

# 45'883 Nationalfonds-Projekte für 10,208,843,482 Franken

## Open Government Data - Eine Fallstudie

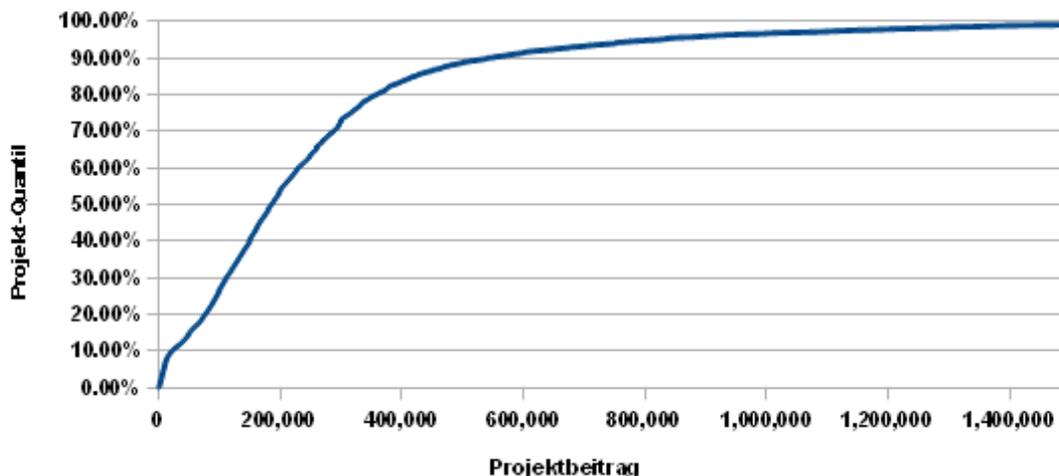
An der opendata.ch 2011Konferenz<sup>1</sup> entzündete sich am Beispiel der Nationalfonds-Daten eine Diskussion über nützliche und weniger nützliche Formate, in denen Behördendaten der Öffentlichkeit preisgegeben werden. Datenjournalisten klagten über eine 40-seitige PDF-Datei, die man nicht auswerten könne. Vertreter des Nationalfonds gaben zu bedenken, dass solche Daten auch missbraucht und fehlinterpretiert werden können, wenn die Auswertenden (Journalisten, Politiker, Öffentlichkeit) nicht genügend Kenntnisse hätten, um solche Daten zu verstehen.

Bei Betrachtung der vom Schweizerischen Nationalfonds (SNF) publizierten Angaben über die subventionierten Personen und Projekte<sup>2</sup> stellt man fest, dass dort die Behördendaten des Schweizerischen Nationalfonds recht vollständig veröffentlicht sind.

## Transparenz ist Ausblenden von Information

Verteilung der Projektbeiträge 1991-2010

Projektbeiträge bis 1,5 Mio Franken

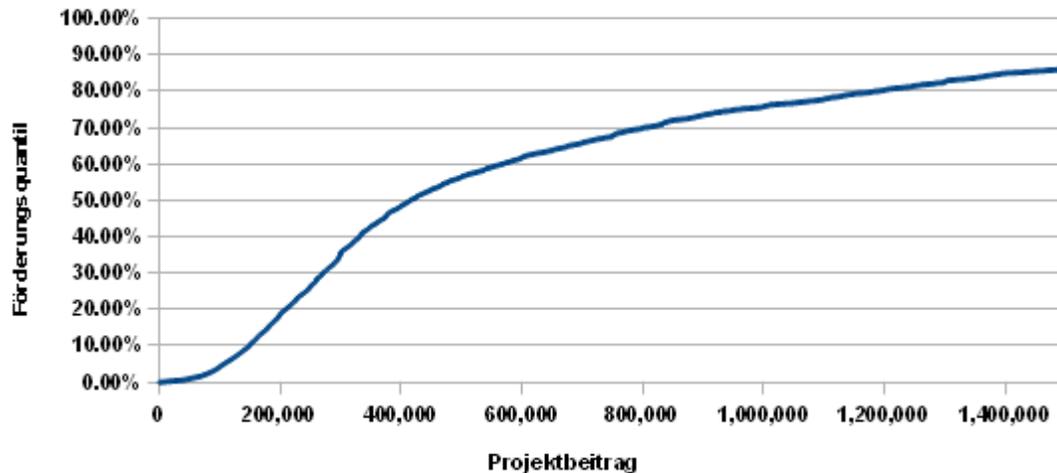


1 <http://opendata.ch/2011/06/opendata2011-medienmitteilung/>

2 <http://www.projectdb.snf.ch/WebForms/PersonProjectSearch.aspx>

## Anteil an Gesamtförderung nach Projektgrösse 1991-2010

### Projektbeiträge bis 1,5 Mio Franken



Das erste Diagramm zeigt, wie die Höhe der Projektbeiträge des Schweizerischen Nationalfonds von 1991 - 2010 verteilt ist. Das zweite Diagramm zeigt, welchen Anteil die Projekte bis zu einer gewissen Höhe des Projektbeitrags an der gesamten Förderungssumme im Zeitraum 1991-2010 erhielten. Bei beiden Diagrammen wurden die wenigen Projekte mit Beiträgen zwischen 1.5 bis 20 Millionen der Übersichtlichkeit halber ausgeblendet.

Diese Fallstudie konzentriert sich auf den schnellen Hack. Statistisch könnten diese Diagramme auf vielfältige Weise verbessert werden. So wären etwa die Projektbeiträge des SNF aus vier Jahrzehnten entsprechend der Kaufkraft des Schweizer Franken zu bereinigen oder in Beziehung zum Bruttosozialprodukt, der Anzahl Einwohner oder der Anzahl Forscher zu setzen.

Man kann aber auch schon aus dieser unbereinigten Version interessante Tatsachen ablesen. So ergibt sich etwa, dass Forschung einerseits aus einigen wenigen Schwerpunktprojekten besteht, andererseits aber der grösste Anteil der Forschungsgelder (75%) in die Masse der „kleinen“ Projekte (bis zu 1 Mio Franken) fliesst.

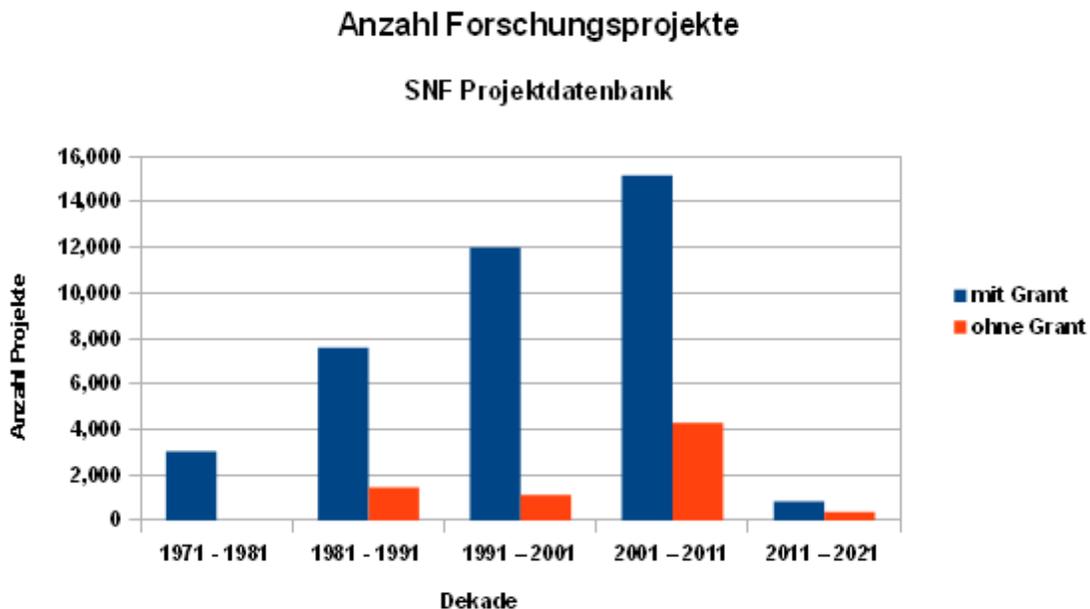
Das obenstehende Diagramm basiert auf den Daten des schweizerischen Nationalfonds, welche dieser auf der oben erwähnten Website publiziert hat. Wer allerdings mit Hilfe der auf dieser Site verfügbaren Daten obiges Diagramm herzustellen versucht, wird seine liebe Mühe damit haben. Die vom Nationalfonds implementierte - und sicher auch teuer bezahlte - Benutzerschnittstelle erlaubt nur wenige standardisierte Abfragen der Projektdatenbank und ist extrem sparsam mit statistisch auswertbaren Resultaten.

## Formate sind wichtig

Um also die obigen Diagramme zu erzeugen, war es notwendig, ein Skript zu schreiben, welches die ganze Datenbank des SNF über die Website ausliest. Die Erstellung eines solchen Skripts kostete rund fünf Arbeitstage. Das Herunterladen der gesamten Datenbank dauerte etwa fünf Stunden. Sie steht nun als MS Access Datenbank, bzw. als SIARD-Datei<sup>3</sup> zur freien Verfügung<sup>4</sup>.

Die verwendeten Skripts sind unter einer GPL-Lizenz am selben Ort gleichfalls frei verfügbar und können jederzeit verwendet werden, um den aktuellen Stand der SNF-Datenbank herunterzuladen – solange sich die Struktur der Website nicht ändert. Die Skripts wie auch die heruntergeladenen Daten sind etwas „struppig“, wie es bei einem schnellen Hack kaum vermeidbar ist: Die Daten wurden nicht von HTML-Elementen (Einheitenreferenzen und Break-Tags) gereinigt, die Skripts sollten anständiger modularisiert und ausführlicher getestet werden.

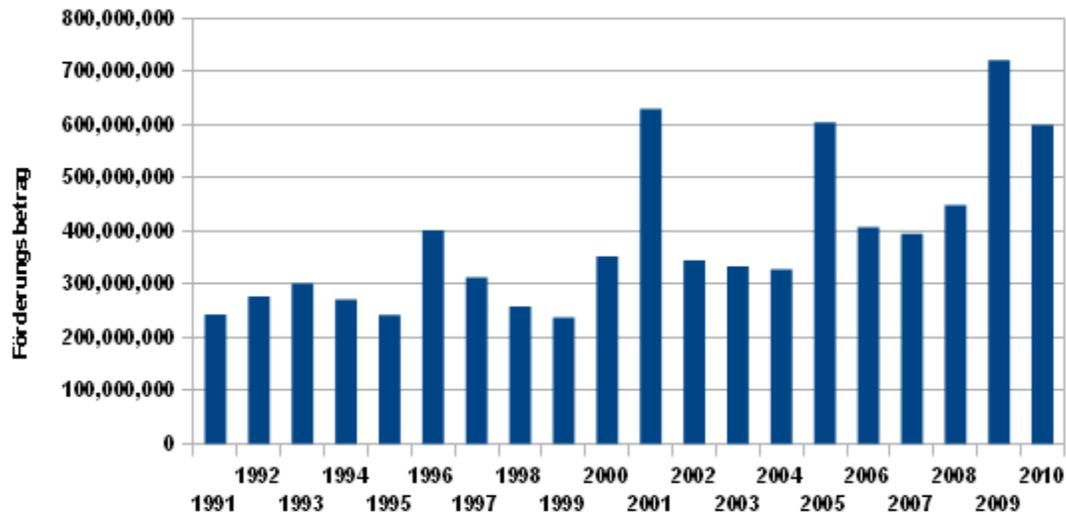
Wer sich diese Datenbank näher anschaut, ist zuerst einmal mit einer sehr grossen Menge von Information konfrontiert und verliert schnell die Übersicht. Mittels Ausblenden von Information schaffen wir mit ein paar Diagrammen etwas Transparenz:



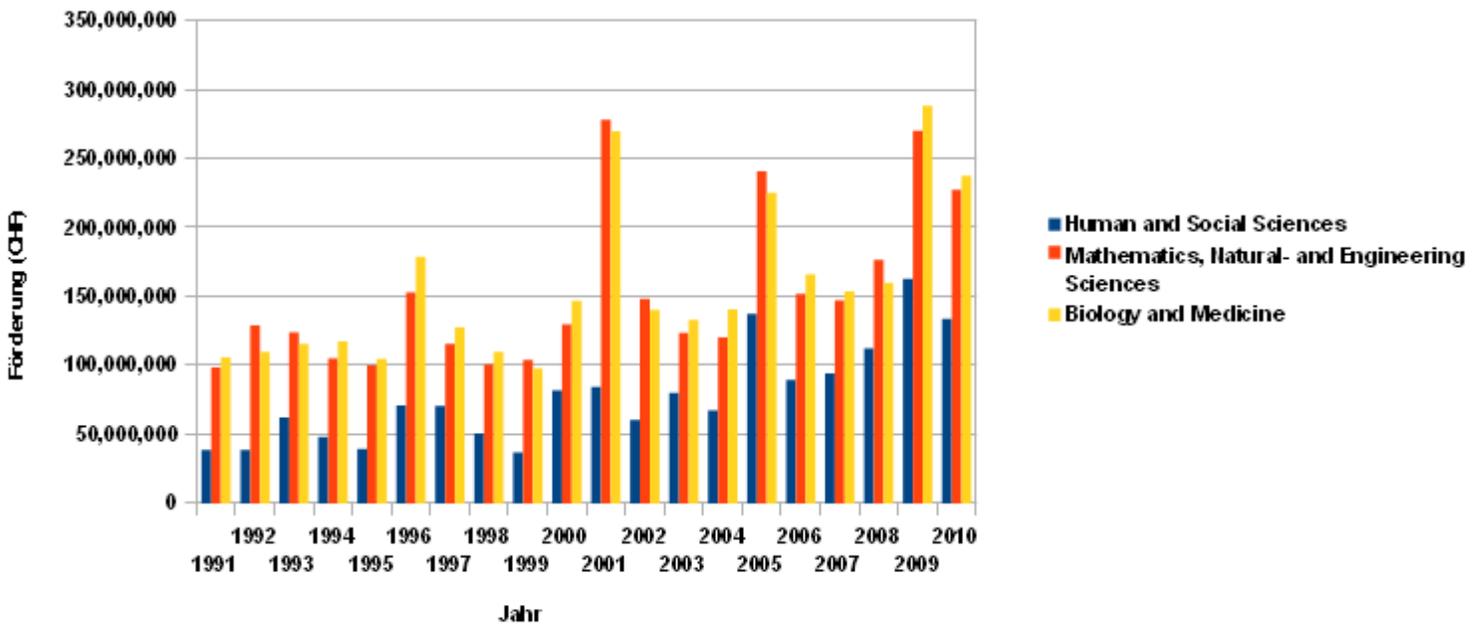
3 SIARD - Software-Independent Archival of Relational Databases wurde vom Schweizerischen Bundesarchiv entwickelt. Die Software „SIARD Suite“, mit deren Hilfe man eine SIARD-Datei in eine MS Access-, Oracle- oder SQL Server-Datenbank laden kann, kann beim Bundesarchiv kostenlos bezogen werden: <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html>.

4 <http://www.enterag.ch/enterag/downloads/index.xhtmll>

## Forschungsförderung 1991-2010



## Förderung 1991-2010 nach Forschungsrichtung



Eines der Probleme einer Opendata-Strategie für Behördendaten besteht darin, dass die Ämter ungern die Kontrolle darüber aufgeben, welche Information wie ausgeblendet wird. Die Verfügbarkeit der Rohdaten alleine schafft noch keine Transparenz. Diese wird erst durch den Datenjournalismus erzeugt, welcher allerdings die Rohdaten als Basis seiner Arbeit benötigt.

## **Datenjournalismus**

Die Verwandlung eines riesigen Datenhaufens in Transparenz ist Aufgabe eines sich momentan erst neu konstituierenden Datenjournalismus. Während die Aufgabe des Journalisten früher darin bestand, Information zu beschaffen, an die man oft nicht leicht herankam, verwandelt sie sich heute immer mehr in die Kunst des klugen Filterns und Ausblendens von im Überfluss vorhandener Information.

So könnte aus dem hier publizierten Datenhaufen der Nationalfondsprojekte herausgefiltert werden, welchen Anteil der Subventionen Fächer, Hochschulen, Förderprogramme erhalten haben und wie sich dieser Anteil über die Jahre entwickelt hat. In Korrelation mit externen Daten der universitären Open Access Repositories könnte die nachträgliche Zugänglichkeit der mit öffentlichen Geldern geförderten Resultate dargestellt werden. Wenigstens diese zugänglichen Daten könnten herangezogen werden, um die Wirkung der eingesetzten Forschungsgelder zu beurteilen.

Selbstverständlich ist nicht ausgeschlossen, dass manche Boulevard-Journalisten unzulässige Vereinfachungen in die Welt setzen und sich darüber amüsieren, wieviele Millionen investiert werden, um Wespen einen Farbpunkt aus Hinterteil zu malen, ohne dass ein erkennbarer Erkenntniszweck damit erreicht wird. Die „Reife“ der Benutzer und Weiterverarbeiter von Behördendaten kann niemand gewährleisten. Anhand der Rohdaten lässt sich aber immerhin überprüfen, wie unzulässig von Einzelnen publizierte Vereinfachungen sind. Transparenz ist also vor allem auch die Überprüfbarkeit der ausgeblendeten Information. Solange die Ämter keinen direkten Zugang zu ihren Datenbanken ermöglichen, ist diese Überprüfbarkeit nicht gegeben.

## **Formate sind unwichtig**

Offenbar genügt eine recht bescheidene Investition um die Rohdaten einer Analyse zugänglich zu machen. Wenn Datenjournalisten pro Diagramm so viel bezahlen würden wie pro Fotografie, wäre die Verwandlung eines ungünstigen Website-Formats in eines, welches vielfältige Analysen erlaubt, schnell finanziert.

Wenn die Rohdaten einmal verfügbar sind, können zahlreiche Auswertungen, wissenschaftlichsoziologische Studien und politische Analysen auf ihnen basieren.

Wenn eine geeignete Form so vielen Nutzern so grosse Vorteile bringt, wie immer behauptet wird, müsste es sich für diese ja lohnen, die paar Tage Aufwand zu finanzieren, welche die Konversion zugänglicher Daten in ein nützliches Format mit sich bringt.

Die in dieser Fallstudie heruntergeladene und öffentlich zugänglich gemachte Projektdatenbank des SNF illustriert also, dass das Jammern über ungeeignete Publikationsformen und -formate von Behördendaten an der Sache vorbeizieht.

## Das Infrastrukturparadox

Aber unsere Institutionen, ja unsere ganze Kultur schreckt davor zurück, Geld in das Etablieren von Infrastruktur zu investieren. Viele wissenschaftssoziologische Forschungsprojekte können viele Millionen Forschungsgelder für Studien beziehen, die auf den hier zugängliche gemachten Daten beruhen. Für deren Zugänglichmachen wird aber kein Rappen aufgewendet.

Dieses Phänomen ist nicht auf das Zugänglichmachen von Daten in unbrauchbaren Formate beschränkt.

- Kaum eine Forschungsförderung finanziert Infrastrukturaufwände für moderne Repositories (z.B.: Online-Verzeichnis der Beschreibungen der zoologischen und botanischen Arten der Welt <http://www.plazi.ch/>). Diese unterscheiden sich von traditionellen Journalsammlungen der Bibliotheken, weil sie jeweils einen wachsenden und zu pflegenden aktuellen Stand der Forschung reflektieren (Beispiel: Genomprojekt). Die Devise lautet: Wir fördern Forschung, nicht Infrastruktur. So gehen viele wissenschaftliche Rohdaten aus der teuren geförderten Forschung für immer verloren.
- Kaum eine Kulturförderung finanziert Infrastrukturaufwände für moderne Repositories digitaler Kunst (z.B.: Online-Repository freier Kultur <http://www.cultlib.ch/>). Diese unterscheiden sich von traditionellen Sammlungen der Bibliotheken und Museen, weil die digitale Archivtauglichkeit ein schwieriges Infrastrukturproblem darstellt. Die Devise lautet: Wir fördern Kultur, nicht Infrastruktur. So gehen viele moderne Kulturleistungen für immer verloren.
- Viele Stiftungen für die Unterstützung von Behinderten, verweigern die Finanzierung von Infrastruktur (z.B.: rollstuhlgängige Aufzüge, Barrierenfreiheit in Websites). Die Devise lautet: Wir helfen Behinderten, wir bauen keine Infrastruktur. So gehen wichtige Vorteile für alle (auch zukünftige) Behinderten verloren.

Die öffentliche Hand – aber nicht nur diese – will offenbar lieber Personen als Einrichtungen fördern. In einer solchen personenfixierten Promikultur wird ehrliche Arbeit gering geschätzt. Das war nicht immer so. Eigentlich ist wohl bekannt, dass Infrastrukturinvestitionen sehr wirksam sind. Der Zustand des schweizerischen Eisenbahnnetzes oder die Dichte der in der Schweiz vorhandenen Memoinstitutionen belegt, dass diese Tatsache früher bekannt war. Auf die digitale Welt wird diese Erkenntnis aber offenbar nicht übertragen.

Weltweit gibt es sehr grosse Bibliotheken, Museen, Archive, die alle grösstenteils von öffentlichen Geldern getragen werden. Für diese Einrichtungen wurden in einer Zeit Gelder zur Verfügung gestellt, als man Infrastruktur noch als wichtigen Faktor der Kultur wahrnahm. Diese traditionellen Memoinstitute haben heute Mühe, den strukturellen Wandel der Digitalisierung der Kultur nachzuvollziehen. Ihr ureigenstes Kerngeschäft ist zwar das nachhaltige Verfügbarmachen von Information zum Nutzen der gesamten Öffentlichkeit. Sobald es aber um digitale Information geht, tun sie sich schwer damit. Ausserdem litten sie in den letzten selber unter dem Infrastrukturparadox. Ihre

Budgets wurden reduziert, während sie für die Erledigung ihrer Aufgaben eigentlich grössere Aufwände benötigen würden.

Einerseits ist heute fast alle neue Information schon bei der Entstehung digital. Andererseits stellen sich deshalb heute andere Probleme beim Verfügbarmachen solcher Information.

## **Das Offenlegen von Behördendaten lohnt sich**

Eine SIARD-Datei zum Download bereitzustellen ist sehr viel preiswerter als die Entwicklung einer Webanwendung wie diejenige des Schweizerischen Nationalfonds. Die Entwicklung der letzteren dürfte zwischen 50'000 und 250'000 Franken gekostet haben. Der regelmässige Abzug einer SIARD-Datei aus der SNF-Datenbank und deren Bereitstellung zum Download dürfte etwa 500 Franken pro Jahr kosten. Durch Publikation der Rohdaten im für den Datenjournalismus geeigneten Format lassen sich also von den Behörden mehr als 99.5% der Ausgaben für die Präsentation der Daten einsparen.

Nicht nur die Subventionierung der wissenschaftlichen Forschung, auch diejenige des kulturellen Schaffens (Bundesamt für Kultur, Filmförderung, ...) oder der Landwirtschaft sind von grossem öffentlichem Interesse. Das Bundesamt für Statistik (BfS) gibt zwar bescheidene Auskunft über diese Behördenaktivitäten. Seine Auswahl, welche Aspekte auszublenden sind, ist aber eher vom Behörden- als vom öffentlichen Interesse geprägt. Das sieht man leicht, wenn man die statistischen Angaben des BfS zur Forschungsförderung mit der hier publizierten Datenbank vergleicht.

Wenn also die Ämter eine konsequente Opendata-Strategie einschlagen, können auf *admin.ch* viele Steuergelder eingespart werden, die heute in die jeweils eigene Webentwicklung des Auftritts eines jeden Amtes fliessen. Man muss dafür allerdings auf die verkrampte Kontrolle der dem Medienkonsumenten aufgedrängten Interpretation der Daten verzichten, die man ohnehin nicht aufrechterhalten kann, weil sie mit einfachen Skripten auszuhebeln ist. Offen ist offen, egal in welchem Format. Die Herstellung von Transparenz mittels Ausblenden kann - wie im vordigitalen Zeitalter - ruhig den Medien (dem Datenjournalismus) und somit der Öffentlichkeit überlassen werden.

Datum: Zürich. 14. Juli 2011

Autor: Hartwig Thomas

Lizenz:  <http://creativecommons.org/licenses/by-sa/2.5/ch/>