

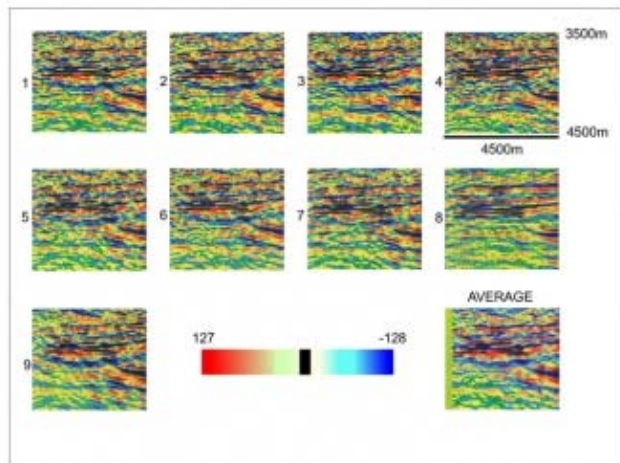
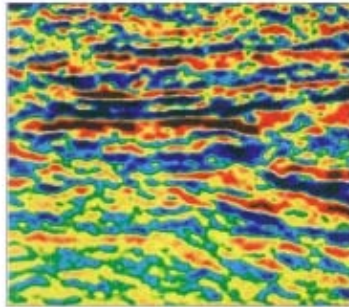
Erhaltung und Publikation von wissenschaftlicher Software

Helmholtz Open Access Workshop 2013

Dr. Jens Klump

Deutsches GeoForschungsZentrum GFZ

Die wissenschaftliche Herausforderung



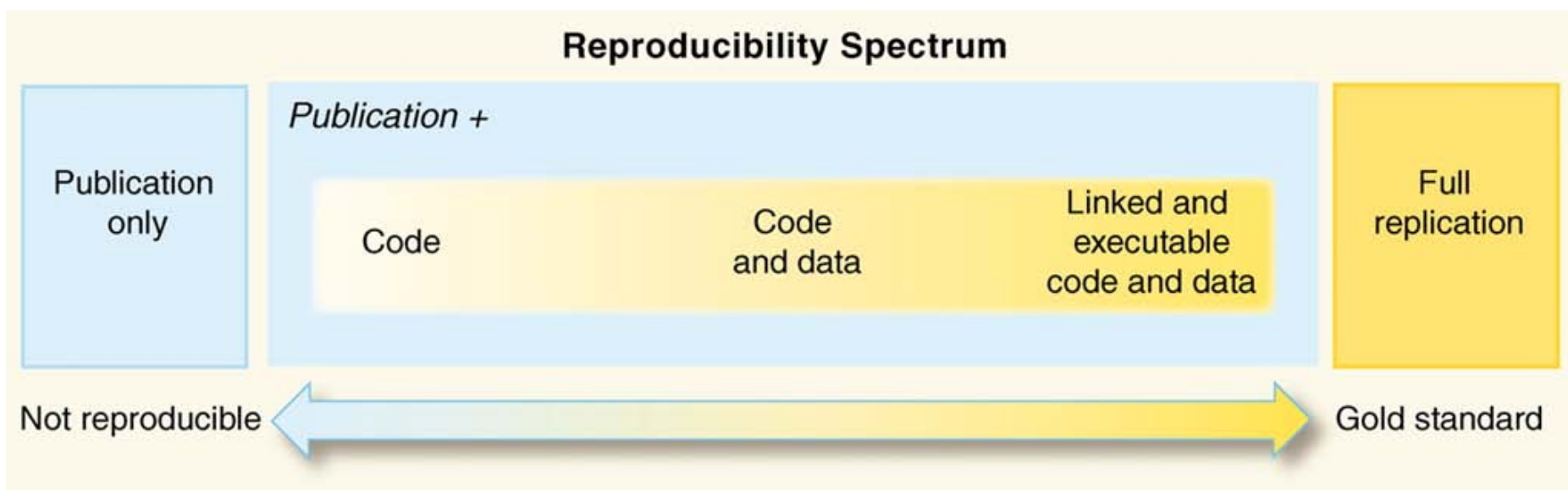
Hatton & Giordani, 2012

Beispiel: Daten aus einem seismischen Profil wurden durch neun verschiedene Kombinationen von Hard- und Software prozessiert.

Ergebnis: neun verschiedene, **signifikant unterschiedliche Ergebnisse**.

Dieses Ergebnis steht in Konflikt mit der wissenschaftliche Methode.

Ist es reproduzierbar?



Peng, *Science*, 2011

Vorschlag zur guten wissenschaftlichen Praxis

POLICY ACTIONS TO ELIMINATE SOURCE CODE WITHHOLDING IN RESEARCH COMPUTATION

Institutional support

Publicly funded research institutions and university TTOs must remove organizational impediments to OSS licensing of computer code and embrace a wider variety of methods for exploiting and sharing their intellectual property. Creating a “standard set” of open software licensing tools within and across institutions that includes established OSS licenses would be an important step toward that goal.

Funding policy

Public funding and policy-setting agencies must explicitly and clearly state their strong preference for open dissemination, sharing, and publication of scientist-created software and source code. Although not an absolute requirement in recognition of the enormous diversity of research receiving public funds, the burden of justifying proprietary research products would be left to the applicant.

Publishing requirement

Scientific journal publishers must enact editorial policies requiring, as a condition of publication, that researchers make available new computer source code generated in the course of the research and necessary to reproduce the published research findings. Policies in place at journals already meeting this requirement (16–18, 36) could provide guidance for wider implementation.

Morin, et al., *Science*, 2012

Offene Fragen

Im Rahmen der Forschung entstehen eine Vielzahl von Algorithmen, Skripte und Software-Anwendungen ... kurz: Code.

- Welche Wege der **Veröffentlichung** gibt es?
- Wie kann ich eine spezifische Version **referenzieren** und **zitieren**?
- Welche **Lizenzen** sind für wissenschaftlichen Software-Code geeignet?
- Welche Plattformen und Kanäle gibt es für den **Vertrieb** für Quelltext und Dokumentation von Software und Modellen?

Die soziale Herausforderung

Ähnlich wie beim Zugang zu Forschungsdaten gibt es auch bei der Offenlegung von Quellcode Vorbehalte:

- Format und Dokumentation des Code
- Tests und Fehlersuche nicht abgeschlossen
- Lizenzierung und Pläne für kommerzielle Nutzung
- Technisch schwierige Implementierung
- Fehlende Anerkennung in der Bewertung wissenschaftlicher Leistungen

Eine Welt ohne Beweise

Stellen wir uns einmal eine Welt vor, in der Veröffentlichungen zur Mathematik keine Beweise enthielten ...

- Niemand erwartet einen Beweis und niemand hat bisher einen Beweis veröffentlicht.
- *Das war schon immer so und die Literatur ist voll von guten Theoremen. Warum sollten wir das ändern?*

Stellen wir uns vor, jemand schlägt vor, Veröffentlichungen sollten Beweise enthalten ...

Vorbehalte

- Der Beweis ist nicht schön.
 - Es wäre zu aufwendig, den Beweis zu überarbeiten, so dass andere ihn nachvollziehen können.
 - Der Beweis zielt nur auf dieses eine Theorem, macht aber keine allgemeinen Aussagen. Er war nur für den eigenen Gebrauch gedacht.
 - Meine Zeit verwende ich lieber darauf, noch weitere Veröffentlichungen zu schreiben, statt noch mehr Zeit auf diese Veröffentlichung zu verwenden. Schließlich habe ich ja schon einen Beweis.

LeVeque, 2011

Noch mehr Vorbehalte

- Der Beweis ist im Detail noch nicht völlig ausgearbeitet.
 - Der Beweis ist noch nicht für alle Spezialfälle ausgearbeitet. Er funktioniert jedoch für die Beispiele in meiner Veröffentlichung.
 - In ein paar Fällen funktioniert der Beweis nicht. So lange ich das nicht erwähne, wird es keiner merken.
 - Eigentlich hat mein Doktorand das Theorem bewiesen.
- Der Doktorand ist inzwischen nicht mehr hier, aber ich bin sicher, der Beweis funktioniert.

LeVeque, 2011

... und noch mehr Vorbehalte

- Der Beweis ist wertvolles geistiges Eigentum.
 - Es hat mich Jahre gekostet, dieses Theorem zu beweisen. Wieso sollte ich den Beweis jetzt kostenlos hergeben?
 - Mit dem selben Ansatz lassen sich weitere Theoreme beweisen. Ich sollte daraus zumindest noch fünf weitere Veröffentlichungen machen dürfen.
 - Jemand könnte den Ansatz in meinem Beweis für weitere Beweise nutzen, ohne mich zu zitieren.
 - Der Ansatz ist großartig. Ich könnte ihn kommerziell nutzen oder verkaufen.

LeVeque, 2011

... und noch mehr Vorbehalte

- Es gibt technische Hürden.
 - Den Beweis in die Veröffentlichung mit aufzunehmen würde sie sehr viel Länger machen. Das druckt kein Verlag.
 - Gutachter wollen keine Beweise auf Gültigkeit prüfen. Dafür findet man keine Gutachter.
 - Der Beweis verlangt raffinierte mathematische Werkzeuge, die den meisten Lesern und Gutachtern nicht zur Verfügung stehen.
 - Mein Beweis bezieht sich auf nicht veröffentlichte oder proprietäre Beweise. Meine Beweis zu veröffentlichen hilft auch nicht weiter, weil die Leser ihn dadurch nicht vollständig nachvollziehen können.

LeVeque, 2011

Spektrum der Anwendungsfälle

- Algorithmen (Matlab, R, ...)
- Eigenständige Software
- Modelle
- Beiträge zu Open Source Software
- System-Images für den Einsatz in der Cloud
- ...

Keine Lösung kann alle Anwendungsfälle abdecken.

Offene Fragen

- Welche Vorgehensweisen eignen sich für welche Art von Software?
- Welche Plattformen eignen sich für welche Art von Software?
- Wie können diese Ergebnisse zitierbar gemacht werden?
- Wie können Interpretation (Literatur), Daten und Software zugänglich gemacht werden – am besten als ausführbare Veröffentlichung.
- Sind vorhandene Plattformen geeignet, Software-Code als Teil der wissenschaftlichen Überlieferung langfristig zu bewahren?
- Wie werden Beiträge zu Open Source Projekten anerkannt?
- Welche Lizenzen sind für wissenschaftliche Software geeignet?

Das Projekt „SciForge“

Das DFG-Projekt „SciForge“ soll untersuchen, wie der Umgang mit wissenschaftlicher Software verbessert werden kann.

- Informationsmaterial und Unterstützung für einen nachhaltigeren Umgang mit wissenschaftlicher Software
- Erhöhung der Transparenz und Reproduzierbarkeit wissenschaftlicher Ergebnisse durch Zitierbarkeit von Software
- Entwicklung von Verfahren für die Veröffentlichung und Zitierbarkeit von Software
- Etablierung von wissenschaftlicher Software als Teil der wissenschaftlichen Überlieferung

Mögliche Ergebnisse aus „SciForge“

Software-Plattform

- Quellcode
- Versionierung
- Bündelung
- Persistente Identifikatoren

Software-Journal

- Dokumentation
- Referenz/Zitat
- Peer-review

Software wird Teil der wissenschaftlichen Überlieferung.

Mit diskutieren? mailto: software-l-on@gfz-potsdam.de
<http://dokuwiki.gfz-potsdam.de/datawiki/doku.php?id=software:start>

Fragen?



Vielen Dank für Ihre Aufmerksamkeit!

Mit diskutieren? mailto: software-l-on@gfz-potsdam.de
<http://dokuwiki.gfz-potsdam.de/datawiki/doku.php?id=software:start>

Literatur

Hatton, L., und A. Giordani (2012), Does computation threaten the scientific method?, *International Science Grid This Week*. [online] Available from: <http://www.isgtw.org/feature/does-computation-threaten-scientific-method>

Ince, D. C., L. Hatton, und J. Graham-Cumming (2012), The case for open computer programs, *Nature*, 482(7386), 485–488, doi: 10.1038/nature10836.

LeVeque, R. J. (2011), Top 10 Reasons to NOT Share your Code and Why you Should Anyway, Reno, NV. [online] Available from: http://faculty.washington.edu/rjl/talks/LeVeque_CSE2011.pdf

Morin, A., J. Urban, P. D. Adams, I. Foster, A. Sali, D. Baker, und P. Sliz (2012), Shining Light into Black Boxes, *Science*, 336(6078), 159–160, doi: 10.1126/science.1218263.

Peng, R. D. (2011), Reproducible Research in Computational Science, *Science*, 334(6060), 1226–1227, doi: 10.1126/science.1213847.