



Peter Kunszt kennt die Herausforderungen bei der Forschung mit Big Data.

Peter Kunszt leitet SyBIT, das Bioinformatik- und IT-Projekt von SystemsX.ch

«SyBIT macht die Forschenden fit für Big Data»

Längst gilt Big Data auch in der Systembiologie als wertvolle Informationsquelle. Doch wichtige Erkenntnisse über die Abläufe in biologischen Systemen lassen sich nur gewinnen, wenn es gelingt, die relevanten Informationen aus der Datenflut herauszufiltern. Peter Kunszt und sein Team unterstützen die Forschenden dabei, solche Prozesse zu automatisieren. Mit dem IT-Supportprojekt SyBIT tragen sie dazu bei, dass die Schweizer Forschenden auch langfristig an der Weltspitze mithalten können.

Woher stammt die Datenflut in der Systembiologie?

Systembiologische Forschung wurde erst möglich durch grosse Entwicklungsschübe in den Technologien, mit denen sich biologische Systeme beobachten lassen. Beträchtliche Fortschritte wurden beispielsweise bei Apparaturen zur DNA-Sequenzierung erzielt. Aber auch in der Massenspektrometrie und jüngst in der Mikroskopie ist die Entwicklung beeindruckend. Alle diese Geräte produzieren immer mehr Daten. Das lässt sich mit den immer leistungsfähigeren Digitalkameras vergleichen. Jedes Jahr erscheinen neue Modelle mit noch mehr Megapixeln und man benötigt neue Speichermedien mit immer mehr Kapazität. In der Forschung gilt dasselbe, aber in ganz anderen Dimensionen. Deshalb sprechen wir auch von Big Data.

Was bedeutet dieser Begriff für Sie?

Neben der Menge zeichnet sich Big Data, insbesondere in der Systembiologie, durch die Komplexität der Daten aus. Oft hat man verschiedene Arten von Messwerten, bei denen erst geklärt werden muss, wie sie zusammenhängen. Solche Daten lassen sich deshalb nicht leicht interpretieren. Ein weiterer Aspekt ist die Geschwindigkeit. Denn wenn man von grossen Mengen an kom-

plexen Daten ausgeht, wird es auch immer schwieriger, diese in einem nützlichen Zeitrahmen auszuwerten. Schliesslich kommt noch die Datenunsicherheit dazu. Damit ist die Qualität und Zuverlässigkeit der Informationen gemeint. Denn die Werte können auch durch Messfehler verfälscht sein.

Welcher dieser Aspekte stellt die grösste Herausforderung dar?

Menge, Komplexität und Geschwindigkeit spielen immer zusammen. Unsere Aufgabe ist es, unter Berücksichtigung dieser drei Faktoren gemeinsam mit den Forschenden die besten Lösungen zu finden, die ihre wissenschaftlichen Fragestellungen mit den heute verfügbaren Techniken rasch beantworten.

Können Sie das konkretisieren? Wie unterstützt SyBIT die SystemsX.ch-Forschenden genau?

Das ist bei jedem Projekt verschieden. Je nach Bedarf der Forschenden helfen wir, die nötige Hard- und Software aufzubauen, oder wir organisieren den Zugang zu Grossrechnern. Wir unterstützen die Wissenschaftler aber auch bei der Analyse und Auswertung sowie bei der Verwaltung und Speicherung ihrer Daten.



Mit dem Projekt SyBIT stellen wir das gesamte Know-how rund ums Datenmanagement zur Verfügung und helfen, die einzelnen Schritte zu automatisieren und effizient zu gestalten. Wir machen die Forschenden sozusagen fit für Big Data, damit sie das Potenzial der neuen Technologien voll ausschöpfen können.

Ist der Bedarf für IT-Support gestiegen?

Ja, ganz klar. Mit der wachsenden Menge und der steigenden Komplexität der Daten wird es für die Forschenden immer schwieriger und zeitintensiver, ihre Daten zu managen. Man stellt sich das zwar einfach vor, aber bei sehr grossen Datenmengen, wie sie beispielsweise bei der Massenspektrometrie entstehen, ist bereits das Erfassen aufwendig. Die Daten aus verschiedensten Experimenten müssen richtig annotiert und strukturiert abgelegt werden. Nur so lassen sie sich auch später noch den richtigen Projekten zuordnen und bei Bedarf reproduzieren. Für die Analyse und Auswertung der riesigen Datenmengen sind oft Algorithmen nötig, welche die Merkmale und die Muster, die von Interesse sind, automatisch auffinden.

Können Sie ein konkretes Beispiel nennen?

Aktuell unterstützen wir MorphogenetiX. In diesem Projekt beobachten die Forschenden die Spezialisierung von Zellen mithilfe der 3D-Mikroskopie. Dank dieser neuen Technologie müssen Proben nicht mehr in Formalin fixiert, sondern können lebend gefilmt werden. Das 3D-Mikroskop nimmt dabei bis zu 700 Bilder pro Sekunde auf. So können die Wissenschaftler verfolgen, wie sich die Zellen teilen, und schliesslich aufzeigen, wie beispielsweise eine spezialisierte Hirnzelle entsteht.

Die so generierten Datenmengen sind riesig. SyBIT unterstützt die Forschenden von MorphogenetiX deshalb bei der rechnergestützten Auswertung der Daten. Einer meiner Mitarbeitenden ist für mehrere Monate fix vor Ort und testet zusammen mit den projekteigenen Spezialisten die entwickelten Algorithmen, um die Auswertung der riesigen Datenmenge zu automatisieren.

Grosse Datenmengen erfordern viel Speicherkapazität. Was wird gespeichert, was nicht?

Um zu entscheiden, was relevant ist, müssen wir die Daten extrem gut kennen. Deshalb geht es bei all dieser datenlastigen Forschung mit neuen Technologien zunächst einmal darum, die Daten überhaupt zu verstehen sowie Muster und Zusammenhänge zu erkennen. Insbesondere in der Grundlagenforschung weiss man anfangs häufig gar nicht, was alles in den Daten steckt. Deshalb möchten die Wissenschaftler meist alles behalten. Oft wird erst gegen Ende eines Projekts klar, welche Daten relevant sind, beziehungsweise welche Daten gelöscht werden können, weil sie sich später einfach und sogar präziser neu generieren lassen.

Und wie stellen Sie sicher, dass die Daten auch in Zukunft zugänglich sind?

Leider ist die langfristige Datensicherung noch ein weitgehend ungelöstes Problem. Für die Forschungsbereiche Genomik und Proteomik gibt es zwar bereits internationale, zentrale Datenbanken, doch deren Finanzierung ist noch nicht langfristig gesichert. Für die Archivierung von Daten aus bildgebenden Verfahren etwa existiert noch gar keine Lösung. Nach dem Auslaufen von SyBIT gibt es keine Institution, die solche Daten weiterverwaltet.

Wer ist denn Ihrer Meinung nach dafür zuständig?

Aus meiner Sicht wäre die Datenarchivierung Aufgabe der Bibliotheken. Es kann nicht sein, dass Forschende dafür zahlen müssen, dass ihre Daten weiterhin aufbewahrt werden. Hier braucht es Lösungen des Staates. Doch das Problem wurde zum Glück bereits erkannt. Zurzeit werden verschiedene Optionen geprüft und auch politisch diskutiert.

SyBIT endet zusammen mit SystemsX.ch 2018.

Wie wird der IT-Support danach gewährleistet?

Durch lokale Supportgruppen. Die Idee stammt aus der Region Arc Lémanique. Dort wurde bereits 2004 die Gruppe «Vital-IT» aufgebaut, die Rechenleistung, Speicher und Bioinformatik-Support bietet. Nach diesem Vorbild konnten wir an den Universitäten Zürich und Basel sowie an der ETH Zürich lokale SyBIT-Partner etablieren. So bleibt die Unterstützung der Forschenden auch über das Ende von SyBIT und das Auslaufen von SystemsX.ch hinaus gesichert.

Und wie gelingt es, das Know-how in der Community zu verankern?

Erfreulicherweise können wir bereits heute beobachten, dass SystemsX.ch-Forschende das aufgebaute Wissen auch in neuen Projekten anwenden. Da die Spezialisten der Supportteams nicht nur SystemsX.ch-Projekte unterstützen, wird das Know-how auch in andere Forschungsgruppen weitergetragen.

Sind die Hochschulen somit für die datenintensive Forschung nach dem Auslaufen von SyBIT gerüstet?

Im Prinzip ja. Die lokalen Supportgruppen sind als Bioinformatik-Services an den Hochschulen fest verankert und das Schweizerische Institut für Bioinformatik übernimmt nach dem Auslaufen von SyBIT deren Koordination. Zudem haben wir in den letzten Jahren die Partnerinstitutionen von SystemsX.ch dabei unterstützt, die nötige IT-Infrastruktur aufzubauen. Nun gilt es, die lokalen IT-Ressourcen untereinander zu vernetzen, damit die Hochschulen auch gegenseitig von den Services und der spezialisierten Infrastruktur profitieren können. Dies trägt dazu bei, dass die Schweizer Forschenden in der Systembiologie auch langfristig an der Weltspitze mit dabei sind.