



Adobe Stock / goodluz

Früh zur Datenkompetenz

An der Universität Erlangen-Nürnberg erlernen Physikstudierende schon im Bachelorstudium Kompetenzen der Datenverarbeitung.

Michael Krieger, Heiko B. Weber und Christopher van Eldik

Daten gelten als entscheidende Ressource des 21. Jahrhunderts. In der Physik gibt es eine große Bandbreite strukturierter digitaler Datenströme aus Großexperimenten in Astronomie oder Teilchenphysik, aber auch hochgradig individuell geprägte Datenaufnahme in Laborexperimenten in der Festkörperphysik oder Optik. Wären all diese Forschungsdaten systematisch erfasst und offen zugänglich, ließen sich daraus neue Erkenntnisse gewinnen. Dafür sind neben den Rohdaten auch die Metadaten bedeutsam, die Messaufbau, experimentelle Parameter, Einheiten oder Kontexte beschreiben. Um eine systematische Datenerfassung zu etablieren, gilt es, entsprechende Kompetenzen schon früh im Studium zu vermitteln.

Als Leitgedanke eines Forschungsdatenmanagements dienen die FAIR-Prinzipien: Daten sollen auffindbar (findable), offen zugänglich (accessible), standardisiert und dialogfähig (interoperable) sowie wiederverwertbar (reusable) sein. Solcherart aufbereitete Daten könnten es zukünftig erlauben, mit Computern möglicherweise völ-

lig neue Zusammenhänge zu entdecken. Bund und Länder haben mit der groß angelegten Initiative für eine Nationale Forschungsdateninfrastruktur (NFDI)¹⁾ erhebliche Fördermittel bereitgestellt, um Konsortien aus verschiedenen wissenschaftlichen Disziplinen diesbezügliche Konzepte ausarbeiten zu lassen.

Doch vermitteln unsere Physikstudiengänge überhaupt die passenden Kompetenzen für die Datenverarbeitung? In der Physik besteht breiter weltweiter Konsens, dass die Studierenden fundierte Kenntnisse in Analysis, linearer Algebra etc. haben müssen, um physikalische Konzepte zu verstehen. Ein solcher Konsens bezüglich Datenkompetenzen scheint dagegen nicht zu existieren. Beim Erstkontakt mit Daten, etwa in den Praktika der Experimentalphysik, üben die Studierenden zwar elementare Datenevaluation und systematische Dokumentation, diese genügen aber

1) www.nfdi.de

2) numpy.org; scipy.org und matplotlib.org

3) jupyter.org

nicht den steigenden Ansprüchen der Forschung und der Berufspraxis, die sich zunehmend mit der Analyse großer Datenmengen befassen. Häufig scheitert es schon an einem Konsens über eine geeignete Programmiersprache.

Am Department Physik der Friedrich-Alexander-Universität Erlangen-Nürnberg haben wir in den letzten Jahren das Curriculum angepasst und die Vermittlung von Datenkompetenz früh im Bachelorstudium platziert. Daraus ergeben sich erhebliche Vorteile für den weiteren Studienverlauf. Die Erfahrung hat gezeigt, dass die Studierenden in puncto Datenkompetenz schnell auf die Überholspur gehen und sich in den Physik-Arbeitsgruppen als treibende Kräfte erweisen, um ein modernes Forschungsdatenmanagement zu etablieren.

Praxis und Praktikum

Wie führt man Physikstudierende möglichst früh an computergestützte Datenauswertung heran? Wie lässt sich der Herausforderung begegnen, dass manche Studierende bereits Programmiererfahrung mitbringen, für andere das Thema aber wie ein Buch mit sieben Siegeln erscheint?

Nach unserer Erfahrung gestaltet sich das Erlernen einer Programmiersprache am einfachsten durch begleitetes Learning-by-doing anhand physikalischer Fragestellungen. Die intensive eigene Beschäftigung in Form kleinerer Programmieraufgaben fördert ein tiefes Verständnis und legt die Grundlage für die Herausforderungen in der Datenauswertung. Das Bearbeiten von Aufgaben in Teams stärkt den gerade zu Beginn des Studiums so wichtigen Austausch mit den Mitstudierenden und fördert den Aufbau einer wissenschaftlichen, lösungsorientierten Diskussionskultur. Dabei dürfen aber die Inhalte des Fachstudiums nicht in den Hintergrund treten. Nur wenn der Anwendungsbezug klar ersichtlich ist, sind die Studierenden motiviert, das Gelernte im weiteren Studienverlauf zu nutzen.

Seit dem Wintersemester 2018/19 gibt es im Bachelor-Studiengang daher ein verpflichtendes Computerpraktikum „Datenverarbeitung in der Physik“ (DV-Praktikum). Dort bearbeiten die Studierenden in Teams unter Anleitung von Tutor:innen Programmieraufgaben und reichen diese elektronisch in Form eines fertigen Programms ein. Ein interaktiver Onlinekurs, der die Grundlagenkenntnisse zur Bearbeitung der nächsten Praktikumseinheiten vermittelt, begleitet das wöchentliche zweistündige DV-Praktikum. Die Kursinhalte beschränken sich nicht auf das Erlernen einer Programmiersprache. Das Hauptaugenmerk liegt darauf, grundlegende Kompetenzen für Datenverarbeitung und -analyse zu vermitteln. So enthält der Kurs einführende Module in Messunsicherheiten und Fehlerfortpflanzung, Interpolation, Fitten von Daten, Statistik-Funktionen und Monte-Carlo-Simulationen, Fourier-Transformationen, numerische Integration, Lösung von Differentialgleichungen sowie grafische Aufbereitung von Ergebnissen. Auch ein Kapitel zu „Guter wissenschaftlicher Praxis“ gehört dazu. Diese Aspekte sind selbst für computeraffine Studierende neu und fordern damit alle gleichermaßen.

Als Programmiersprache nutzen wir Python, da die Grundlagen auch für Anfänger schnell zu erlernen sind.

Zudem erleichtern es gute Python-Kenntnisse, sich andere gängige Sprachen wie C++ oder Java anzueignen. Wichtiger noch für unsere Zwecke: Neben der umfangreichen Python-Standardbibliothek stehen weitere gut kuratierte Basis-Bibliotheken zur numerischen Datenverarbeitung, Statistik und grafischen Aufbereitung frei zur Verfügung – nämlich numpy, scipy und matplotlib.²⁾ Mit den verwendeten Jupyter-Notebooks³⁾ lässt sich Python-Code innerhalb eines Webbrowser-Fensters verarbeiten. Dabei werden Python-Code und Ergebnisse bestehend aus Text und Grafiken gemeinsam in einem Dokument dargestellt und gespeichert. Dies erlaubt es, diese leicht mit anderen zu teilen. Jupyter-Notebooks verfügen zudem über umfangreiche (Text-)Formatierungswerkzeuge, inklusive Formelsatz, Tabellen und Einbindung von Bildern (**Abb. 1**).

Nach einer sehr erfolgreichen Pilotphase haben wir das DV-Praktikum auf Wunsch der Studierenden im ersten Fachsemester verankert und das Grundpraktikum daher in das zweite Semester verlagert. Dies bietet den Studierenden die perfekte Gelegenheit, die erworbenen Datenverarbeitungs-Kompetenzen im gesamten Studium einzusetzen.

Spielerischer Zugang zur Physik

Durch die Einbindung des DV-Praktikums in das Erlanger Curriculum der Physik können Lehrende in Vorlesungen, Übungen und Praktika auf diese Kompetenzen zurückgreifen. Mit jeder Anwendung vertiefen sowohl Studierende wie auch Lehrende ihre Kenntnisse. Das ermöglicht völlig



Abb. 1 Jupyter-Notebooks stellen Python-Code, Dokumentation und Grafiken auf einer Ebene dar, hier eine Analyse zur Bestimmung der Fallbeschleunigung.

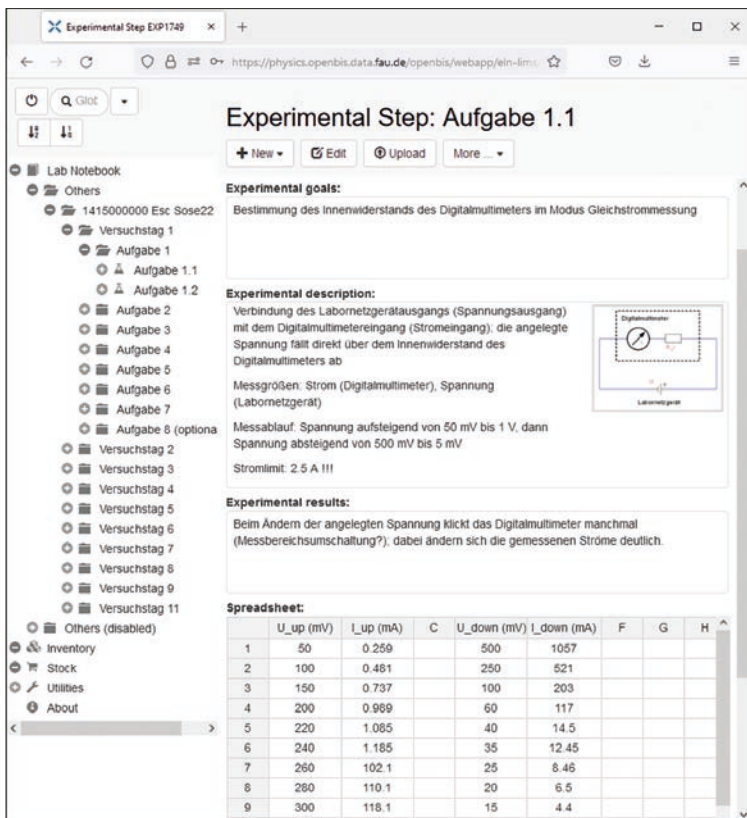


Abb. 2 In der Baumansicht im elektronischen Laborbuch ist der Praktikumsablauf vorgezeichnet. Die Studierenden haben während der Vorbereitung Ziele und Durchführung der einzelnen Praktikumsexperimente inklusive Skizzen in den entsprechenden Eingabefeldern notiert. Ergebnisse und Beobachtungen erfassen sie im gleichen Formular. Hierfür stehen ein Freitextfeld, eine Tabelle für numerische Messdaten sowie der Dateupload für Messergebnisse zur Verfügung.

neue Aufgabentypen, die komplexere, manchmal auch realitätsnähere oder aktuellere physikalische Probleme zulassen statt nur auf analytische Kompetenzen zurückzugreifen. Jede Lehrperson kann selbst entscheiden, an welcher Stelle eher analytische oder numerische Verfahren sinnvoll sind.

Bei der Visualisierung bietet die Numerik einen spielerischen Zugang: Mit ihrer Hilfe lassen sich Parameter schnell variieren und die Ergebnisse in verschiedenen grafischen Auftragsarten darstellen – in dieser Qualität und Geschwindigkeit ist dies mit traditionellen Methoden innerhalb einer Übung nicht zu schaffen. So können Studierende bei der Fourier-Analyse testen, was passiert, wenn sie im Signal nur gerade oder nur ungerade Terme zulassen. Idealerweise sollten Studierende in der Lage sein, solche Fragen analytisch und auch numerisch zu lösen beziehungsweise beide Methoden zu vergleichen.

Auch gilt es, dem Misstrauen aufmerksamer Studierender gegenüber Näherungen, die für analytische Lösungswege häufig unvermeidbar sind, zu begegnen. Numerisch quantifiziert und visualisiert ist oft viel leichter nachzuvollziehen, ob und unter welchen Umständen diese Näherungen gut begründbar sind. So profitieren Lehrveranstaltungen der experimentellen und der theoretischen Physik gleichermaßen von dieser Art spielerischer Datenkompetenz.

4) github.com/FAU-PHYSIK-EP/load_openBIS_data_into_python

Elektronisches Laborbuch

Für ein zukunftsweisendes Forschungsdatenmanagement eignet sich das handgeschriebene Laborbuch nicht. Daten und Metadaten sind nur durch sorgfältige elektronische Protokollierung nachhaltig verfügbar, etwa in Form eines elektronischen Laborbuchs (Electronic Lab Notebook, ELN). Dieses ermöglicht von Beginn an eine strukturierte Datenerfassung. In einem Pilotprojekt unserer Universität setzen wir das ELN openBIS ein (**Infokasten**), das aktuell in verschiedenen Forschungsgruppen ausprobiert und evaluiert wird.

Als Vertiefung des DV-Praktikums führen wir dieses ELN-System im Physikstudium ein. Dafür eignet sich das bestehende Elektronikpraktikum mit seinen elf Versuchstagen im 4. Fachsemester. Dieses bearbeitet ein Programm von einfachen elektrischen Messungen über komplexer werdende analoge und digitale Schaltungen bis hin zu Microcontrollerprogrammierung und Analog/Digital-Wandlung. Bislang haben die Studierenden die Messdaten auf einem Netzwerklaufwerk gespeichert und Vorbereitung sowie Metadaten in einem Papierlaborbuch dokumentiert. Der Übergang zum elektronischen Laborbuch im Frühjahr 2022 war technisch gesehen relativ einfach, da die Arbeitsplätze ohnehin mit einheitlichen Computern ausgestattet sind und die Lernziele sowie Inhalte des Elektronikpraktikums unberührt blieben.

Beim ersten Login sehen die Studierenden ihren Laborbuch-„Space“, auf den die Zweiertteams gemeinsamen Zugriff haben. Darin sind bereits alle Versuchstage als „Projects“ sowie die einzelnen Experimente und Aufgaben als „Experiments“ bzw. „Experimental Steps“ gegliedert. Die gesamte Struktur legen wir per Pythonskript an.

Die elektronischen Laborbuchseiten enthalten drei vorgegebene Freitextfelder für (i) die Ziele des Experiments, (ii) die Experimentbeschreibung und (iii) die Dokumentation der Resultate. Messdaten lassen sich bei kleineren Messreihen per Hand in Tabellen erfassen oder als Rohdatendatei in das ELN laden. Auf diese Weise führen die Studierenden während des Experimentierens die beschreibenden Metadaten sowie die Rohdaten an einem Ort zusammen (**Abb. 2**).

Electronic Lab Notebooks (ELN)

Unter den verfügbaren Open-Source-Softwarelösungen fiel die Entscheidung in Erlangen während der Pilotphase auf openBIS (openbis.ch), das viele wichtige Funktionalitäten mit sich bringt und es erlaubt, Datenstrukturen ähnlich einer Datenbank zu definieren und zu verknüpfen. Dabei gilt es bei der Nutzung, die Balance zwischen Effizienz (einfache Formulare, Freitexteingabe) und Perfektion (experimentenspezifische Formulare und resultierende strukturierte Metadatenablage) oder Schnittstellen zu Forschungspartnern zu finden. Mit solchen Fragen beschäftigen sich auch die verschiedenen NFDI-Konsortien, etwa FAIRmat.

Eine wichtige Funktionalität eines ELN-Systems ist die Exportschnittstelle. openBIS bietet sowohl die Möglichkeit, gut lesbare, formatierte Dokumente zu erstellen, als auch die Option, strukturierte Daten (JSON-Format) zu exportieren. Diese Interoperabilität ist die Voraussetzung für die Umsetzung der FAIR-Prinzipien.

Der Vorteil des ELN zeigt sich bei Nutzung der Python-schnittstelle: Mit wenigen Programmzeilen lassen sich die Daten direkt aus dem ELN zur weiteren Auswertung und Visualisierung laden (Abb. 3) – im Praktikum genauso wie zu Hause. Hierzu steht ein Pythonmodul⁴⁾ zur Verfügung, das Befehle der PyBIS-Bibliothek kombiniert.

Die Studierenden fanden schnell heraus, wie sie Fotos von ihrem Smartphone in das ELN einbinden können, um handgezeichnete Skizzen oder Fotos des Versuchsaufbaus zu integrieren. Gleiches gilt für Screenshots aus externen Dokumenten, wie Benutzerhandbüchern von Messgeräten. Obwohl die Nutzung des elektronischen Laborbuchs in der Pilotphase freiwillig war, haben die Studierenden es sofort akzeptiert und ausschließlich genutzt.

Darüber hinaus bietet das elektronische Laborbuch die Gelegenheit, Abläufe effizienter zu gestalten und etwa die Vorbereitung und Durchführung der Experimente besser zu verzahnen: So können die Studierenden ihre Antworten auf vorbereitende Fragen sowie weitere Notizen in die Freitextfelder (Ziele und Experimentbeschreibung) der entsprechenden Experimente eintragen. Vorbereitung, Fragenkatalog und Antworten sind am Ende übersichtlich dargestellt und wohlgeordnet. Zudem können sich die Betreuerinnen und Betreuer durch ein Pythonskript eine exzerpierte Version der Vorbereitung ansehen und sich damit einen Überblick über die Vorbereitung der verschiedenen Gruppen verschaffen, ohne sich durch die einzelnen Experimente im ELN klicken zu müssen.

Ein elektronisches Laborbuch hilft auch, Nachbesserungen, Plagiate oder Fälschungen zu verhindern. Denn in einem elektronischen System ist jede Handlung prinzipiell genau nachvollziehbar. Somit kommt erst gar nicht der Gedanke auf, Daten zu „frisieren“. Der Rohdatenzugriff ermöglicht es den Betreuern, die Versuchsaufbauten im Verlauf der Zeit bei etwaigen Ergebnisabweichungen zu überprüfen und technische Defekte zu entdecken. Damit die Studierenden nach Abschluss der Lehrveranstaltung eine vollständige Dokumentation haben, können sie die Inhalte des elektronischen Laborbuchs exportieren.

Wirkung auf die Forschungsebene

Nachdem die Studierenden den Umgang mit dem ELN im 4. Semester erlernt haben, bringen sie ihre Kompetenzen während der Forschungsphase zur Bachelor- oder Masterarbeit unmittelbar in die Arbeitsgruppen ein. Dort gilt es allerdings, die Forschungsdaten wesentlich strukturierter und an die jeweiligen Arbeitsabläufe angepasst zu erfassen. Diese Aufgabe ist mehrheitlich noch zu lösen und Gegenstand der NFDI-Initiative. Auch wir sind dabei, unsere Arbeitsabläufe in der Forschung systematisch zu erfassen und abzubilden. Dabei helfen Konzepte aus der Informatik, wie sie für Datenbanken existieren, etwa Entity-Relationship-Diagramme. Auch hier nützt der ELN-Einsatz im Studium: Wir werden Konzepte mit stärker strukturierten elektronischen Laborbüchern zunächst im Praktikum erproben und evaluieren. Aufgrund der abgeschlossenen Einheiten sind dort kurze Lern- und Innovationszyklen möglich, deren Erfahrungen wir in die längerfristig angelegte und komplexere Dokumen-

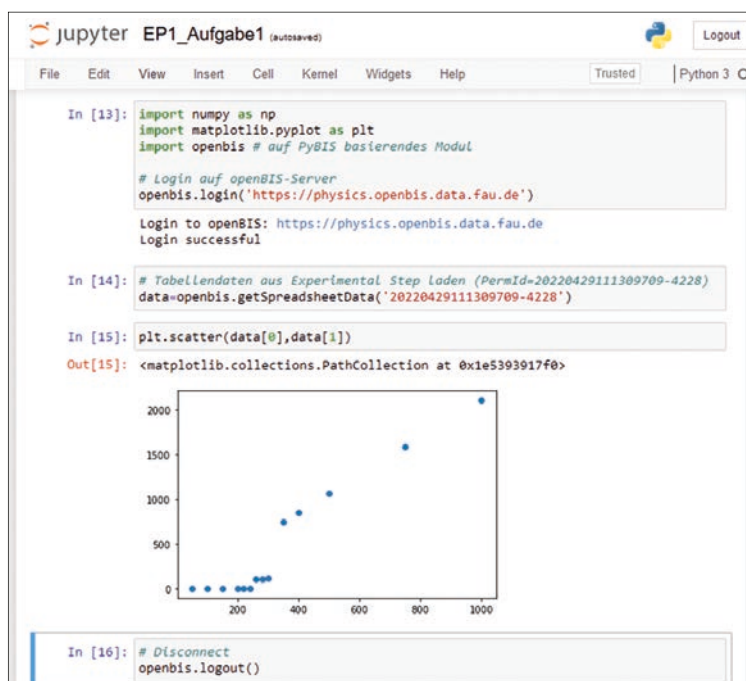


Abb. 3 Mit wenigen Programmzeilen laden die Studierenden ihre im Praktikum erfassten Messdaten direkt aus dem ELN in ein Jupyter-Notebook, um sie dort zu analysieren oder grafisch zu visualisieren.

tation der Forschung in unserer Arbeitsgruppe einbringen können.

Wir sehen es als ein Erfolgsrezept, frühzeitig im Physikstudium zeitgemäße Datenkompetenz zu vermitteln. Die durchgängige Verwendung einer einheitlichen und in den Naturwissenschaften mächtigen Programmiersprache wie Python verschafft neue Möglichkeiten und verändert die Lehre und die Forschung nachhaltig. Zudem sehen wir einen sehr positiven Einfluss auf die sich verändernde Forschungslandschaft, die viel stärker als früher von einem aktiven Forschungsdatenmanagement geprägt sein wird.

Die Autoren



Heiko B. Weber (FV Tiefe Temperaturen) und **Michael Krieger** (FV Halbleiterphysik) forschen an festkörperphysikalischen Fragestellungen, leiten das Elektronikpraktikum und arbeiten aktiv im NFDI-Konsortium FAIRmat mit.

Christopher van Eldik (FV Teilchenphysik) forscht an astroteilchenphysikalischen Fragestellungen. Als Studiendekan hat er die Umstellung des Curriculums mitgeprägt und leitet die Lehrveranstaltung „Datenverarbeitung in der Physik“.

Dr. Michael Krieger, Prof. Dr. Heiko B. Weber, Lehrstuhl für Angewandte Physik, Department Physik, Friedrich-Alexander-Universität Erlangen-Nürnberg; **Prof. Dr. Christopher van Eldik**, Erlangen Centre for Astroparticle Physics, Department Physik, Friedrich-Alexander-Universität Erlangen-Nürnberg