

# Update April - Juli 2021

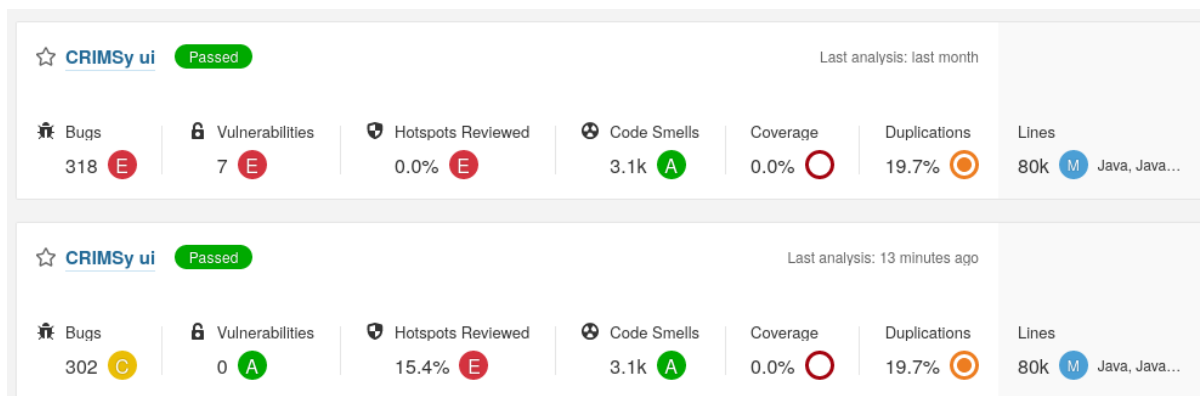
CRIMSy – Leibniz Bioactives Cloud – Frank Broda ([fbroda@ipb-halle.de](mailto:fbroda@ipb-halle.de))

Wie im letzten Update vom März 2021 angekündigt, haben wir uns stark auf die Erstellung einer auslieferungstauglichen Version konzentriert. Diese Bemühungen haben Früchte getragen und Ende Juli sind die Knoten in Halle (IPB) und Freising (LSB@TUM) mit der neuen Version online gegangen. Wie in den vergangenen Newslettern bereits angekündigt, stehen in der neuen Version eine Vielzahl neuer Funktionen zur Verfügung. Die Instanz in Halle enthält zudem mehrere tausend Datensätze aus der lokalen InhouseDB. Dabei wurden Datensätze, die auch in PubChem enthalten sind, allgemein zugänglich gemacht.

Wir werden in den nächsten Wochen weitere Partner kontaktieren, um das Netzwerk auszudehnen. Wir gehen jedoch davon aus, dass es urlaubszeitbedingt zu Verzögerungen kommen wird.

## Qualitätssicherung

In den über 320 Commits der letzten Monate wurde nicht nur das Feature zur Bildannotation fertig gestellt, sondern auch eine Vielzahl von Fehlern und störenden Eigenheiten ausgemerzt. An vielen Stellen wurden Unit-Tests ergänzt, so dass die Testabdeckung jetzt 62 Prozent beträgt. Ebenso wurde mit der Überarbeitung der Dokumentation begonnen, damit es nicht zu Diskrepanzen zwischen Dokumentation und Programmverhalten kommt.



Einen Qualitätssprung hat unser Projekt infolge einer Analyse mit dem Werkzeug SonarQube gemacht. SonarQube konnte eine Reihe problematischer Stellen in unserem Source-Code aufdecken, wobei wir keine der angezeigten Schwachstellen in die Hochrisikogruppe einstufen würden. Wie der Vergleich zeigt, konnte ein guter Teil der problematischen Stellen bereits abgearbeitet werden. Viele der verbleibenden problematischen Stellen lassen sich auf das Tool fineUploader zurückführen, für das wir bereits am Austausch arbeiten.

## Public Relations

Um Anwendern einen schnellen Überblick über unser Projekt zu verschaffen, haben wir ein Video produziert, in dem einige der Fähigkeiten von CRIMSy exemplarisch dargestellt sind. Um niemanden zu langweilen haben wir uns auf 2 Minuten und 45 Sekunden beschränkt:

[https://youtu.be/PDGm0\\_Wsly0](https://youtu.be/PDGm0_Wsly0)

Außerdem werden wir unser Projekt Ende September auf den folgenden (virtuellen) Veranstaltungen präsentieren:

- Leibniz Conference on Bioactive Compounds 2021 in Halle und
- Als Software-Demo im Rahmen des Workshops *Computersoftware for Biodiversity* auf der INFORMATIK 2021 in Berlin

# Update Februar / März 2021

---

CRIMSy – Leibniz Bioactives Cloud – Frank Broda ([fbroda@ipb-halle.de](mailto:fbroda@ipb-halle.de))

In den vergangenen Wochen ist es uns gelungen, den größten Teil unserer „blocking issues“ abzuarbeiten. Dabei sind über 200 Commits in verschiedenen Repositorien entstanden. Wir beginnen jetzt mit intensiven Tests, um die Auslieferung vorzubereiten. Meilensteine unserer Arbeit waren:

## Suche

Die verteilte Suche in komplexen Objektbeziehungen unter Berücksichtigung der jeweiligen Zugriffsrechte konnte erfolgreich überarbeitet werden. Durch die komplexe Datenstruktur (Experiment enthält Sample, Sample besteht aus Material, Material hat Strukturformel usw.; jeweils mit Berechtigungsinformation) kommen schnell zweistellige Anzahlen von Datenbanktabellen zusammen. Unser Code kann Abfragen so aufbauen, dass nur die tatsächlich notwendigen Tabellen einbezogen werden. Gleichzeitig wurden die Funktionen gekapselt, so dass die Komplexität auch weiterhin beherrschbar bleibt.

## User Interface

Bisher stand nur ein Moleküleditor mit stark eingeschränkter Funktionalität zur Verfügung. Durch das von Frank Lange geschaffene Paket MolecularFaces (<https://github.com/ipb-halle/MolecularFaces>) können Nutzer jetzt zwischen mehreren Moleküleditoren wählen (z.B. OpenChemLib-JS oder MolPaintJS; bei Vorhandensein einer Lizenz auch MarvinJS von ChemAxon). Der Editor MolPaintJS wurde deutlich verbessert und es wurden erste Sondierungen zur Darstellung von Spektrendaten vorgenommen.

In der ausgelieferten Version von CRIMSy wird zudem eine Möglichkeit zur Speicherung und Annotation von Bilddaten vorhanden sein.

## Container-Infrastruktur

Aus Sicherheitsgründen und aufgrund des technischen Fortschritts müssen die Container, in denen die Anwendung ausgeführt wird, regelmäßig einer Revision unterzogen werden:

- zur Vereinfachung der Softwaredistribution wird von uns ein spezielles Container-Image für den Proxy-Container auf Docker-Hub zur Verfügung gestellt. In diesem Image ist eine Reihe von JavaScript-Bibliotheken, z.B. für chem. Strukturformeln und Image-Annotation, gebündelt.
- Es wurde ein Review und Benchmarking von verschiedenen Datenbankerweiterungen zur Speicherung und Indexierung von chemischen Strukturen durchgeführt. Als Ergebnis wurde der Datenbankcontainer von pgchem::tigress auf Bingo (<https://github.com/epam/Indigo>) umgestellt.

## Ausblick

In den nächsten Wochen planen wir, unser System mit größeren Datenbeständen zu testen. Zusätzlich zu den Strukturen und taxonomischen Informationen werden wir hierzu weitere Teile (Experimente, Samples) aus der InhouseDB des IPB importieren. Wichtigstes Nahziel bleibt die Auslieferung an unsere Kooperationspartner.

# Update Dezember 2020 / Januar 2021

CRIMSy – Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

Nach wie vor liegt unser Fokus auf möglichst schneller Auslieferung eines einsatzfähigen Produkts. Durch den Jahreswechsel, die aktuellen COVID-Beschränkungen und verschiedene Spezialaufträge sind wir mit unseren Plänen in Rückstand gekommen. Im Dezember 2020 und im Januar 2021 haben wir vorrangig daran gearbeitet, die einzelnen Module vom Prototyp-Stadium zur Anwendungsreife weiterzuentwickeln und Fehler auszumerzen. Ein Großteil der fast 100 Commits der letzten zwei Monate diente diesem Zweck.

Daneben haben wir die Möglichkeit zur Verlinkung von Datensätzen geschaffen, um im weiteren Verlauf Teile der In House DB des IPB importieren zu können. Im November hatten wir über den Teilimport von Strukturen berichtet. Zukünftig wird auch die Verknüpfung dieser Informationen zu Experimenten, Proben (Samples) und Organismen möglich sein.

Experiment-Einträge

6 -- 2020-09-15 13:00:16

## References and Links

In December 2020 / January 2021 a feature was added to enable the linking of materials (compounds), items (samples, probes) and possibly other entities. The feature is a great step towards the planned import of external data sources. It works by adding link records and mentioning the link label with a hash tag (“#”) in the text. In view mode the link is rendered as a clickable link:

- [#BENZOL](#) - a compound
- [#CODEINE](#) - another compound
- [#ASS\\_SAMPLE](#) - a sample
- possibly more links in the future (documents)

If clicked, the links open a modal dialog displaying the linked data.

[BENZOL](#) [ASS\\_SAMPLE](#) [CODEINE](#)

### Codein

Alternative names	Indices
Codein	CAS not available
	SMILES not available
	InChI not available

In diesem Zusammenhang ist uns auch aufgefallen, dass die feingranulare und damit hochkomplexe Rechteverwaltung bei Datenbankabfragen noch einmal nachgeschärft werden muss. Dies wird voraussichtlich Gegenstand des nächsten Newsletters sein.

Herr Lange hat seine Arbeit bei uns aufgenommen. Über seine ersten Ergebnisse wird ebenfalls im nächsten Newsletter berichtet werden.

# Update Oktober / November 2020

CRIMSy – Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

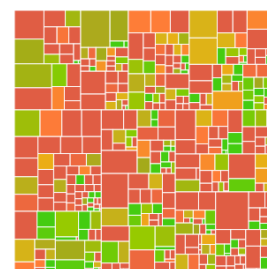
Im Oktober und November 2020 hat es über 200 Änderungen (Commits) in unserem öffentlichen Repository (<https://github.com/ipb-halle/CRIMSy>; Entwicklungszweig LBAC-60-CloudLIMS) gegeben:

- Ein Großteil der Änderungen zielte darauf ab, unseren Code produktionsreif zu bekommen. Viele Features, die im Laufe des letzten Jahres prototypisch implementiert wurden, müssen jetzt in eine präsentable Form gebracht werden. Dies betrifft zum Teil die Programmstabilität; den größten Aufwand erfordert jedoch die Präsentationsschicht (Layout usw.). Die Fehlersuche in der Präsentationsschicht ist zudem wesentlich aufwändiger als auf der Server- und Datenbankseite. Bei der Einführung neuer Features haben wir uns strengste Zurückhaltung auferlegt.
- Das im letzten Newsletter angekündigte Framework für Datenbankabfragen wurde umgesetzt und funktioniert wie erwartet. Abfragen der Nutzer können in der Cloud verteilt werden und die Knoten liefern (im Gegensatz zu früher) nicht nur Dokumente, sondern auch Datensätze mit Materialien, Gebinden und experimentellen Ergebnissen zurück.

The screenshot shows the Bioactives Cloud web interface. At the top, there is a navigation bar with 'BioCloud' and 'LIMS' on the left, and 'Logout' and 'Einstellungen' on the right. The main header features the 'Leibniz Bioactives Cloud' logo. Below the header is a search bar containing the text 'codein'. To the right of the search bar are buttons for 'Aktualisieren' and 'Dokument zum Upload'. Below the search bar is a table with the following columns: Name, Location, Type, and Relevanz. The table contains several rows of data, including 'Codein' and various 'IPB\_Jahresbericht\_2001.pdf' through '2005.pdf' files. To the right of the table is an 'Advanced Search' panel with various filters and checkboxes. Below the table is a 'Suchen' input field and a 'Zurück' button. On the right side of the interface, there is a chemical structure viewer showing a complex organic molecule with a nitrogen atom and a hydroxyl group.

Name	Location	Type	Relevanz
Codein	NWC04170	MATERIAL-STRUCTURE	
IPB_Jahresbericht_2001.pdf	NWC04170	DOCUMENT	0.322
IPB_Jahresbericht_2000.pdf	NWC04170	DOCUMENT	0.488
IPB_Jahresbericht_2002.pdf	NWC04170	DOCUMENT	0.423
IPB_Jahresbericht_2003.pdf	NWC04170	DOCUMENT	0.526
IPB_Jahresbericht_2004.pdf	NWC04170	DOCUMENT	0.543
IPB_Jahresbericht_2005.pdf	NWC04170	DOCUMENT	0.285

- Es wurde begonnen, testweise Datensätze aus der *In House Datenbank* der Abteilung Natur- und Wirkstoffchemie am IPB zu importieren. In einem weitgehend automatisierten Prozess wurden fast 8000 Strukturen importiert. Für einen vollständigen Import (knapp 21.000 Strukturen) müssen einige Datensätze vorab manuell kuratiert werden, da das verwendete Datenbank-Plugin einige Regeln strikter auslegt als die Originaldatenbank.
- Wir haben unserem Projekt auf GitHub einen „Badge“ von codecov.io hinzugefügt, der die Testabdeckung des Projekts anzeigt. Außerdem sind über diesen Badge einige qualitätsrelevante Codemetriken (z.B. Größenverteilung der Quellcode-Dateien, siehe Abb.) abrufbar.



## Personelle Verstärkung

Wir freuen uns, dass Frank Lange im Februar 2021 erneut (leider befristet bis Ende 2021) zu unserem Team stoßen wird. In der Anfangsphase (2016 / 2017) hat er dem Projekt wertvolle Impulse verliehen.

# Update August / September 2020

---

CRIMSy – Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

Die vergangenen beiden Monate waren einerseits durch umfangreiches *Refactoring* gekennzeichnet, mit dem wir uns von vielen Altlasten getrennt haben:

- Mit einer Ausnahme haben wir unsere Datenbankschlüssel von *Universal Unique Identifiers* (UUIDs) auf Integer-Datentypen umgestellt. UUID-Schlüssel sind i.d.R. instanzübergreifend eindeutig, während Integer-Datentypen dies nicht sicherstellen können. Dafür sind Integer-Zahlen (z.B. „27“) für Menschen viel leichter benutzbar als UUIDs (z.B. „92c47aaa-539b-48d2-aa20-dbb062121637“).  
Durch ein geändertes Konzept sind instanzübergreifend eindeutige Schlüssel seit längerem (fast) nicht mehr notwendig und UUIDs seitdem folglich nur noch hinderlich. Die Umstellung erforderte Änderungen an ungefähr 1900 Codezeilen in über 400 Dateien und konnte dank unserer Unit-Tests innerhalb weniger Tage vollzogen werden.
- HTML Sanitation wurde auf ein neues Framework (OWASP Java HTML Sanitizer) umgestellt, um bei der Verarbeitung von formatiertem Text keine CSS-Sicherheitslücken (CSS = Cross-Site-Scripting) zu riskieren. Der bislang verwendete Code war im Funktionsumfang zu sehr eingeschränkt; eine Erweiterung des bisherigen Codes erschien sowohl unter dem Effizienz- als auch dem Sicherheitsaspekt nicht sinnvoll.
- Unsere Anwendung kommt seit Anfang September ohne die Komponente Apache SOLR aus. Obwohl Apache SOLR eine hervorragende Software für die Indexierung großer und größter Datenbestände ist, war die Nutzung in der Leibniz Bioactives Cloud von Anfang an auch irgendwie eine Notlösung. Der Zugriff auf die SOLR-Datenbank konnte nur über eine interne REST-Schnittstelle erfolgen und bestimmte Anwendungsszenarien (z.B. die globale Relevanzmetrik und die WordCloud-Suche) erforderten eine doppelte bzw. sogar dreifache Datenhaltung (SOLR, PostgreSQL und Dateisystem). Wir konnten Apache SOLR jetzt durch eine von uns in einem anderen Projekt entwickelte konfigurierbare Text Processing Pipeline ersetzen. Die Rohextraktion der Dokumente (PDFs, Office, ...) erfolgt, wie in SOLR auch, mittels Apache Tika. Die weiteren Stufen (Stemming, Stop Word Filtering, ...) erfolgen jedoch durch unsere Pipeline.

Andererseits haben wir etliche neue Funktionen implementiert und vorhandene Komponenten erweitert:

- Es ist jetzt möglich, Nutzervorlieben und zuletzt gewählte Einstellungen zu speichern
- Ein neues Formular ermöglicht, Zugriffsrechte für verschiedene Ressourcen zu verwalten
- Wir sind dabei, ein Framework zu schaffen, welches uns gleichzeitig einfache wie auch komplexe Suchen über die verschiedensten Ressourcen (Materialien, Gebinde, Dokumente, Experimente, Nutzer etc. pp.) ermöglichen wird. Eine einfache Suche wird wie bei einer Suchmaschine durch die Eingabe von Schlüsselworten gekennzeichnet sein. In komplexen Suchen können Anfragen wie „**Zeige alle Assay-Ergebnisse (Leseberechtigung vorausgesetzt) der letzten 12 Monate für Targets aus der Gattung *Bacillus* bei denen eine  $\beta$ -Lactam-Substruktur getestet wurde.**“ formuliert<sup>1</sup> werden. Das Framework wird auch erlauben, bislang nur lokal verfügbare Ressourcen (Materialien, Gebinde etc.) mit **berechtigten (!)** externen Nutzern zu teilen.
- Bei den Integrations-Tests mit Selenium hat es Fortschritte gegeben. Wir haben mit SeleniumIDE erste Testfälle erzeugt und die Distributions- und Installationskripte können jetzt vollautomatisch eine Testumgebung aufsetzen. Die Mechanik unseres Web-Frameworks verhindert momentan noch den vollen Erfolg, da die Selenium-Tests zu schnell ausgeführt werden.

---

<sup>1</sup> Natürlich zunächst nicht in natürlicher Sprache

# Update Juni / Juli 2020

CRIMSy – Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

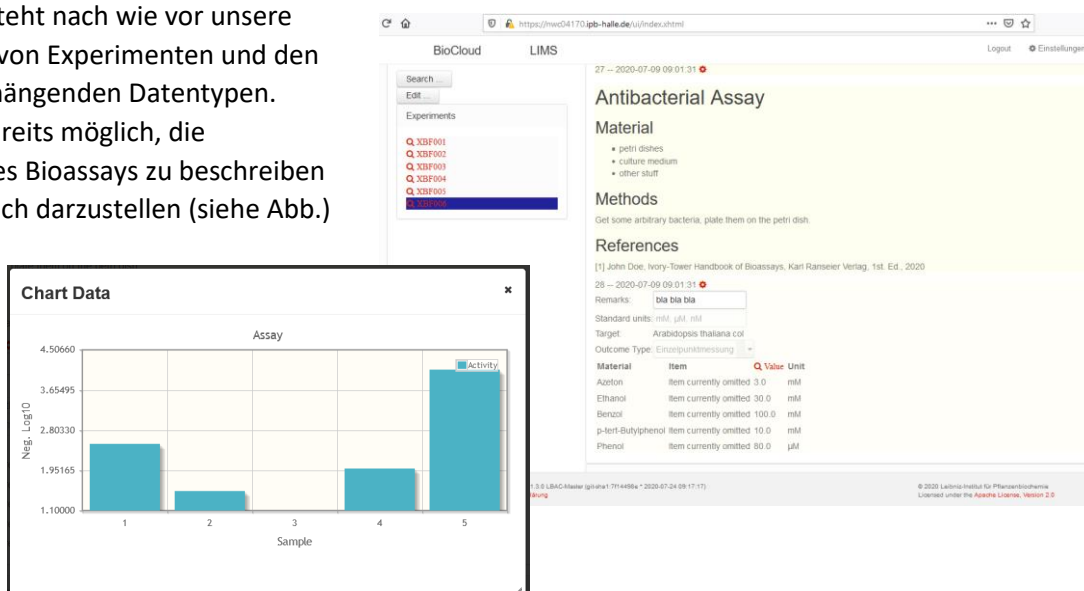
## Entwicklungsfortschritt

Bei der Umsetzung unserer Pläne sind wir gut vorangekommen. Viele notwendige Funktionen wie

- Substruktursuche
- Verwaltung von Substanzen, Biomaterial (Organismen, Gewebe, ...)
- Verwaltung von Gebinden und Containern
- Verwaltung von „Projekten“ (ein Konstrukt zum vereinfachten Rechte-Management)
- Ansteuerung von Barcode-Druckern zum Drucken von Gebinde-Etiketten

sind bereits implementiert. Weite Teile des Codes berücksichtigen die jeweiligen Zugriffsrechte und speichern versionierte Daten, so dass sich Änderungen rückverfolgen lassen.

Ganz am Anfang steht nach wie vor unsere Implementierung von Experimenten und den damit zusammenhängenden Datentypen. Trotzdem ist es bereits möglich, die Durchführung eines Bioassays zu beschreiben und Daten graphisch darzustellen (siehe Abb.)



In einem der nächsten Schritte wollen wir unseren Code netzwerkfähig machen, d.h. die verschiedenen Datentypen bzw. Objekte (Substanzen, Gebinde, Experimente, ...) über das Netzwerk austauschen.

Als besonders aufwändig hat sich die Erstellung des User-Interface herausgestellt. Fehler in einer der verwendeten Bibliotheken (Mojarra) haben uns etwas Zeit gekostet und schließlich zum Austausch gegen MyFaces geführt. Außerdem erschweren uns die Komplexität des Nutzerinterface und die asynchrone Verarbeitung von Anfragen (AJAX) gelegentlich das Vorankommen. Mit der Kapselung von Darstellungseinheiten in *Composite Components* haben wir aber auch hier einen Weg gefunden, die Komplexität zu bändigen und uns die Entwicklung zu erleichtern.

## Statistik

Nach aktuellem Stand besteht unser Projekt aus mehr als 68.000 Zeilen Java Code und mehr als 8700 Zeilen HTML. Hinzu kommen geschätzt 15.000 Zeilen Shell-Skripte, Datenbankkonfiguration, JavaScript, CSS, Dokumentation usw. (Dateien aus fremden Quellen nicht eingerechnet).

Die Testabdeckung hat sich leicht auf 48 % verbessert.



# Update Februar / März 2020

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Feature Update

In den vergangenen Wochen ist es uns gelungen, das im letzten Newsletter beschriebene MultiCloud-Feature produktionsreif zu bekommen und an unsere Knoten in Dortmund, Freising, Jena, Magdeburg und Halle zu verteilen. Wir hatten eine weitere tiefgreifende Änderung (DTOs, s.u.) in dieses Update gepackt und sind daher froh und stolz, dass das automatische Update aller Knoten im Wesentlichen reibungslos funktioniert hat.

## Einführung einer DTO-Abstraktionsschicht

Die ständig zunehmende Komplexität und Verflechtung unserer Datenstrukturen und das Zusammenspiel mit Hibernate hat uns Kopfschmerzen bereitet. Hibernate ist ein modernes, umfangreich dokumentiertes und häufig genutztes Framework, um Objekte in einer Datenbank zu persistieren. Die Lage kann jedoch unübersichtlich werden, wenn wie in unserem Fall sehr komplexe Objekthierarchien in einer verteilten Umgebung bearbeitet werden. Um dauerhaft die Übersicht zu behalten, haben wir uns entschlossen, die Datenschicht (Hibernate Entities) systematisch von der Logikschicht zu trennen und Datentransferobjekte (DTOs) einzuführen. Dies machte umfangreiche Änderungen an unserem Quelltext erforderlich. Unsere Unit-Tests haben geholfen, die Umstellung schnell und in hoher Qualität durchzuführen.

## CloudLIMS

Herr Mauz hat außerdem viel Kraft in die Entwicklung des CloudLIMS (siehe Newsletter Oktober / November 2019) gesteckt. Unser Hauptaugenmerk liegt aktuell auf der Implementierung der Konzepte:

- **Material:** darunter verstehen wir die konzeptionelle bzw. theoretische Beschreibung von Objekten um sie zu klassifizieren. Dabei ist unerheblich, ob diese Objekte real existieren. Beispiele für Klassen von Materialien sind u.a.
  - chemische Reinverbindungen, z.B. Aceton
  - Substanzgemische; bei Extraktfraktionen unter Angabe des Quellorganismus
  - Enzyme, Plasmide, Primer (→ ggf. mit Sequenzinformation)
  - Organismen (Taxonomie)
- **Gebinde:** sind real existierende Gefäße (Flaschen, Tubes, Vials, Beutel, ...), die ein bestimmtes Material enthalten
- **Container:** sind Aufbewahrungssysteme (Regale, (Kühl-)Schränke, Racks, Magazinsysteme, ...) für Gebinde mit bestimmter oder unbestimmter Kapazität

Die softwaretechnische Umsetzung erfolgt so, dass die Zugriffsrechte für alle Objekte feingranular verwaltet werden können. Ebenso wird die Historie der Objekte gespeichert, so dass rückverfolgbar ist, welche Änderungen im Verlauf der Zeit vorgenommen wurden. In Verbindung mit der vergleichsweise einfach umzusetzenden Blockchain-Technologie werden die gespeicherten Daten hohe Beweiskraft erhalten.

## GitHub

Die Corona-Pandemie hat den Entschluss unseren Code auf GitHub zu veröffentlichen (<https://github.com/ipb-halle/CRIMSy>) beschleunigt, weil wir uns davon eine Erleichterung der Arbeit im Homeoffice versprechen. Der CloudLIMS-Zweig ist allerdings noch nicht enthalten.



# Update Dezember 2019 / Januar 2020

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Sicherheit und Datenschutz

Seit dem letzten Bericht haben wir uns intensiv um die Verbesserung der Sicherheit der Leibniz Bioactives Cloud und um Datenschutzaspekte gekümmert. Ein Teil dieser Verbesserungen geht auf Anregungen aus dem Leibniz-Institut für Lebensmittel-Systembiologie an der Technischen Universität München (Leibniz-LSB@TUM) zurück. Das Leibniz-LSB@TUM ist in einer besonderen Situation, da es die Rechenzentrumsinfrastruktur der TU München nutzt und insbesondere keine eigene DMZ<sup>2</sup> betreibt.

Wir planen, diese Verbesserungen in den nächsten Wochen zusammen mit dem Multi Cloud Feature (siehe unten) auszurollen. Unter anderem haben wir folgende Maßnahmen umgesetzt:

- Optionale Aktivierung von HSTS (HTTP Strict Transport Security)
- Content Security Policy und X-Frame-Options (die die Einbindung in Frames fremder Seiten verbieten)
- Intruder Lockout (Accounts und Internetadressen werden nach mehreren fehlerhaften Login-Versuchen gesperrt, um das Erraten von Passwörtern zu erschweren)
- Eine vom externen Datenschutzbeauftragten des IPB geprüfte Datenschutzerklärung
- Verbesserungen hinsichtlich Datensparsamkeit und Anonymisierungsmöglichkeiten
- Lokale Spiegelung von JavaScript-Bibliotheken (z.B. JQuery), die vorher individuell vom Nutzerbrowser geladen wurden

## Workshop

Am 29. Januar fand am IPB in Halle ein Workshop zur Leibniz Bioactives Cloud statt. Wir haben dort unseren Gästen aus Dortmund und Braunschweig den aktuellen Stand präsentiert und unsere Pläne vorgestellt. Wir haben während des Workshops ein Video aufgezeichnet, das Sie sich ebenso wie die zugehörige Präsentation über die Webseite des Forschungsverbunds „Wirkstoffe und Biotechnologie“ ansehen können:

[https://www.leibniz-wirkstoffe.de/projects/lbac\\_project/](https://www.leibniz-wirkstoffe.de/projects/lbac_project/)

## MultiCloud

Die Leibniz Bioactives Cloud ist im sachsen-anhaltinischen Verbund „Autonomie im Alter“ auf Interesse gestoßen, an dem neben dem IPB z.B. auch die Universität und das LIN in Magdeburg beteiligt sind. Um zukünftig die parallele Nutzung unserer Software in mehreren Forschungsverbänden zu ermöglichen und gleichzeitig die strikte Trennung der Datenverkehre als auch die jeweiligen Zugriffsbeschränkungen durchzusetzen, haben wir uns das sogenannte MultiCloud-Feature ausgedacht. Durch die Verwendung unterschiedlicher Zertifizierungsstellen können wir für jede Cloud eine strikte Abschottung nach außen garantieren. Gleichzeitig kann ein Knoten jedoch Mitglied in mehreren Clouds sein, der Knoten benötigt lediglich ein Zertifikat der entsprechenden Zertifizierungsstelle einer Cloud und wenige Zeilen Konfigurationsdaten.

---

<sup>2</sup> Demilitarisierte Zone (ein Netzwerkbereich, der durch Firewalls sowohl von außen als auch von innen zugriffsbeschränkt ist)

# Update Oktober / November 2019

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Auslieferung

Im Oktober und November haben wir uns intensiv darum gekümmert, Knoten in unseren Partnerinstituten in Betrieb zu nehmen. Stand Ende November sind Knoten in Dortmund (ISAS), Jena (HKI), Magdeburg (LIN), Freising (Leibniz-LSB@TUM) und Halle (IPB) in Betrieb. Weitere Institute (IPK Gatersleben und FLI Jena) haben Interesse signalisiert.

Die Diversität der IT-Landschaften an den beteiligten Instituten hat einige kleinere Schwachstellen in unserem Code offenbart, die wir zügig beheben konnten.

## Workshop

Zukünftig wird es darauf ankommen, Wissenschaftler ins Boot zu holen, die die Knoten in ihren Instituten mit Daten füttern. Wir möchten weiterhin darum bitten, uns KollegInnen für die Institute zu nennen, die mitarbeiten möchten und als Ansprechpartner fungieren könnten. Eine gute Gelegenheit, Werbung zu machen bietet der „Leibniz Cloud“-Workshop Ende Januar in Halle. Wissenschaftler und IT-Fachkräfte aus den Instituten des Forschungsverbands sind herzlich eingeladen, ihre Teilnahme zu registrieren: <https://events.ipb-halle.de/event/56/>. Die Reise- und Hotelkosten für 2 Teilnehmer pro Institut werden vom Forschungsverbund übernommen.

## Leibniz Jahrestagung

Auf der Mitgliederversammlung des Forschungsverbands „Wirkstoffe und Biotechnologie“ haben wir unseren Entwicklungsstand präsentiert, Rechenschaft über unsere Arbeit abgelegt und unsere weiteren Pläne skizziert.

## CloudLIMS

Durch die Investitionsbank Sachsen-Anhalt wurde dem IPB das Projekt „ProCognito“ im Rahmen der Initiative „Autonomie im Alter“ bewilligt. Dieses Projekt enthält ein Arbeitspaket zur Schaffung einer verteilten Infrastruktur, über die die Projektpartner beispielsweise Zugriff auf die IPB-Substanzdatenbank (und damit indirekt auf die physische Substanzbibliothek) erhalten sollen. Die Infrastruktur der Leibniz Bioactives Cloud ist hervorragend zur Realisierung dieses Vorhabens geeignet und wir werden die aktuelle „In House Datenbank“ des IPB, unsere Chemikalienverwaltung „KICKS“ und die Leibniz Bioactives Cloud zu einem CloudLIMS verschmelzen. Durch umfassendes und feingranulares Rechtemanagement werden die teilnehmenden Institute völlig frei sein, diese Infrastruktur ausschließlich für sich selbst zu nutzen oder frei bestimmbare Mengen an Daten und Ressourcen mit Partnerinstituten zu teilen.

# Update August / September 2019

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Dokumentenhandling

Wir haben uns mit der Internationalisierung und dort vor allem mit der Behandlung von Dokumenten in unterschiedlichen Sprachen auseinandergesetzt. Im Verlauf der Indexierung von Dokumenten unterzieht man den Text üblicherweise einem Stemming (d.h. Rückführung auf die Stammform: „gegangen“ → „gehen“) und einer Stopwort-Filterung (d.h. Entfernung von Pronomen, Präpositionen usw.). Dies muss sprachspezifisch erfolgen. Vor allem hat aber die konkrete Konfiguration Einfluss darauf, welche Begriffe in der Word Cloud auftauchen. Zu aggressives Vorgehen führt zu seltsamen Artefakten, weshalb wir für den aktuellen Kompromiss etwas probieren mussten.

Außerdem haben wir ein Maß für die Relevanz der Suchergebnisse entwickelt und implementiert, da die üblichen TF / IDF basierten Metriken in einem verteilten Suchindex nicht funktionieren. Das Problem war uns schon lange bewusst; als Platzhalter für das echte Relevanzmaß hatten wir bislang Zufallszahlen verwendet.

## Nutzerforum

Vor Auslieferung haben wir noch ein einfaches Nutzerforum eingebaut. Unser Vorbild war u.a. die Webseite <https://slack.com>. Aufgrund der verteilten Architektur gibt es keine fertigen Lösungen. Durch Beschränkung auf das Wesentliche (d.h. Textnachrichten) gelang uns die Minimierung des Implementierungsaufwands.

## Bugfixes

In Vorbereitung der Auslieferung haben wir das System gründlich getestet und über 25 Bugs (darunter viele ärgerliche Kleinigkeiten) behoben. Einige Fremdbibliotheken (v.a. JavaScript), die wir in der Vergangenheit eingebunden hatten, haben wir ersetzt. Dies geschah vor allem dann, wenn die Sicherheitsauswirkungen durch die Bibliotheken schwierig zu beurteilen waren.

Wir sind optimistisch, nächsten Monat erste Tests mit unseren Partnern im Forschungsverbund durchführen zu können.

# Update Juli 2019

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Rechtmanagement

In den vergangenen Wochen haben wir weiter konsequent an der Implementierung einer nutzer- (bzw. administratoren-)freundlichen Rechteverwaltung gearbeitet. Nach vielen Vorarbeiten „unter der Haube“ gibt es jetzt erste Browserdialoge zum Ausprobieren.

## Engineering

Einige unserer Java-Klassen waren unübersichtlich geworden, weil Code für Geschäftslogik und Nutzerinterface zu sehr vermengt waren. Darunter hat unter anderem auch die Testbarkeit gelitten – es ist kaum möglich, Tests für solch komplexe Klassen zu schreiben. Die Erfahrung von Herrn Mauz ermöglicht uns, besser zu werden. Dabei verlieren wir nicht den Fokus auf möglichst baldige Auslieferung, sondern kümmern uns praktisch ausschließlich nur Code, der sowieso überarbeitet werden muss.

## Internationalisierung etc.

Bei der Internationalisierung („I18N“) unserer Anwendung gehen wir jetzt konsequenter vor. Obwohl die Internationalisierung (Deutsch, Englisch) von Projektbeginn an vorgesehen war, war diese Anforderung zwischenzeitlich etwas vernachlässigt worden.

Ebenso wollen wir die sprachspezifische Indexierung von Dokumenten verbessern. In diesem Zusammenhang hat sich Herr Mauz auch das Ranking der Ergebnisse vorgenommen und einen Schätzer für die Relevanz auf der Basis von Okapi BM25 ([https://en.wikipedia.org/wiki/Okapi\\_BM25](https://en.wikipedia.org/wiki/Okapi_BM25)) implementiert.

## Spin-Off: MolPaintJS

Der im Mai 2017 vorgestellte Editor für chemische Strukturformeln hat sich weiterentwickelt und kann jetzt von anderen Projekten genutzt werden. Der Source-Code steht unter Apache 2.0 Lizenz und kann von Interessenten unter <https://github.com/ipb-halle/MolPaintJS> heruntergeladen werden. Und nur mal schnell anschauen und ausprobieren? Geht ganz einfach über den Link <https://ipb-halle.github.io/MolPaintJS>.

# Update Juni 2019

Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

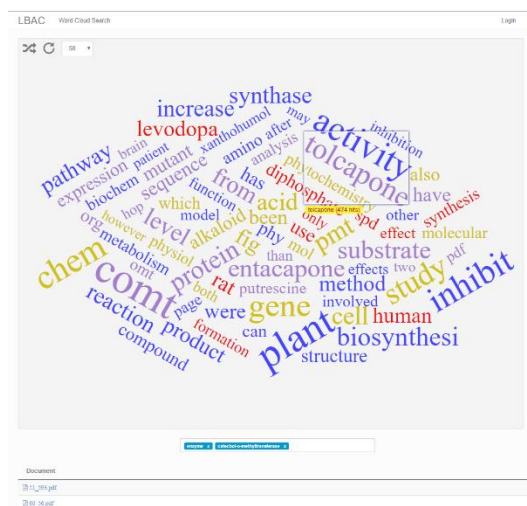
Der Zeitraum seit dem letzten Bericht vom März 2018 war nicht einfach für unser Projekt. An einigen Stellen traten unerwartet Schwierigkeiten auf und der Fortschritt verlief langsamer als geplant.

## Personelle Wechsel

Zum Dezember 2018 hat Herr Teuscher eine unbefristete Stelle beim Landesamt für Statistik angetreten. Er hat das Projekt mit seinem Wirken entscheidend vorangebracht, wofür wir ihm auch an dieser Stelle nochmals unseren Dank aussprechen wollen. Es war nicht einfach, gleichwertigen Ersatz für ihn zu finden. Nach zwei Ausschreibungsrunden konnten wir jedoch mit Herrn Fabian Mauz einen Diplom-Wirtschaftsinformatiker für unser Projekt rekrutieren. Herr Mauz bringt 10 Jahre Berufserfahrung als Programmierer mit und war auf dem Gebiet mathematischer Modellierung und Statistik tätig.

## Word Cloud Search

Zum Ende seiner Tätigkeit hat Herr Teuscher das Feature „Word Cloud Search“ fertiggestellt. Dieses Feature erlaubt, die Häufigkeit von Begriffen in den Ergebnisdokumenten visuell darzustellen und die Suchanfrage interaktiv zu verfeinern. Der eigentliche Wert der Arbeit liegt dabei nicht unbedingt im implementierten Modul „Word Cloud Search“ sondern im besseren Verständnis der Technologie von Apache Solr, der Prozessierung von natürlichsprachlichen Dokumenten usw.



## Rechtmanagement

Nachdem das Projekt bislang nur über rudimentäres Rechtmanagement verfügte, haben wir im Februar 2019 mit der Neukonzeption begonnen. Diese Arbeiten stehen kurz vor dem Abschluss und werden die Administratoren der Leibniz Bioactives Cloud befähigen, Berechtigungen anhand von *Access Control Lists* und mittels verschachtelter Gruppenmitgliedschaften zu verwalten, wie man das z. B. von Windows kennt. Dabei war eine ganze Reihe von Randbedingungen zu beachten, um vertrauliche Informationen zu schützen und den dezentralen Ansatz des Projekts nicht *ad absurdum* zu führen.

## Review, Dokumentation, Testbarkeit

Wir haben die Einarbeitungsphase von Herrn Mauz für einen Review des Projekts und das Ausfüllen von Dokumentationslücken genutzt. Dabei wurden kleinere Schwachstellen aufgedeckt, die teilweise schon behoben sind, z. B. die Authentifizierung von Anfragen entfernter Knoten, oder die Problematik, Websocket-Verbindungen einer Nutzer-Session zuzuordnen.

Fortschritte hat es auch bei der Testabdeckung gegeben. Auch wenn die Erreichung vollständiger Testabdeckung für unsere verteilte Webanwendung unerreichbar sein wird, hilft jeder neue Test das Vertrauen in die Qualität der Software zu stärken. Das seit März 2017 verwendete Arquillian-Framework leistet uns dabei gute Dienste. Ein echtes UI-Testing (z.B. mit Selenium) konnten wir aufgrund des notwendigen Aufwands bislang noch nicht realisieren.

# Update Feb. / März 2018

Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

## Machine Learning

Bei der Vorhersage biologischer Aktivitäten haben wir Fortschritte gemacht. Nachdem wir zunächst mit einfachen 1024 bit Fingerprints getestet hatten, haben wir Atompaar-Deskriptoren nach Carhart et al.[1] und Kearsley et al.[2] implementiert. Wir mussten jedoch feststellen, dass die von uns untersuchten, im Paper von Dahl et al.[3] aufgeführten, PubChem-Datensätze nicht für die Vorhersage der Höhe der Aktivität geeignet sind, sondern nur eine binäre Klassifikation AKTIV / INAKTIV erlauben. Einen besseren Erfolg haben wir mit dem Datensatz Kinase SARfari aus ChEMBL[4] erzielt (siehe Diagramme). Besonders glücklich sind wir, dass unser Ansatz wesentlich besser als das in MOE[5] implementierte QSAR-Verfahren funktioniert hat.

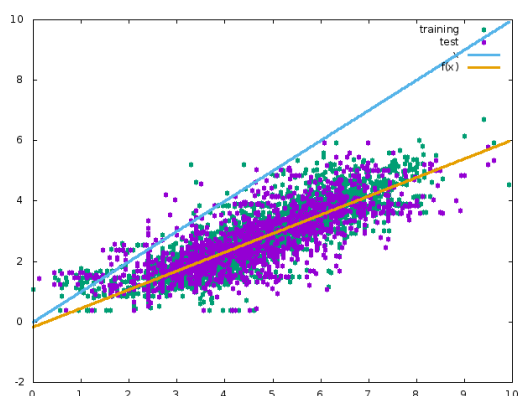


Abb. 1: tatsächliche Aktivität (x) vs. vorhergesagte Aktivität (y) für Test- und Trainingsdaten. Blaue Gerade:  $x=y$ . Beide Skalen logarithmisch.

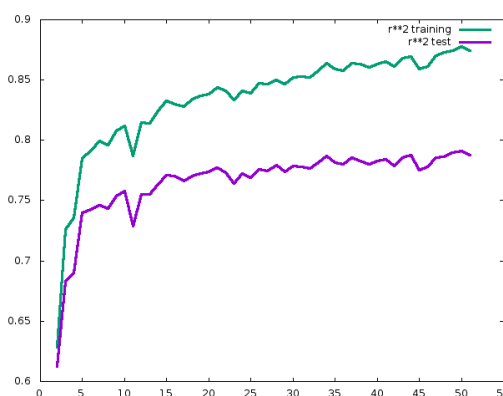


Abb. 2: Lernkurve für den Trainingsdatensatz. Im Vergleich das Bestimmtheitsmaß ( $r^2$ ) für die Testdaten (x = Anzahl Trainingsepochen, y =  $r^2$ )

## Rollout

Zahlreiche zusätzliche Verpflichtungen (Inventur, Vorträge, ...), die Grippewelle und Urlaube haben unseren Zeitplan durcheinander gebracht. Ein weiterer Schwerpunkt unserer Arbeit lag dennoch auf der Vorbereitung des Rollout. Rund ein Drittel der 34 Commits der letzten beiden Monate dienten der Fehlerbereinigung und zur Erhöhung von Wartbarkeit und Robustheit.

Das IPB betreibt seit ein paar Wochen einen von außen erreichbaren Cloud-Knoten. Herr Hinrichs vom ISAS hat Anfang April zugesagt, demnächst einen Knoten in Dortmund in Betrieb zu nehmen. Ebenso haben wir im HKI und am DSMZ für einen Testbetrieb angefragt. Wir hoffen daher, zu den **Wirkstofftagen am 23. / 24. April in Halle** einen Cluster im Testbetrieb präsentieren zu können.

## Literaturverzeichnis

- [1] Carhart RE, Smith DH, Venkataraghavan R. Atom Pairs as Molecular-Features in Structure Activity Studies - Definition and Applications. *J Chem Inf Comp Sci.* 1985;25:64-73.
- [2] Kearsley SK, Sallamack S, Fluder EM, Andose JD, Mosley RT, Sheridan RP. Chemical similarity using physiochemical property descriptors. *J Chem Inf Comp Sci.* 1996;36:118-27.
- [3] George E. Dahl NJ, Ruslan Salakhutdinov. Multi-task Neural Networks for QSAR Predictions. arXiv:14061231 [statML]. 2014.
- [4] Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* 2014;42:D1083-D90.
- [5] Chemical Computing Group ULC SSW, Suite #910, Montreal, QC, Canada, H3A 2R7. Molecular Operating Environment (MOE), 2013.08. 2018.

# Update Dez. 2017 / Jan. 2018

Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

## Lizenzierung

Nach etwas Überzeugungsarbeit haben wir es geschafft: Wir dürfen die Software des Projekts „Leibniz Bioactives Cloud“ unter der Apache 2.0 Lizenz veröffentlichen. Auf viele wichtige Fragen (z.B. Nachnutzung von Quellcode, Haftung, ...) gibt es damit eine Antwort.

## Technische Neuerungen

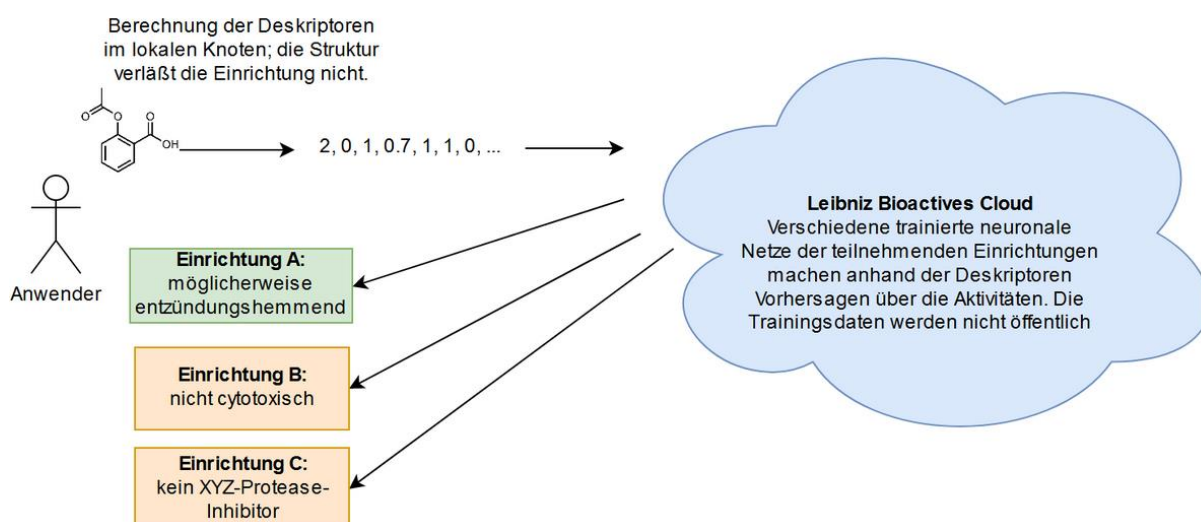
Im Dezember und Januar haben wir an folgenden technischen Neuerungen gearbeitet:

- Docker-Image für PgChem (ein Datenbankplugin für chemische Strukturen)
- „becoming smart“ – wir haben erste Experimente zu Machine Learning unternommen
- Verwaltungsfunktionen (Installation / Updates, Backup, Accounts, Collections, LDAP-Anbindung, ...)

Besonders in die zwingend benötigten Verwaltungsfunktionen ist viel Kraft und Energie geflossen, obwohl die Nutzer selten bis nie damit konfrontiert sein werden.

## Machine Learning

Wir wollen in jedem Knoten ein künstliches neuronales Netz installieren, das mit lokal vorliegenden Assay-Ergebnissen trainiert wird. Als Eingaben für das Netzwerk sollen QSAR-Deskriptoren (und Fingerprints usw.) dienen, anhand derer das künstliche neuronale Netz das Assay Outcome (z.B. Aktiv/Inaktiv, IC<sub>50</sub> o. ä.) vorhersagt, nachdem es mit bekannten Outcomes trainiert wurde. Nutzer aus anderen teilnehmenden Instituten können die QSAR-Deskriptoren für Ihre Verbindungen lokal bei sich berechnen lassen und senden die Deskriptoren an die Knoten der Cloud. Als Rückmeldung erhalten sie die Information, für welche Assays ihre Substanz möglicherweise infrage kommt.



Die größte Hürde stellt momentan die Berechnung der Deskriptoren dar, da einfache Fingerprints allein nicht geeignet sind. Wir wollen uns hierzu das Chemistry Development Kit (CDK) ansehen.

# Update November 2017

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Präsentationen

Im November durften wir unser Projekt zweimal präsentieren: Am 9. November einer Delegation vom Thailand Center of Excellence for Life Sciences (TCELS, siehe Abb.) und am 28. November den Mitgliedern des Leibniz Forschungsverbands “Wirkstoffe und Biotechnologie” – unseren Auftraggebern – auf der Jahresversammlung der Leibniz Gemeinschaft.

Die (technische) Vorbereitung beider Präsentationen beanspruchte einen großen Teil unserer Arbeitskraft.



## Technische Neuerungen

Wie im Oktober-Newsletter angekündigt, haben wir uns mit der Einführung von technischen Neuerungen im November zurückgehalten. Unsere Life-Demo auf der Mitgliederversammlung scheiterte daher auch nicht wegen eines Programmfehlers, sondern wegen Verbindungsabbruchs – wahrscheinlich weil wir unsere Internetverbindung durch den Upload mehrerer Dokumente überlastet hatten.

Als neue Funktionen haben wir unter anderem eine Zertifikatssperrliste (Certificate Revocation List, CRL), die automatische Installation von Cron-Jobs und einen Backup-Mechanismus realisiert.

## Projektwebseite

Die Projektseite der Leibniz Bioactives Cloud auf der Website des Forschungsverbands wurde aktualisiert ([http://www.leibniz-wirkstoffe.de/projects/lbac\\_project/](http://www.leibniz-wirkstoffe.de/projects/lbac_project/)). Passwortgeschützt („lbac“ - „cloud“) finden sich dort neben den gesammelten Newslettern die Projektdokumentation und Anleitungen zur Inbetriebnahme eines Knotens.



# Update Oktober 2017

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

Der Monat Oktober war für uns aufgrund von Krankheit und Urlaub ein eher kurzer Monat.

## Deployment

Wir haben erste Rückmeldungen von unseren Ansprechpartnern am HKI, am DSMZ und am ISAS bekommen. Darunter waren einige wertvolle Hinweise, wie wir in die zukünftige Entwicklung einfließen lassen werden. Der Ton war allgemein wohlwollend, so dass wir uns auf dem richtigen Weg zu befinden glauben.

Der Proxy-Container eines Knotens wird zukünftig optional sein und die Anbindung des Knotens an zentral gemanagte Proxies erlauben. Insbesondere Institute, die über ein Intrusion Detection System (IDS) verfügen, können so von einer erhöhten Sicherheit profitieren. Die Kehrseite der Medaille bei Nutzung zentral gemanagter Proxies ist natürlich der etwas größere manuelle Installationsaufwand zur Inbetriebnahme eines Knotens.

## Feature Freeze

Für die Vorstellung des Projektstands auf der Mitgliederversammlung des Forschungsverbands am 28. November 2017 in Berlin haben wir einen „Feature Freeze“ beschlossen. Im verbleibenden Monat wollen wir die angefangenen Features fertigstellen und uns der Fehlerbeseitigung widmen, damit wir in der Mitgliederversammlung eine ordentliche<sup>3</sup> Show abliefern können.

## Upload

Beim Dateiapload auch mehrerer Dateien über das JQuery haben wir die prinzipiellen Probleme gelöst. Die Webanwendung nimmt Dateien entgegen, legt sie auf einem dedizierten Speicher ab und leitet sie an Solr zur Indexierung weiter. Die Zeit der schmutzigen Hacks mit Shell-Skripten ist damit vorbei.

---

<sup>3</sup> Bitte im Sinne von redlich, wahrhaftig, echt zu verstehen.

# Update September 2017 (Draft)

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Dokumentation

Eine gute Dokumentation ist wichtig, um das Vertrauen der Administratoren zu erlangen, die für die Installation und den Betrieb der Cloud verantwortlich sind. Wir haben mit der Erstellung von Handbüchern begonnen, die alle Aspekte des Projekts umfassend beleuchten sollen. Dazu gehören die Beschreibung des Managements, der Architektur, der übertragenen Daten und der verwendeten Protokolle als auch eine Bedrohungsanalyse mit Risikobewertung und Maßnahmen zur Risikominimierung. Bislang hat das Handbuch einen Umfang von nur 25 Seiten, wir haben uns aber vorgenommen, kontinuierlich daran weiterzuarbeiten.

Außerdem haben wir die Projektseite [http://www.leibniz-wirkstoffe.de/projects/lbac\\_project/](http://www.leibniz-wirkstoffe.de/projects/lbac_project/) aktualisiert.

## Deployment

Wir haben unsere Ansprechpartner in DSMZ, HKI, ISAS und LIN gebeten, erste Tests für uns durchzuführen.

## Statistik

Im letzten Monat hatten wir rund 20 Commits in unser Git-Repository. Nach diversen Umstrukturierungen und Aufräumaktionen umfasst unser Quellcode derzeit rund 140 Dateien (zzgl. Grafiken) mit insgesamt über 13000 Codezeilen.

Die Änderungen betrafen vor allem

- Die Umstellung des User-Interface auf ein modernes, geräteunabhängiges Design (*responsive design*) mit BootFaces
- Vorbereitungen für den Upload von Dokumenten mit JQuery
- Das Management von Nutzern, Collections usw.
- Die Entsorgung diverser Altlasten
- Die Skripte für Konfiguration und Installation

# Update Juli / August 2017

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Neuer Mitarbeiter

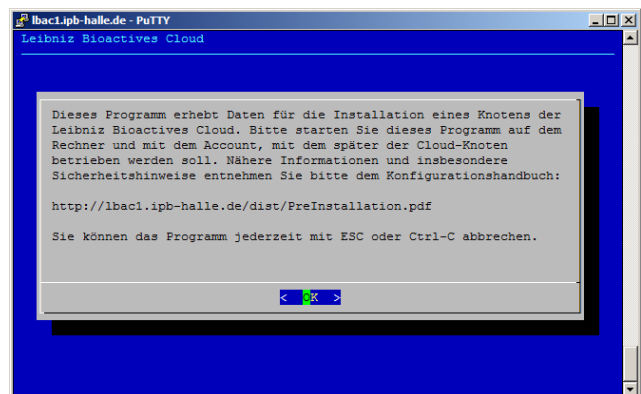
Seit 1. August verstärkt Herr Henrik Teuscher das Entwicklerteam der Leibniz Bioactives Cloud. Bedingt durch die Komplexität des Vorhabens ist er momentan noch dabei, sich in das Projekt einzuarbeiten. Trotzdem konnten wir einige Designentscheidungen der Vergangenheit bereits auf den Prüfstand stellen und werden dies sukzessive korrigieren. Größere Schwierigkeiten als erwartet bereitet uns die Einrichtung der integrierten Entwicklungsumgebung (IDE): Nachdem sich die kostenlose IDE Eclipse als zu instabil erwiesen hat, wechselt Herr Teuscher jetzt zu IntelliJ IDEA.

## Dokumenten-Upload

Für den Upload von Dokumenten werden wir zunächst auf eine JQuery-basierte Lösung setzen. Das WebDAV-Protokoll und damit auch die Apache Jackrabbit Bibliothek werden damit erst einmal nicht benötigt. Falls zu einem späteren Zeitpunkt die Implementierung eines Java Content Repository (JCR) zur Dokumentenverwaltung notwendig wird, liegen erste Vorarbeiten in unserem Git-Repository.

## Deployment

Fortschritt hat es bei der Konfiguration und Installation eines Knotens der Leibniz Bioactives Cloud gegeben. Der gesamte Prozess erfolgt skriptgesteuert, so dass der manuelle Aufwand für die Inbetriebnahme eines Knotens minimiert wird. Durch die Verwendung von NCurses-Dialogen wird die Einstieghürde bewusst niedrig gelegt. Um erste Prototypen noch dieses Jahr ausliefern zu können, haben wir unsere Fühler zum DSMZ ausgestreckt.



## Dokumentation

Parallel haben wir unsere Infrastruktur auf die Produkte der Firma Atlassian umgestellt. Anstelle der bisherigen Quellcode-Verwaltung Subversion verwenden wir jetzt das Git-basierte Produkt Bitbucket. Das Issue-Tracking und das Wiki wurden von Trac auf Confluence und JIRA umgestellt. Mit Bamboo verfügen wir jetzt erstmalig über eine Lösung für Continuous Integration. Die Vorteile der kommerziellen Lösungen liegen in höherer Nutzerfreundlichkeit und gesteigerter Effizienz. Außerdem kann die Erstellung von Handbüchern mit LaTeX entfallen, da Confluence den Export nach Word und PDF erlaubt.

# Update Juni 2017

---

*Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)*

## Docker / Containerisierung

Die einzelnen Komponenten des Cloud-Prototyps wurden weitgehend erfolgreich in Container aufgeteilt. Nach ersten Experimenten mit Docker Swarm wurde entschieden, dass die Leistung eines einzelnen Docker-Hosts vorerst ausreichend sein wird. Die Containerisierung fand deshalb mit Docker Compose statt, wodurch die Gesamtkomplexität (Wegfall der Komponenten verteilter Speicher, Apache Zookeeper, ...) deutlich reduziert werden konnte. Der bislang eingesetzte Web-Proxy (Apache HTTPD) ist im Rahmen der Containerisierung vorerst entfallen. Ebenso gibt es momentan noch keinen Ersatz für den automatischen (aber primitiven) Dokumenten-Upload via CIFS.

## Sicherheit

Mit der Containerisierung wurden auch die Kommunikationspfade sauber getrennt, so dass die Kommunikation mit dem Backend nur noch nach gegenseitiger zertifikatsbasierter Authentifizierung möglich ist. Auch die Kommunikation mit dem User Interface (UI) wird zukünftig verschlüsselt erfolgen. Hierfür sollen möglichst offiziell signierte Zertifikate zum Einsatz kommen, die von allen Browsern verifiziert werden können. Möglicherweise müssen die Knoten zukünftig wieder um einen Web-Proxy (evtl. nginx) ergänzt werden; z.B. um effizienter statische Inhalte auszuliefern und die Nutzer auf die verschlüsselten Seiten umzuleiten.

## Dokumenten-Upload

Aufgrund von Vorbehalten des HKI gegen das CIFS-Protokoll wurde nach Alternativen gesucht. Favorisiert wird momentan der Upload über WebDAV, da dieses Protokoll einige Vorteile bietet und durch alle relevanten Betriebssysteme unterstützt wird. Insbesondere kann der fragile Indexierungsmechanismus beim CIFS-Upload, der auf der Auswertung des Samba-Audit-Protokolls beruhte, durch eine robustere Alternative ersetzt werden. Ein weiteres wichtiges Kriterium ist die Eignung für den (beschränkten) Massupload von Dateien. Die Verwendung offiziell signierter Zertifikate im UI würde auch den WebDAV-Upload über den Windows Explorer ermöglichen. Momentan werden die Bibliotheken Milton.io und vor allem Apache Jackrabbit bzw. Apache Jackrabbit Oak auf ihre Eignung für unser Vorhaben geprüft.

## Dokumentation

Das Projekt Leibniz Bioactives Cloud ist jetzt auf der Homepage des Forschungsverbunds vertreten ([http://www.leibniz-wirkstoffe.de/projects/lbac\\_project/](http://www.leibniz-wirkstoffe.de/projects/lbac_project/)). Zukünftig wollen wir dort unseren Projektfortschritt dokumentieren und technische Unterlagen für die IT-Abteilungen der beteiligten Institute bereithalten.

## Urlaubszeit

Der nächste Newsletter wird voraussichtlich Ende August 2017 erscheinen.

# Update April / Mai 2017

Leibniz Bioactives Cloud – Frank Broda (fbroda@ipb-halle.de)

Seit den Wirkstofftagen in Freising hat das Leibniz-Bioactives-Cloud-Projekt Fortschritte gemacht:

## Test Framework

Es gelang, das Test-Framework Arquillian in das Projekt zu integrieren und insbesondere erste Tests unter Beteiligung der Java Persistence API zu etablieren. Die Verfügbarkeit automatischer Tests ist wichtig, um sich dem Paradigma der testgetriebenen Entwicklung (**test driven development**) anzunähern. Im Idealfall entlasten diese automatischen Tests die Entwickler von der Durchführung trivialer Funktionstests und zeigen frühzeitig Probleme bei der Interoperabilität von Komponenten auf.

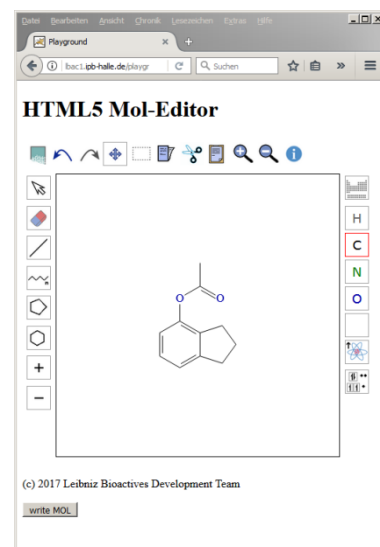
## Sicherheit

In Vorbereitung des Deployments erster Cloud-Testknoten wurde mit der Verschlüsselung der Netzwerkverbindungen und der gegenseitigen zertifikatsbasierten Authentifizierung der Cloud-Knoten experimentiert. Darüber hinaus wurde Kontakt mit Frau Matthies vom HKI Jena hergestellt. Wie erwartet, gibt es berechtigte Sicherheitsbedenken, die aktiv adressiert werden müssen. Aufgrund der Hinweise von Frau Matthies wird aktuell die Containerisierung der Komponenten mittels **Docker** vorangetrieben. Die Containerisierung wird zudem ermöglichen, bei der Entwicklung und dem Betrieb der Cloud der **DevOps**-Philosophie zu folgen. Die Kommunikation(-swege) in und mit der Cloud wurde(n) präzisiert, dies wird jedoch durch die Containerisierung in nächster Zeit ein Update erfahren.

## MolEditor

Die Leibniz Bioactives Cloud soll perspektivisch chemische Strukturformeln verarbeiten können. Dies erfordert die Fähigkeit, Strukturen anzuzeigen und zu editieren. Mit der Deaktivierung des JavaPlugins in praktisch allen Webbrowsern stehen einige bislang verfügbare Editoren nicht mehr zur Verfügung. Der leistungsfähige JavaScript basierte Editor MarvinJS der Firma ChemAxon darf kostenlos nur in öffentlich frei zugänglichen Web-Angeboten benutzt werden. Es gibt noch eine Reihe weiterer kostenloser Editoren, die jedoch von unterschiedlicher Qualität sind. Mit dem nebenstehend abgebildeten Editor wird dem Projekt ggf. ein eigener Editor zur Verfügung stehen, der u.a. eine History-Funktion und die Übertragung von Strukturen mittels Copy-Paste aus ChemDraw unterstützt.

Dieses Teilprojekt umfasst derzeit ca. 3000 Zeilen Quelltext und ist noch nicht abgeschlossen (u.a. fehlen noch Funktionen wie Templates (z.B. 5- und 6-Ringe), Stereochemie, Snap-In auf Gitterpunkte). Die Fertigstellung des Editors hat momentan jedoch keine Priorität.



## Offene Fragen

Das Projekt braucht Kontakt zu IT-lern, die die Teststellung erster Cloud-Knoten aktiv begleiten wollen. Idealerweise sollten die Kontaktpersonen über den Forschungsverbund und das Vorhaben Bioactives Cloud grob orientiert sein. Frau Finsterbusch-Kettner stellt dazu gerade Unterlagen zusammen.