte Vorlagen unterstützt, die für einen schnellen Bericht über ein chemisches Verfahren in einer standardisierten Weise verwendet werden könnten. Zusätzlich wurden drei weitere Eingabeseiten „Eigenschaften“, „Literatur“ und „Analyse“ für die Bereitstellung weiterer Informationen geschaffen.

Austausch und Freigabe von Daten

Um einen möglichst einfachen Austausch von Forschungsinformation zu ermöglichen, werden innerhalb des ELNs Export- und Importfunktionen für einzelne Proben, Reaktionen und Sammlungen erstellt. Verfügbare Dateiformate für diese Anwendungen sind Excel (.xlsx) und SDF (.sdf). Die Details des Exportvorgangs können vom ELN-Benutzer über eine Auswahl der Spalten bestimmt werden, die in der exportierten Datei angegeben werden sollen. Neben dem Import und Export wurden zusätzlich zwei Funktionen („Teilen“ und „Synchronisieren“) implementiert, um Informationen mit anderen Forschern teilen zu können. Die Benutzeroberfläche erlaubt die Organisation einzelner Gruppen nach ihrem Status und nach gewünschten Rechten. Der Benutzer des ELNs und Eigentümer der übermittelten Daten kann dazu eigene Rollen definieren und Gruppen zuweisen oder eine vordefinierte Rolle verwenden. Die Definition von Freigabeberechtigungen umfasst die Auswahl einer Berechtigungsstufe für zulässige Aktionen, die vom Lesen bis zum Besitz und der Ermittlung der verfügbaren Detailstufe für jedes der verfügbaren Elemente reichen. Während die Auswahl der Benutzerrolle und der Rechte für das Teilen- und Synchronisierungswerkzeug gleich sind, unterscheiden sich die beiden Aktionen hinsichtlich der Inhalte der bereitgestellten Forschungsdaten: Die gemeinsame Nutzung von Informationen durch Teilen ermöglicht den Zugriff auf einen gegebenen Satz Forschungsdaten. Der ausgewählte Kollege kann die Informationen z.B. lesen, schreiben, teilen und löschen hat aber keinen Zugriff auf laufende Änderungen der gemeinsamen Sammlung. Dies ist nur über die Erstellung von synchronisierten Sammlungen verfügbar. Synchronisierte Sammlungen werden erstellt, um einen permanenten Zugriff auf andere ELN-Benutzer auf den ausgewählten Satz von Forschungsdaten zu ermöglichen, einschließlich der Sichtbarkeit von Änderungen, die nach der Synchronisation vorgenommen wurden.

Ablage von Messwerten

Neben den unter einer Nutzerkennung eingetragenen Werten fallen in Laboratorien eine Vielzahl digital verfügbarer Daten an, die zur vollständigen Dokumentation des Forschungsprozesses mit im elektronischen Laborbuch erfasst werden müssen. Um eine Nachnutzbarkeit und eine nachvollziehbare Dokumentation zu gewährleisten, müssen diese Daten strukturiert abgelegt werden können. Das heißt, die Daten müssen zu den entsprechenden Analysen zugeordnet werden können. Ist eine Assoziation der Daten zu einem Benutzer nicht gegeben, wie beispielsweise bei

der automatischen Speicherung von Daten durch Messgeräte, muss dem Nutzer des elektronischen Laborbuchs eine Möglichkeit gegeben werden, diese Daten den entsprechenden Versuchen zuzuordnen zu können.

Die Daten können auf dem lokalen Filesystem des Servers gespeichert werden, auf dem auch die elektronische Laborbuch Anwendung installiert ist. Um ein Management größerer Datenmengen gewährleisten zu können, wurde eine Anbindung der Large Scale Data Facility (LSDF) ausgewählt. Da die Daten in einem elektronischen Laborbuch aktiv genutzt werden, das heißt die Daten häufig abgerufen, verändert oder ergänzt werden, sollten die Daten auf dem Datenspeicher in einer einfachen Ordnerstruktur vorliegen, um ein häufiges umkopieren der Daten zu vermeiden. Für die spätere Langzeitarchivierung ist es jedoch notwendig über das Datenmodell strukturierte Datencontainer inklusive der Metadaten (zum Beispiel im BagIt-Format) erzeugen zu können. Mit der Anbindung eines Langzeitarchivs können die Daten dann kostengünstig nachhaltig gesichert werden. Um die Nachhaltigkeit weiter zu adressieren, wird die Integrität der Daten mit ihrer Erzeugung erfasst und bis zur Ablage ins Archiv kontrolliert. Damit ist ein Nachweis über die Unversehrtheit der Daten über den gesamten Forschungszyklus hinweg möglich.

Chemotion-Repository zur Bereitstellung von Forschungsdaten

Das Repository für chemische Forschungsdaten ist mit den gleichen allgemeinen und fachspezifischen Funktionen ausgestattet, die das Chemotion-ELN bietet. Einige wenige Funktionen wurden hinzugefügt, um eine Datenübermittlung zur Registrierung der Forschungsdaten über DataCite zu ermöglichen und eine Indexierung in der Datenbank PubChem zu initiieren. Die Installation der Software muss auf einem allgemein zugänglichen Server durchgeführt werden, um die Verfügbarkeit der bereitgestellten Daten auch außerhalb der eigenen Forschungseinrichtung zu ermöglichen. Zu diesem Zweck wurde am KIT die Installation auf einer virtuellen Maschine des Rechenzentrums mit öffentlich zugänglicher IP-Adresse gewählt. Während in einer frühen Version des Chemotion-ELNs Daten noch einzeln eingegeben werden mussten, gelingt durch die Verwendung der Freigabe-Funktion des ELNs explizit für die Sammlung Chemotion-Repository eine einfache und schnelle Bereitstellung der erhaltenen Forschungsdaten. Zu diesem Zweck können die bereits innerhalb des ELNs zur Verfügung stehenden Module der Rechteverwaltung genutzt und auf den Transfer der Daten und den gewünschten Detaillevel hin eingestellt werden.

Suche nach Forschungsdaten

Eines der Hauptargumente für die Verwaltung von Forschungsdaten mit einem ELN ist die digitale Verfügbarkeit von Informationen und die damit verbundene Möglichkeit, nach Daten und Informationen zu suchen, wenn die Organisation und Pflege des ELNs bzw. des Repositoriums dies in geeigneter Weise unterstützt. Die Chemotion Produkte ermöglichen Text- und Struktursuche innerhalb diverser Inhalte des ELNs. Die Suche nach Textfragmenten oder chemischen Strukturen kann weiter auf unterschiedliche Elemente (Proben, Reaktionen) beschränkt werden, um die Auswertung der Ergebnisse zu erleichtern. Die Textsuche ist dazu bestimmt, nach dem Vorhandensein von Text- oder Formelfragmenten in Proben als Basiseinheit zu durchsuchen. Die meisten der nicht-numerischen Eigenschaften wie z.B. Name, Molekülformel, IUPAC-Name, oder kano-

nischer Smiles-String können als Suchparameter verwendet werden. Der zugehörige Inhalt in Reaktionen wird auf der Grundlage des Suchergebnisses gefiltert. Für Reaktionen sind der Name und die Nummer der Reaktion suchbar.

Die Struktursuche im Gegensatz zur Textsuche kann entweder durch die Suche nach einer Unterstruktur oder einer Ähnlichkeitssuche erfolgen (Bajusz et al. 2015). Diese Suchmethoden sind Fingerprint-basierte Methoden. Innerhalb des ELNs und Repositoriums wurde eine pfadbasierte Fingerabdruckmethode implementiert, die als FP2 (OpenBabel) bezeichnet wird. Dieser Fingerprint ist identisch mit den Daylight Fingerprints (James & Weininger 2006) die als Standard in vielen Publikationen verwendet werden.

Literaturangaben

- Winkler-Nees, S. 2013. “Status of Discussion and Current Activities: National Developments”. In *Digital Curation of Research Data Experiences of a Baseline Study in Germany*, H. Neuroth, S. Strathmann, A. Oßwald, J. Ludwig (Hrsg.). VWH. Kapitel 2, 18–33.
- Coles, S. J., J. G. Frey, C. L. Bird, R. J. Whitby, A. E. Day. 2013. “First steps towards semantic description of electronic laboratory notebook records”. *Journal of Cheminformatics* 5 (52).
- Rees, I., E. Langley, W. Chiu, S. J. Ludtke. 2013. “EMEN2: an object oriented database and electronic lab notebook”. *Microsc. Microanal* 19 (1): 1–10.
- Barillari, C., D. S. M. Ottoz, J. M. Fuentes-Serna, C. Ramakrishnan, B. Rinn, F. Rudolf. 2016. “openBIS ELN-LIMS: an open-source database for academic laboratories. *Bioinformatics* 32 (4): 638–640.
- Rubacha, M., A. K. Rattan, S. C. J. Hosselet. 2011. “A review of electronic laboratory notebooks available in the market today”. *Lab. Autom.* 16 (1): 90–98.
- Day, A. E., S. J. Coles, C. L. Bird, J. G. Frey, R. J. Whitby, V. E. Tkachenko, A. J. Williams. 2015. “ChemTrove. Enabling a Generic ELN To Support Chemistry through the Use of Transferable Plug-ins and Online Data Sources”. *Journal of Chemical Information and Modeling* 55: 501–509.
- Willoughby, C., C. L. Bird, S. J. Coles, J. G. Frey. 2014. “Creating Context for the Experiment Record. User-Defined Metadata: Investigations into Metadata Usage in the LabTrove ELN”. *Journal of Chemical Information and Modeling* 54 (12): 3268–3283.
- Milsted, A. J., J. R. Hale, J. G. Frey, J. G. C. Neylon, C. 2013. “LabTrove: A Lightweight, Web Based, Laboratory ‘Blog’ as a route towards a Marked Up Record of Work in a Bioscience Research Laboratory”. *PLOS One* 8: e67460. Online verfügbar unter: <https://doi.org/10.1371/journal.pone.0067460>. Zuletzt geprüft am 12.08.2017.

- Rudolphi, F., L. J. Goossen. 2012. “Electronic laboratory notebook: the academic point of view”. *Journal of Chemical Informations and Modeling* 52 (2) 293–301.
- Kelly, R., R. Kidd. 2015. “Editorial: ChemSpider - a tool for Natural Products research”. *Nat. Prod. Reports* 32: 1163–1164.
- Bajusz, D, A. Rácz, K. J. Héberger. 2015. „Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?“. *Journal of Cheminformatics* 7 (20).
- James, C.A., D. Weininger. 2006. Daylight theory manual: Daylight Chemical Information Systems.