

Fit für „Big Data“: stufenweise interaktives wissenschaftliches Programmieren mit Jupyter Notebooks für Geowissenschaftler

Antragsteller: J. Florian Wellmann

Inhalt

Motivation – Warum...?	0
Erfahrungen des Antragsstellers	2
Wie sollen die interaktiven Programmierübungen in den Lehrbetrieb integriert werden?	3
Erwartete Ergebnisse (einschließlich Innovationspotential)	6
Wie kann der Erfolg des Vorhabens gemessen werden? Was gibt es für Risiken? ...	6
Welche Zielgruppen und Veranstaltungen werden adressiert?	7
Wie soll das Vorhaben verstetigt werden?	7
Übertragbarkeit des Innovationsvorhabens	8
Referenzen und Kommentare	9

Motivation – Warum...?

Die Verfügbarkeit großer Datenmengen in den Geowissenschaften nimmt stetig zu – und damit auch die Notwendigkeit, diese Daten bearbeiten zu können [1,2]. Nun muss man allerdings sagen, dass gerade im Bereich der Geowissenschaften die Affinität zu wissenschaftlicher Programmierung und quantitativer Datenanalyse unter den Studierenden nicht gerade stark ausgeprägt ist. Daher experimentiere ich in meiner Lehre bereits seit einigen Jahren mit neuen Programmierkonzepten, die den Einstieg in die wissenschaftliche Programmierung deutlich erleichtern – basierend auf dem Konzept der „Jupyter Notebooks“, die in der Zwischenzeit für viele Programmiersprachen (u.a. Python, R, Julia, Matlab) vorhanden sind. Der bisherige Erfolg dieser Konzepte zeigt sich auch darin, dass die entsprechenden Lehrveranstaltungen durchgehend sehr positiv bewertet wurden (mit einer Gesamtdurchschnittsnote von 1,2) und ganz besonders freue ich mich, in diesem Jahr dafür sogar den „Lehrpreis der Fachschaft für die Beste Innovation in der Lehre“ erhalten zu haben.

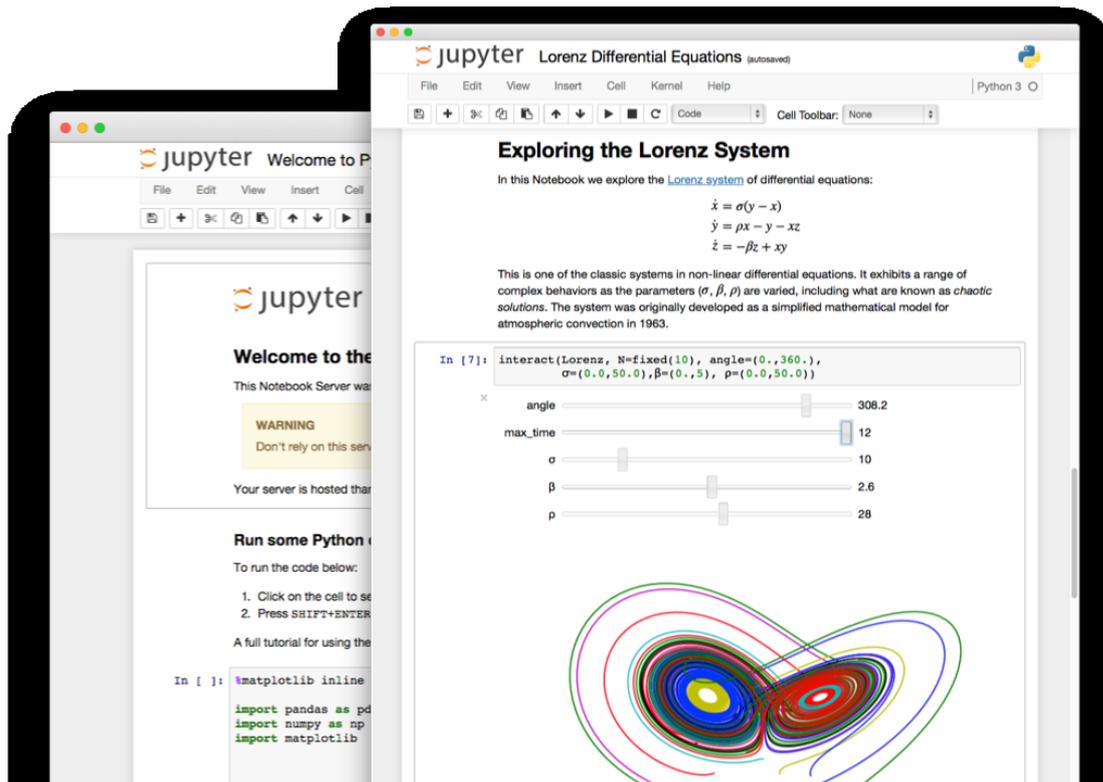


Abbildung 1 Beispiel eines Notebooks zur Darstellung von Text, Code und Abbildungen in einem Dokument (quelle: jupyter.org)

Die Attraktivität dieser Notebooks liegt darin, dass auf elegante Weise Instruktionen, wissenschaftlicher Code und (graphische) Ergebnisse direkt in einem Dokument verbunden werden können (Siehe Abb. 1). Diese Methoden erleichtern den Studierenden bereits jetzt den Einstieg in wissenschaftliche Programmierung und Datenanalyse. Abgesehen von der Lehre werden Jupyter Notebooks auch zunehmend in der wissenschaftlichen Kommunikation verwendet. Auch wächst die Anzahl an Lehrbüchern und Online-Kursen (z.B. in den Geowissenschaften: <http://earthpy.org/>), die in Jupyter Notebooks verfasst werden, stetig. All diese Entwicklungen zeigen, dass das grundlegende Konzept der Notebooks ideal für die Verwendung in der digitalen Lehre geeignet ist.

Mit diesem Antrag möchte ich auf diesem erfolgreichen Konzept und meinen bisherigen, positiven Erfahrungen aufbauen. Der Grund ist vor allem, dass sich die Methoden rund um die „Jupyter Notebooks“ momentan rasant weiterentwickeln, so dass sich ständig neue Möglichkeiten ergeben, diese auch in der Lehre zu verwenden. So ist es jetzt zum Beispiel möglich, eine Vielzahl unterschiedlicher interaktiver Steuerungselemente einzubinden, die z.B. eine Parameterstudie sehr intuitiv und einfach machen. Weiterhin gibt es jetzt Möglichkeiten, Code teilweise auszublenden, um so den Fokus auf den Inhalt zu lenken. Eine ideale Grundlage, um Studierenden der Zugang zu Programmierung und Datenanalyse schrittweise nahezubringen. Neuerungen und Erweiterungen für Jupyter Notebooks erscheinen mittlerweile fast im Wochenrhythmus. Das Fellowship würde mir daher die perfekte und zeitlich ideale Möglichkeit geben, diese neuen Methoden für die Lehre zu testen

und weiter zu entwickeln – und deren Verwendung auch über meine Lehre und die Geowissenschaften hinaus noch weiter bekannt zu machen.

Mit diesem Antrag möchte ich ganz konkret zwei Hauptziele verwirklichen: (a) die Entwicklung von Konzepten und deren technische Realisierung, um Programmierübungen entsprechend der individuellen Kompetenzniveaus der Studierenden erstellen zu können und so insbesondere den Einstieg ins Programmieren zu erleichtern, sowie (b) eine Cloudumgebung für die Studierenden zu implementieren, welche als Lernplattform zum Programmieren dient, komplexe Aufgaben ohne komplizierte Installationsvorgänge zulässt und kooperatives Arbeiten an Projekten ermöglicht. Weiterhin erwarte ich viele neue Entwicklungen um Jupyter Notebooks und werde im Rahmen dieses Projektes schnell und flexibel auf diese Entwicklungen reagieren, um diese für die Verwendung in der Lehre anzupassen und optimieren zu können.

Der bedeutende Innovationsaspekt ist, dass neueste Methoden um das erfolgreiche Konzept der „Jupyter Notebooks“ speziell für die Lehre angepasst und optimiert werden, in einer Umgebung, die Programmierung auf intuitive Weise und ohne technische Hürden ermöglicht – mit dem Ziel, dass Studierende Programmierung spielerisch erfahren und erlernen und dabei einen positiven Einstieg in den Sinn und das Potential der wissenschaftlicher Programmierung und Datenanalyse in den Geowissenschaften bekommen.

Erfahrungen des Antragsstellers

Die technischen und methodischen Grundlagen für diesen Antrag – basierend auf den „Jupyter Notebooks“ (<http://jupyter.org>) wurden bereits erfolgreich in meinen Lehrveranstaltungen eingesetzt und kontinuierlich weiterentwickelt, auch unterstützt durch einen erfolgreichen „Exploratory Teaching Space“-Antrag, den ich RWTH-intern dafür bekommen habe. Die Zustimmung von Seiten der Studierenden war durchgehend positiv (Durchschnittsnote 1,2; Lehrpreis der Fachschaft).

Ergebnisse des Projektes haben wir auch schon auf einer wissenschaftlichen Konferenz vorgestellt, auf dem diesjährigen European Geosciences Union General Assembly (<http://meetingorganizer.copernicus.org/EGU2017/EGU2017-17191.pdf>). Auch dort war das Interesse groß und die Rückmeldung durchgehend sehr positiv, was zeigt, dass die Methoden auf Interesse bei anderen Lehrenden stoßen – und auch, dass sie bisher gerade im Bereich der Geowissenschaften noch nicht weit verbreitet sind.

Abgesehen von der Lehre verwenden wir die „Jupyter Notebooks“ in meiner Arbeitsgruppe auch intensiv in der Forschung, z.B. zur Dokumentation wissenschaftlicher Experimente, zur reproduzierbaren Analyse wissenschaftlicher Daten, sowie in der Form von Tutorials zu eigenen wissenschaftlichen Codes (hier ein Beispiel in statischer Form: <http://pynoddy.readthedocs.io/en/latest/notebooks/5->

[Geophysical-Potential-Fields.html](#)). Ich baue daher mit den geplanten Methoden auf eine breit angelegte Erfahrung der Konzepte und auf Kenntnisse der neuesten Entwicklungen in diesem Bereich auf.

Wie sollen die interaktiven Programmierübungen in den Lehrbetrieb integriert werden?

Teilziel (a): Mit diesem Teilziel wird die Einbindung von Jupyter Notebooks als „blended learning“ (engl: blend = mixen) Komponente in den bestehenden Lehrbetrieb entwickelt. Frontale Programmierübungen sollen in eine zentralisierte Arbeits- und Präsenzphase umgestellt werden, in der die Lehrperson nicht mehr frontal, sondern zentral positioniert ist, und die einzelnen Studierendengruppen unterstützt, bzw. die Gruppen sich gegenseitig unterstützen. Die Notebooks sollen also als digitaler Mix in den bestehenden Lehrbetrieb eingehen.

Dabei soll der Inhalt der Notebooks stufenweise vorlesungsbezogenen Inhalt als Programmierübung beinhalten. Studierende sollen individuell auswählen können, welchen Schwierigkeitsgrad die Übungen haben (Abb. 2). Dadurch können sie ein und dasselbe Notebook mehrmals mit steigendem Schwierigkeitsgrad verwenden. Da Notebooks auf Zellen mit eigenem Typ basieren, lässt sich dieses Vorhaben mittels Filtern realisieren. Anfängliche Erfolge im Durcharbeiten der Notebooks sollen die Studierenden ermuntern, wachsende Schwierigkeitsgrade zu bearbeiten. Im Semesterverlauf werden die Inhalte der Notebooks entsprechend der Lehrveranstaltung komplexer. Auch, um so der steigenden Kompetenz der Studierenden gerecht zu werden und eine Unterforderung zu verhindern.

Bei der Umsetzung dieses Teilziels steht der Themenbezug der Programmierübungen im Vordergrund. Es soll für die Studierenden von vorn herein ersichtlich sein, wofür die Übungen, und das daraus gewonnene Wissen nützlich sind. Die wissenschaftliche Programmierung und quantitative Datenanalyse wird so gewissermaßen nebenbei erlernt und es soll den Studierenden so auch vermittelt werden, dass diese Fähigkeiten Möglichkeiten und Freiheiten geben, die über die einfache Verwendung von festgelegten Analysemethoden herausgehen.

Teilziel (b): Der erste Schritt um Programmieren zu lernen ist immer die Installation der notwendigen Softwarepakete der jeweiligen Programmiersprache. Heutzutage geht dies bei vielen Sprachen und Systemen zwar einfach, jedoch ist jede individuelle Installation eine zusätzliche Hemmschwelle und birgt das Risiko von Problemen, z.B. in der Form von Versionsunterschieden.

Um dieses Risiko zu vermindern und auch den Studierenden den Schritt der Installation und Einrichtung zu erleichtern, planen wir die Implementierung einer cloudbasierten, multi-user Plattform, auf der die Studierenden simultan programmieren können.

Konzept der stufenweisen interaktiven Programmierung

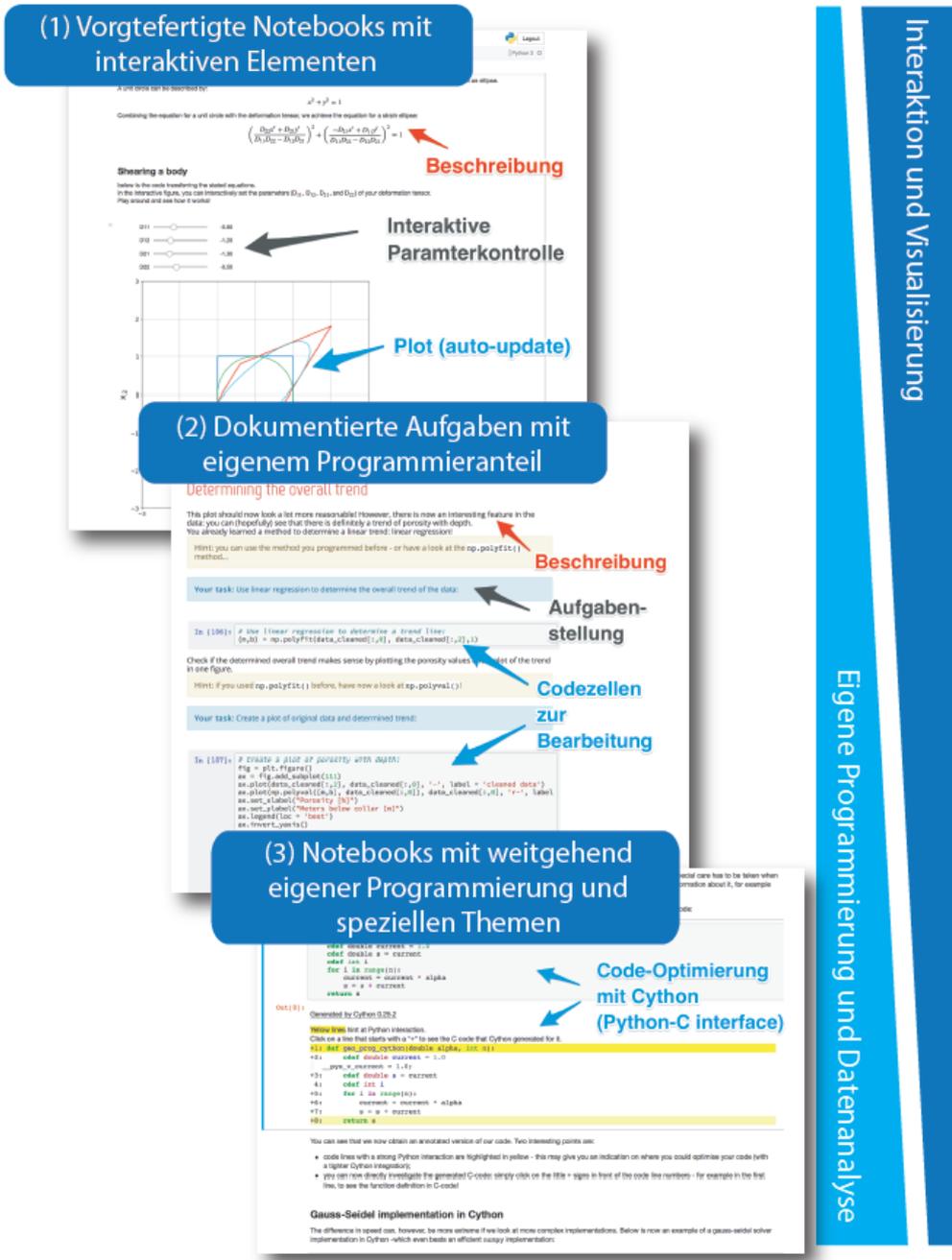


Abbildung 2: Schematisches Konzept zum Einsatz von Jupyter Notebooks mit unterschiedlichen Abstraktionsebenen. Zum vereinfachten Einstieg bestehen vorgefertigte Notebooks mit interaktiven Elementen. Programmierkenntnisse der Studierenden sind dort nicht verlangt. Diese werden nach und nach gestuft gefordert (siehe (2) und (3)).

Mit dieser Plattform wird eine zentrale Kontrolle über den Lehr- und Prüfungsstoff in Form von Notebooks gegeben. Studierende werden in diesem Umfeld Möglichkeiten haben, die eigenen Entwicklungen zu testen (im Rahmen von Self-Assessment

Optionen). Den Lehrenden soll diese Cloudumgebung weiterhin die Möglichkeit geben, zentral Prüfungsleistungen (in Form von Programmierprojekten mit automatischer Bewertung) freizugeben, zu benoten, und (datenschutzrechtskonform) den einzelnen Studierenden Feedback zu geben. Notebooks können zur Bearbeitung freigeschaltet, zentral modifiziert und gesperrt werden. Weiterhin gibt diese Umgebung den Lehrenden flexible Möglichkeiten, den einzelnen Studierenden individuell Feedback auf Ihre Programmierung zu geben. Jeder Studierende hat auf der Plattform nur Zugriff auf den persönlichen Arbeitsordner und einen Austauschbereich. Nicht zu benotende Aufgaben können von den Studierenden kooperativ im Austauschbereich bearbeitet werden. Angedacht ist weiterhin, eine Erweiterung für Notebooks zu implementieren, sodass mehrere Studierende zeitgleich an ein und demselben Notebook arbeiten können, und Änderungen jeweils durch User-IDs gekennzeichnet sind. Insgesamt ergeben sich damit vielfältige neue und innovative Möglichkeiten, um wissenschaftliche Programmierung in der Lehre attraktiver zu gestalten und flexibler einzubinden.

Eine Entwicklung auf technischer Ebene ist die Einbindung der Notebooks in die persönlichen Arbeitsordner der Studierenden, individualisiert über die RWTH-Lernplattform L²P. L²P wurde in der Vergangenheit schon erfolgreich als Authentifizierungsdienst genutzt. Diese Methode ist daher in erster Linie speziell optimiert für den Einsatz in der Lehre an der RWTH, aber da Lernplattformen auch darüber hinaus weit verbreitet sind, ist das keine Limitierung der Übertragbarkeit (s.u.).

Weitere Ziele: Wie bereits erwähnt, werden Jupyter Notebooks zur Zeit in vielen Bereichen weiterentwickelt – und so ergeben sich ständig neue und spannende Möglichkeiten, die neu entwickelten Konzepte in die Lehre einzubinden. Im Rahmen des Projektes werde ich die Möglichkeit haben auf dem neuesten Stand dieser Entwicklungen zu bleiben und diese auch sofort für die Lehre anzupassen und gegebenenfalls im Lehrbetrieb zu testen. Bereits absehbar (auf Grund laufender Projekte) sind beispielsweise folgende Möglichkeiten:

- Die interaktive Bearbeitung von Diagrammen und Plots: dies ist bereits jetzt eingeschränkt möglich, allerdings noch fehlerbehaftet und momentan in aktiver Entwicklung;
- Eine bessere Möglichkeit, um Daten in 3-D anzuzeigen: dieser Aspekt ist natürlich besonders im Bereich der Geowissenschaften sehr spannend und relevant;
- Die Analyse und Darstellung geographischer Daten: bereits jetzt gibt es dafür Möglichkeiten, aber gerade eine bessere Kombination mit Open-Source GIS-Projekten verspricht spannende weitere Entwicklungen in diesem Bereich.

Diese Ziele sind daher kaum planbar, aber absehbar – und gerade diese Fellowship würde mir die Möglichkeit geben, auf diese Entwicklungen flexibel und schnell zu reagieren.

Erwartete Ergebnisse (einschließlich Innovationspotential)

In erster Linie versprechen wir uns Studierende aus den Geowissenschaften eine entscheidende Kompetenz zu vermitteln: die Fähigkeit zu wissenschaftlicher Programmierung und Datenanalyse - als Vorbereitung auf einen Arbeitsmarkt, in dem diese Kompetenzen als „key skill“ immer mehr gefragt sind.

Ein hohes Innovationspotential liegt dabei ganz besonders in der Erarbeitung der beiden Teilziele, wie oben definiert. Sowohl die interaktive Bearbeitung unterschiedlicher Leistungsniveaus inklusive einer Selbstkontrolle, als auch die Einbindung in eine Cloud-Umgebung, die speziell für den Lehrbetrieb angepasst wird, enthalten an sich schon Komponenten mit hohem Innovationspotential – sowohl auf der konzeptionellen, als auch auf der rein technischen Ebene. Da sich aus den aktuellen Entwicklungen rund um die Jupyter Notebooks momentan vielfältige Möglichkeiten ergeben, um spannende neue Lehrkonzepte zu entwickeln, erwarte ich viele weitere Entwicklungen, die auch Studierenden ohne besonderen Bezug zu Programmierung und quantitativer Datenanalyse einen einfachen und intuitiven Einstieg in diese Materie ermöglichen.

Wie kann der Erfolg des Vorhabens gemessen werden? Was gibt es für Risiken?

Als Maßstab für den Erfolg der Methoden werde ich weiterhin sehr genau auf die Rückmeldung der Studierenden, sowohl im Rahmen der standardisierten Evaluationsmethoden an der RWTH als auch im persönlichen Gespräch, achten. Auch die Rückmeldung von Kolleginnen und Kollegen, die diese Methoden auf unsere Entwicklungen hin jetzt in der Lehre verwenden, ist entscheidend, um die Einsetzbarkeit auch in Fächern zu bestimmen, die keine „klassischen“ Programmierübungen enthalten (wie bereits jetzt z.B. in der Strukturgeologie, in der auf unsere Initiative hin diese Konzepte jetzt verwendet werden).

Risiken, also potenzielles Scheitern der Teilziele, oder des Vorhabens sind vorhanden, aber nicht als groß einzuschätzen. Es existiert das technische Risiko, dass die Cloudplattform einem simultanen Ansturm vieler Studierender nicht gewachsen ist. Dieses Risiko wollen wir minimieren, indem wir stetig die Zugriffe auf die Plattform via L²P beobachten (Nutzungsstatistiken lassen sich über L²P bereits jetzt auslesen). Da die Cloudplattform als VM modular aufgebaut ist, können proaktiv weitere Prozessoren, oder weiterer Speicher zugeschaltet werden.

Ein Risiko von Teilziel (b) besteht in mangelnder Teilnahme der Studierenden, wie prinzipiell in jeder Lehrveranstaltung. Basierend auf den bisherigen Erfahrungen (alle Kurse, die diese Methoden verwenden, waren bisher komplett ausgelastet und überausgebucht) und der positiven Bewertungen ist das allerdings nicht zu erwarten.

Weiterhin wird versucht, Anreize in Form von Themenbezug, und digitaler „Barrierefreiheit“ (Verringerung der Hemmschwelle) zu schaffen. Es zeigt sich in der Zwischenzeit auch deutlich, dass vielen Studierenden bewusst ist, dass wissenschaftliches Programmieren und Datenanalyse auch in den Geowissenschaften immer mehr zu einer bedeutenden Fachkompetenz werden – und genau diese Kompetenz wird hier aufgebaut.

Welche Zielgruppen und Veranstaltungen werden adressiert?

Die Zielgruppe sind explizit Studierende, die mit wissenschaftlicher Programmierung und Datenanalyse bisher wenig Kontakt hatten. Im Bereich der Geowissenschaften ist die Anwendung daher in einigen Studiengängen gut möglich. Die geplante Lehrinnovation soll daher in den Studiengängen „Georessourcenmanagement“ (Bachelor), „Angewandte Geowissenschaften“ (Master) und „Applied Geophysics“ (Master) an der RWTH Aachen exemplarisch stattfinden. Wie oben beschrieben, wurden die grundlegenden Konzepte bereits in weiteren Lehrveranstaltungen angewendet (z.B. „Geothermics“, „Statistik und Programmieren/Modellieren (SoSe)“, oder „Endogene Dynamik I (WiSe)“) und auch unsere geplanten Weiterentwicklungen stoßen bereits auf großes Interesse (z.B. „Grundlagen der angewandten Geophysik I und II (WiSe und SoSe)“).

Wie soll das Vorhaben verstetigt werden?

Aus der eigenen Erfahrung heraus ist mir der anfängliche Mehraufwand deutlich bewusst – aber auch die gute Möglichkeit, die entwickelten Methoden später ohne großen Aufwand weiterzuführen. Ein großer Teil des Vorhabens besteht darin, neue Methoden rund um Jupyter Notebooks zu testen und für den Einsatz in der Lehre weiterzuentwickeln. Sobald diese Methoden zur Verfügung stehen sollte sich eine weitere Anwendung in der Zukunft im Rahmen der üblichen Anpassungen von Lehrveranstaltungen bewegen, ohne große zusätzliche Arbeit. Natürlich kann erwartet werden, dass es auch in der Zukunft (und nach dem Ablauf der Förderungsphase) spannende Weiterentwicklungen um diese Konzepte geben wird, aber mit der Grundlage, die hier gesetzt wird, erwarte ich auch in Zukunft weitere Änderungen einfach einsetzen zu können. Wenn der erste Mehraufwand durch dieses Fellowship ermöglicht wird, dann sehe ich daher keine großen Schwierigkeiten bei der Verstetigung.

Der initiale Aufwand für den technischen Teil des Vorhabens, eine cloudbasierte Programmierplattform für die Lehrveranstaltung mit Kopplung an L2P einzurichten ist sicherlich nicht gering. Sobald das System jedoch erfolgreich genutzt wird, ist die Verstetigung nur eine Frage der Weiterfinanzierung der Plattform. Es kann auch erwartet werden, dass durch diese technische Weiterentwicklung der Umstieg für andere Lehrende auf diese Methoden deutlich erleichtert wird.

Übertragbarkeit des Innovationsvorhabens

Aus den vorangegangenen Projekten besteht bereits eine positive Erfahrung zur Übertragbarkeit der Methoden. Unter Anderem werden die Konzepte (gerade auch mit den spezifischen Anpassungen für die Lehre) jetzt in Übungen zur Geothermie, Statistik und sogar in der Strukturgeologie verwendet. Testweise werden wir in diesem Jahr sogar Übungen im Kurs „Karten und Profile“ auf die Notebooks umstellen. Aufgrund dieser durchgehend positiven Erfahrungen kann erwartet werden, dass auch neue Methoden gut auf unterschiedliche Lehrveranstaltungen angewendet werden können.

Die Übertragbarkeit der technischen Aspekte des Vorhabens ist direkt gegeben. Notebook-Erweiterungen, sowie die erstellte Struktur von Programmierplattform und L²P lassen sich problemlos und direkt auf andere Disziplinen übertragen. Bei erfolgreicher Implementierung des JupyterHub auf einem virtuellen Server des ICT-Centers ist eine Übertragbarkeit auf andere Veranstaltungen an der RWTH nahezu problemlos möglich. Für Institute mit eigenen Serverkapazitäten sollten die getesteten Konzepte ebenfalls direkt anwendbar sein. Die Verwendung von L²P, das als fest integriertes Hilfssystem allen Veranstaltungen zur Verfügung steht, vereinfacht weiterhin eine direkte und sichere Einbindung in Lehrveranstaltungen.

Die inhaltliche Übertragbarkeit des Innovationsvorhabens kann über mehrere Kanäle realisiert werden. Als direkter Kontakt dient dazu das Forum des Programms, um sich mit interessierten Fellows auszutauschen. Angedachte Fellowtreffen können zur vertieften Kommunikation des Vorhabens und zur Skizze möglicher inhaltlicher Übertragungen genutzt werden. Ich erhoffe mir dadurch, auch über die Hochschule hinaus mit den erarbeiteten Konzepten einen wertvollen Beitrag für die Lehrinnovation leisten zu können, und interessierte Fellows bei der Implementierung von Jupyter Notebooks in ihre Lehre zu unterstützen. Des Weiteren ist ein „Training Day“, vorerst an der RWTH, geplant, um die Methoden einem breiten Feld Lehrender aus allen Fachrichtungen vertraut zu machen. Potenziellen Teilnehmern des „Training Day“ soll ein intensiver Einstieg in Jupyter Notebooks gegeben werden, der auch die Implementierung auf Server für mehrere Nutzer umfasst. Während des Training Days sollen mit den Teilnehmern zusammen erste Konzepte für integrative Notebooks ihrer Lehrveranstaltungen erstellt werden. Während eines solchen Training Days, können nicht nur die technischen Aspekte des Vorhabens erläutert, sondern auch in direkter Zusammenarbeit mit den Teilnehmern geeignete Themen für interaktive Programmiernotebooks identifiziert werden. Diese geeigneten Themen sollten direkten Bezug zur jeweiligen Lehrveranstaltung haben, und über das Potenzial verfügen, sinnvoll in unterschiedlichen „Schwierigkeitsstufen“ implementiert zu werden. Wenn dieses Vorhaben erfolgreich ist, dann ist durchaus vorstellbar, den Teilnehmerkreis für solch einen „Training Day“ auch auf andere Universitäten auszuweiten.

Referenzen und Kommentare

[1] Als Beispiel kann hier allen voran die Turing Conference for Big Data in Geoscience angeführt werden <https://www.turing.ac.uk/events/big-data-in-geoscience/>

[2] Lin, J.W.B., 2012. Why Python is the next wave in earth sciences computing. *Bulletin of the American Meteorological Society*, 93(12), pp.1823-1824.