

Big Data im Labor

Ein Update

F. Hauer

Eines der meistgehörten Schlagwörter der letzten Monate in allen Branchen ist 'Big Data' – auch die Laborumgebung bildet da keine Ausnahme. Doch je öfter der Begriff fällt, desto weniger kann man sich sicher sein, was damit eigentlich gemeint ist und welche Implikationen für den eigenen Wirkungsbereich zu erwarten sind.



© U.S. Fish and Wildlife Service Southeast Region



Bei aktuellen Geräteanschaffungen im Labor wird klar, dass die Menge und die Struktur der Daten, die diese Geräte produzieren, an Umfang und Komplexität zunehmen

Die Lockungen der omnipräsenten Datenverfügbarkeit sind offensichtlich. Gerade in Laboren nimmt die Datenmenge stetig zu, und nicht immer sind vor Ort Lösungen zur Hand, mit deren Hilfe man die Datenberge erklimmen um vom Ausblick auf die Informationsströme gewinnbringende Schlüsse ziehen lassen. So bleibt oft die Frage im Raum: Was ist Vision, was ist Realität – und was erwartet mich?

Zeit also für ein Update: Welche aktuellen Entwicklungen finden statt, und welche Trends sind in naher Zukunft zu erwarten?

Aktuelle Trends

Bei aktuellen Geräteanschaffungen im Labor wird klar, dass die Menge und die Struktur der Daten, die diese Geräte produzieren, an Umfang und Komplexität zunehmen. Es lässt sich allerdings auch beobachten, dass die Schnittstellen von Geräten immer versatiler werden: Die Entwicklung von Augabeformaten geht längst über die Kompatibilität mit Standardprogrammen wie MS Office, PDFs und annotierten Bildern hinaus. Die neuesten Hardwaregenerationen sind in der Lage, Webservices anzubieten und mit Filesystemen und Datenbanken bis hin zu LIMS direkt zu kommunizieren.

Auch die Benutzeroberflächen orientieren sich zunehmend an Standards ausserhalb des Labors, und sind mehr und mehr inspiriert vom Design und von der Usability von mobilen Anwendungen. Die direkte Anbindung an mobile Apps ist ebenfalls ein Konzept, das bei Nutzern zunehmend auf Resonanz stößt. Eine Vielzahl von

Anbietern stellen schon heute mobile Apps bereit, die Forscher im Labor bei Planung, Durchführung, Dokumentation und Auswertung ihrer Laborprozesse unterstützen – und es ist zu erwarten, dass dieses Angebot weiter wachsen wird.

Eine wichtige Voraussetzung für den Austausch von Daten zwischen Geräten, Softwarelösungen und Nutzer stellen dabei Standards zur Strukturierung von Daten dar. Nachfolgend sollen daher einige Verbundprojekte vorgestellt werden, die sich mit der Prüfung und Weiterentwicklung von Standards für das Labor befassen.

Die Allotrope Foundation

... ist ein Zusammenschluss von großen forschenden Pharmaunternehmen, die sich zusammen mit Netzwerkpartnern aus der Softwareindustrie das Ziel gesetzt haben, ein Framework bereitzustellen, das die Grundlage bilden soll für ein intelligentes, automatisiertes Analytiklabor. Innerhalb dieses Frameworks werden Metadatenbibliotheken bereitgestellt, die eine einheitliche Annotation von Daten und Organisationen erlauben. Dabei konzentriert sich die Allotrope Foundation insbesondere auf die Evaluation von bestehenden Standards und deren Integration in das eigene Framework.

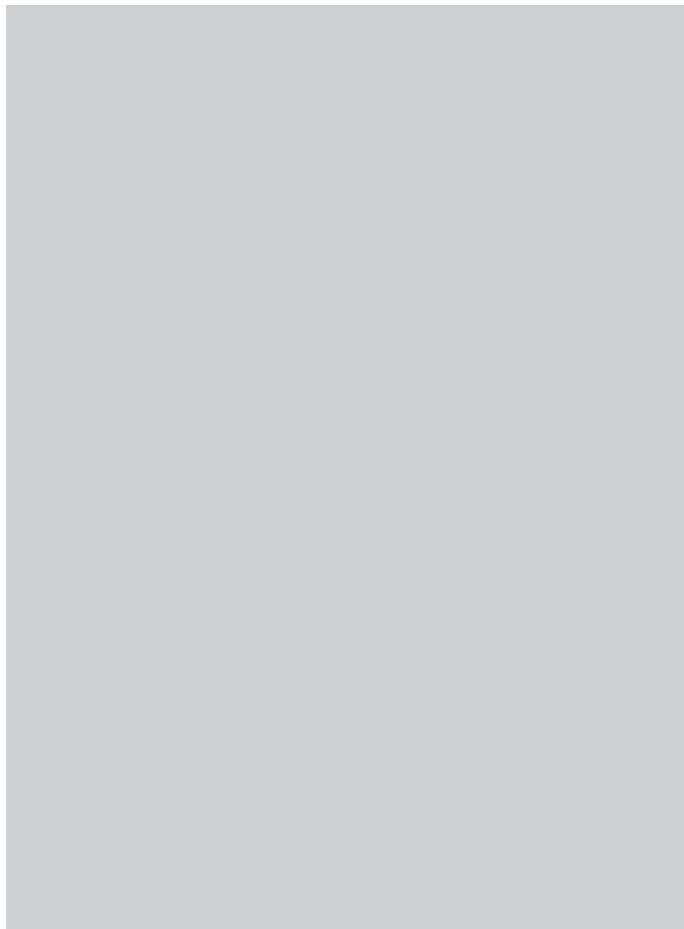
Die Research Data Alliance

... ist eine Organisation aus akademischen Forschungseinrichtungen, Bibliotheken, Datenzentren und wissenschaftlichen Verlagen, die es sich zum Ziel gemacht hat, die Verbreitung und gemeinsame Nutzung von Forschungsdaten zu fördern. Zu

diesem Zweck hat die Research Data Alliance internationale Arbeitsgruppen gebildet, die gemeinsam Tools, Best Practices und Produkte entwickeln sollen, die eine breite Anwendung in Forschungsgruppen, Projekten und Organisationen finden sollen. Aufgaben sind dabei insbesondere die Harmonisierung von Datenstandards, die Förderung des Austausches von Daten, die Schaffung von Möglichkeiten für eine bessere Durchsuchbarkeit von Daten und weiterentwickelte Möglichkeiten zum Datenmanagement und zur langfristigen Datenarchivierung.

Die Pistoia Alliance

... ist ein Verbund aus Pharmaunternehmen und Software- und Geräteherstellern mit der Mission, Innovationsbarrieren in der Forschung abzubauen über einen offenen Austausch der prekompetitiven Grundlagenforschung. Dazu haben die Mitglieder der Pistoia Alliance offene Software-Frameworks entwickelt, die einen Austausch von Daten und Erkenntnissen vereinfachen. Dazu gehört zum Beispiel die Initiation der transSMART Foundation, die eine Softwareplattform für das Wissens- und Datenmanage-





ment und den Datenaustausch in der angewandten Medizinforschung ermöglicht. Weitere Beispiele für Plattformen, die innerhalb der Pistoia Alliance verwirklicht worden, sind z.B. the Analysome zur Beschleunigung der Wirkstoffforschung, die Controlled Substance Expert Community (CSCS) und die Hierarchical Editing Language for Macromolecules, die eine An-

notation von Makromolekülen in chemischen Datenbanken erlaubt.

Im SiLA Rapid Integration Consortium

... haben sich Hersteller von Laborgeräten und -Software zusammengeschlossen, die aktive Entwicklung von Kommuni-

kationsprotokollen für Laborgeräte voranzutreiben. Diese Entwicklung umfasst die Implementierung von Standards für Geräteinterfaces und für das Datenmanagement von Messdaten für die schnelle Integration von Laborhardware in ein automatisiertes Anwendungsumfeld. Eine Harmonisierung der Datenströme wird hier erreicht durch die Bereitstellung eines ‚Common Command Sets‘, durch das alle mit dem Standard kompatible Endgeräte in einem einheitlichen Format angesteuert werden können. Dadurch wird nicht nur die Steuerung von verschiedenen Geräteklassen (vom Pipettierroboter über die Waage bis hin zum Spektrometer), sondern auch die automatisierte Auslesung von Meßergebnissen und Wartungszuständen ermöglicht. SiLA-kompatible Geräte können daher schnell und ohne Anpassungsaufwand in ein SiLA-kompatibles Prozessmanagementsystem (PMS), zum Beispiel auch über ein den Standard unterstützendes LIMS, integriert werden. Als Ergebnis der Aktivitäten des SiLA-Kon-

sortiums steht bereits heute eine wachsende Anzahl von kompatiblen Gerätetreibern und Steuerungssoftware zur Verfügung.

Wie geht es weiter?

Wenngleich diese Auswahl die Beschreibung der Biosphäre der Konsortien und Verbände, die sich mit der Bereitstellung und Evaluation von Standards befasst, keinesfalls erschöpft, bildet sie dennoch eine repräsentative Stichprobe, aus der sich die aktuelle Entwicklung ablesen lässt. Dabei fällt ins Auge, dass eine Vielzahl der Verbände nicht nur gemeinsame Mitglieder, sondern auch überlappende Zielsetzungen haben. Es stellt sich also die Frage, welche Trends und Mechanismen entschieden werden, welche Initiativen und Standards sich am Markt und in der Anwendung durchsetzen werden.

Zu den internen Faktoren, die für eine breite Adaption sorgen werden, gehört sicherlich die offene Verfügbarkeit der entwickelten Tools und Inf-



rastrukturen und die Aktivität der Verbundmitglieder an der Entwicklung. Hier überzeugen insbesondere das SiLA-Konsortium und Entwicklungen, die aus der Pistoia Alliance hervorgehen: einfache und unkomplizierte Verfügbarkeit und eine gute Dokumentation sowie eine bereits an einigen Stellen erfolgreiche Implementierung werden für eine weitere Verbreitung sorgen.

Als externe Faktoren spielen insbesondere die Bedürfnisse des Marktes eine große Rolle. Aufgrund des wachsenden Bedarfes an einer einheitlichen Dateninfrastruktur zwischen Elementen der Laborhardware im Hinblick auf eine fortschreitende Automatisierung von Laborprozessen hat hier das SiLA-Konsortium insbesondere aufgrund der in der Praxis bereits demonstrierten Anwendbarkeit des Entwicklungsansatzes gute Chancen, zu

einem internationalen Standard beizutragen, der in nicht allzu ferner Zukunft obligatorischen Charakter entwickeln könnte.

Ein weiterer Trend, der allgemein in der Entwicklung von Software zu beobachten ist, begünstigt die Entwicklung von Schnittstellen zwischen verschiedenen Softwareprodukten, Datenbanken und Geräten: Für die Entwicklung von Softwarelösungen ist die Abkehr von monolithischen Strukturen hin zu offenen Plattformen zu einem unumgänglichen Imperativ geworden (siehe auch GIT Ausgabe 8/2015: „Vom Monolith zur Plattform“). Mit Schnittstellen, die einen Austausch von Daten zwischen Plattformen erlauben, ist ein weiterer wichtiger Schritt in die Verwirklichung der Visionen von Big Data gemacht: ein ungehinderter Datenfluss zwischen Softwarelösungen aus allen Bereichen des

Labors wie Probeneingang, Analyse und Reporterstellung wird anbieterübergreifend möglich – und damit auch bedeutend preisgünstiger.

Eine weiterhin bestehende Herausforderung verbleibt in der Harmonisierung von Forschungsdaten und analytischen Daten, die im Labor erzeugt werden. Zwar erfreuen sich in dieser Domäne hierarchisch strukturierte Dateiformate (z.B. in Adaptionen des XML-Formates beispielsweise durch AniML) wachsender Beliebtheit, die Komplexität der Annotationen stellt aber viele Teilnehmer immer noch vor ernste Ressourcenprobleme. In naher Zukunft wird sich also zeigen, ob eine Annotation bereits auf Hardwareebene hier den Aufwand mindern kann, oder ob andere, nicht-hierarchische Systeme als Alternative an Attraktivität und Applikabilität gewinnen können. Eine generelle Voraussetzung für weitere Entwicklungen auf der Ebene der Datenverarbeitung wird es zunächst sein, durch die Interkompatibilität von einzelnen Elementen einer bis dahin noch heterogenen Softwareinfrastruktur entweder durch geeignete Schnittstellen oder durch die Bereitstellung einer übergeordneten Plattform eine weitere Harmonisierung auf der Ebene der Datenverarbeitung und -Analyse zu erreichen. Mit jedem Schritt, mit dem wir diesem Ziel näher kommen, wachsen die Möglichkeiten von Big Data im Labor.

KONTAKT |

Dr. Florian Hauer
COO
Labfolder GmbH
Berlin
fh@labfolder.com
www.labfolder.com



Weitere Beiträge zum Thema:
<http://bit.ly/LIMS-GIT>



Mehr Informationen:
<http://bit.ly/Allotrope>
<http://bit.ly/GIT-Alliance>
<http://bit.ly/Pistoia>
<http://bit.ly/GIT-Sila>
<http://bit.ly/Transsmart>