

NEPS-Daten mit DOIs referenzieren

Wenzig, Knut

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Empfohlene Zitierung / Suggested Citation:

Wenzig, K. (2012). *NEPS-Daten mit DOIs referenzieren*. (RatSWD Working Paper Series, 202). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://hdl.handle.net/10419/75343>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Wenzig, Knut

Working Paper

NEPS-Daten mit DOIs referenzieren

RatSWD Working Paper, No. 202

Provided in Cooperation with:
German Data Forum (RatSWD)

Suggested Citation: Wenzig, Knut (2012) : NEPS-Daten mit DOIs referenzieren, RatSWD Working Paper, No. 202, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin

This Version is available at:

<http://hdl.handle.net/10419/75343>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

RatSWD Working Paper Series

www.ratswd.de

RatSWD 

Rat für Sozial- und
Wirtschaftsdaten

202

NEPS-Daten mit DOIs referenzieren

Knut Wenzig

Juli 2012

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

NEPS-Daten mit DOIs referenzieren*

Knut Wenzig†

Inhaltsverzeichnis

1	Motivation	2
2	Aufbau der NEPS-DOIs	2
2.1	Aktueller Stand	2
2.2	Entwicklungsmöglichkeiten	4
3	Registrierung der DOIs bei der da ra	5
3.1	Aktueller Stand der übermittelten Metainformationen	5
3.2	Entwicklungsmöglichkeiten	5
4	Die <i>landing page</i>	6
5	Ausblick	6
6	Anhang	8
6.1	XML zur Registrierung der DOI	8
6.2	<i>Landing page</i> für doi:10.5157/NEPS:SC6:1.0.0	9

* Eine frühere Version dieses Textes wurde am 8. Mai 2012 auf der Tagung “Metadata and Persistent Identifiers for Social and Economic Data” in Berlin vorgestellt. Für Anmerkungen und Hinweise zu früheren Versionen danke ich den Teilnehmer_innen sowie Eva Akins, Tobias Koberg und Jan Skopek aus dem NEPS-Datenzentrum.

† Nationales Bildungspanel, Datenzentrum, Wilhelmsplatz 3, 96047 Bamberg, knut.wenzig@uni-bamberg.de

1 Motivation

Das Nationale Bildungspanel (Blossfeld, Roßbach und von Maurice 2011) führt 6 Panelstudien, in denen jeweils eine Startkohorte begleitet wird, und 2 Zusatzstudien durch. Die Daten aus diesen Erhebungen werden zur wissenschaftlichen Nutzung zur Verfügung gestellt. Damit sind Untersuchungen zu Kompetenzentwicklung, Bildungsprozessen, Bildungsentscheidungen und Bildungsergebnissen in formalen, nicht-formalen und informellen Kontexten über die gesamte Lebensspanne möglich.

Die hieraus veröffentlichten Ergebnisse quantitativer Datenanalyse sollten zur Sicherung guter wissenschaftlicher Praxis auch der Reanalyse zugänglich gemacht werden. Dass in der Veröffentlichung die jeweils verwendeten Daten korrekt zitiert und auf sie mit Hilfe dieser Angaben auch zugegriffen werden kann, ist hierfür unabdingbare Voraussetzung, die durch das Datenzentrum des Nationalen Bildungspanels (NEPS) geschaffen wird.

Um dieses Ziel zu erreichen, werden *persistent identifier* (PID) empfohlen, weil sie neben der eindeutigen Identifikation des verwendeten Datenmaterials folgende Vorteile bieten: einfache Zitation, Sichtbarkeit der Forschung, Nachweismöglichkeit des Impacts (Datenzitationsraten), Nachvollziehbarkeit der Forschung, Unterstützung für Sekundäranalyse, Sichtbarkeit für Datenanbieter (Hausstein, Zenk-Möltgen, Wilde und Schleinstein 2011, S. 7). Werden zu diesem Zweck DOIs (*digital object identifier*) verwendet, kann über den PID auch der Weg zu den Daten geebnet werden: Das DOI-System verknüpft den Identifikator mit der Lokation eines Objekts (Hausstein 2012, S. 3) Dieses Papier dokumentiert, wie das Datenzentrum des NEPS DOIs aktuell einsetzt und skizziert wie das NEPS heute und in Zukunft davon profitieren kann.

Derzeit erfolgt die Registrierung der DOIs über die da|ra, der Registrierungsagentur von GESIS. Hierzu wurde am 23.05.2011 ein *service level agreement* unterzeichnet, das am 27.02.2012 um eine Vereinbarung über die Weitergabe der zugehörigen Metadaten ergänzt wurde. Die GESIS fungiert hierbei als Mitglied im DataCite-Konsortium, das seinerseits Mitglied in der International DOI Foundation ist.

Neben DOIs stehen grundsätzlich auch andere PIDs zur Verfügung, etwa URNs (*uniform resource name*) oder ARKS (*archival resource key*). Der Entscheidung der GESIS, die als großes deutsches Infrastrukturinstitut für die Sozialwissenschaften auch für den eigenen Datenbestand DOIs als PID verwendet, wurde in so fern eine gewisse Leitbildfunktion zugerechnet und Alternativen nicht näher untersucht.¹

2 Aufbau der NEPS-DOIs

2.1 Aktueller Stand

DOIs bestehen aus einem Präfix und einem Suffix, das durch einen Schrägstrich abgetrennt wird. Das Präfix wird von der da|ra für den *publication agent* bzw. *registrant* (also das NEPS) vergeben. Es besteht aus einer 10 (als sog. *directory indicator*), einem Punkt und einer vierstelligen Nummer, dem *registrant code*. Das Präfix für das NEPS lautet „10.5157“. Das Suffix ist prinzipiell frei wählbar: Es steht der Zeichenvorrat UCS-2 zur Verfügung, die Länge ist beliebig (International DOI Foundation 2012). Bei der Vergabe sollte jedoch berücksichtigt werden, dass die DOI ggfs. auch per Tastatur eingegeben werden soll, was die Verwendung beispielsweise von Umlauten als wenig empfehlenswert erscheinen lässt.

¹ Die Überlegung, dass für die Eignung eines PID-Typs „weniger die technische als die organisatorische Ausgestaltung relevant“ (Hausstein 2012, S. 3) sei, scheint sehr plausibel.

In vielen Bereichen sind systemfreie Identifikatoren zu bevorzugen und Informationen über etwaige Eigenschaften der zu identifizierenden Objekte sollten in zugehörigen Metadaten abgelegt und verarbeitet werden. Der Identifikator selbst ist dann nicht sprechend, was seine nicht-automatische Verwendung möglicherweise erschwert. Nichtsdestotrotz sollte von dem Grundsatz nur abgewichen werden, wenn die Struktur der zu identifizierenden Objekte (möglichst) abschließend bekannt ist.

Für die *scientific use files* (SUF) des NEPS trifft dies nach heutiger Einschätzung zu. Diese SUFs sind Zusammenstellungen von mehreren Einzeldatensätzen, mit folgenden Eigenschaften:

- Das NEPS veröffentlicht in der ersten Förderphase von 2008 bis 2013 die SUFs für sechs Panelstudien und für zwei Zusatzstudien.
- Mit Ausnahme der Zusatzstudie für Thüringen werden im zeitlichen Verlauf für alle Studien weitere SUFs zu veröffentlichen sein, die im wesentlichen den Panelwellen entsprechen. Die Daten aus den Panelwellen (bzw. den aufeinanderfolgenden Querschnittserhebungen in Baden-Württemberg) werden jeweils in einer möglichst integrierten Form veröffentlicht, die Daten von allen vorherigen Erhebungen umfasst.
- Um künftige Aktualisierungen der einzelnen SUFs verwalten zu können, verwendet das NEPS-Datenzentrum ein 3-stelliges Versionierungssystem. Die erste Stelle entspricht hierbei in der Regel der Welle (*main release*), die anderen beiden Stellen werden für *major* und *minor updates* verwendet.
- Die Daten des NEPS werden auf drei Zugangsplattformen veröffentlicht (*download*, RemoteNEPS und *on-site*). Die hier jeweils zur Verfügung stehenden Daten unterscheiden sich lediglich im Anonymisierungsgrad.
- Die SUFs werden derzeit für zwei Softwareplattformen zur Verfügung gestellt: Stata und SPSS.
- Von allen produzierten Datensätzen wird eine deutsche und – mit Blick auf die internationale *scientific community* – eine englische Variante produziert, die sich etwa in der Sprache der *label* von Variablen und Werten unterscheiden.

Die Frage der angemessenen Granularität der DOIs ist in der entsprechenden Fachöffentlichkeit Gegenstand intensiver Debatten. Es empfiehlt sich hierbei das zu unterscheiden, was unterschieden werden soll.² Wenn die Replikation von Forschungsergebnissen im Vordergrund steht, bedeutet dies zunächst, dass Varianten für unterschiedliche Sprachen und Softwareplattformen nicht zu unterscheiden werden brauchen: Dass die (erfolgreiche) Replikation von Forschungsergebnissen nicht von der Sprachversion der Datensätze abhängen sollte, wird ebenfalls unmittelbar einleuchten. Und gerade weil Analysealgorithmen in Softwarepaketen nicht selten als *black box* wahrgenommen werden, sollten die Schätzergebnisse für ein Modell von der verwendeten Software unabhängig sein bzw. sich Unterschiede mit Blick auf die jeweilige Implementierung wenigstens aufklären lassen.

Nicht ganz so eindeutig mag die Entscheidung ausfallen, ob die Zugangsplattform und damit der Anonymisierungsgrad in der DOI berücksichtigt werden soll.³ Nicht zuletzt wegen der – gemessen an der Gesamtzahl der Variablen – nur marginalen Modifikation in den Datensätzen,

² “The indecs Principle of Functional Granularity is that ‘it should be possible to identify an entity whenever it needs to be distinguished.’” (Paskin 2003)

³ Das Anonymisierungskonzept des NEPS sieht vor, dass Variablen, deren Inhalt sich auf den drei unterschiedlichen Zugangsplattformen unterscheidet, mit bis zu drei Suffixen (O, R, D) in den Datensätzen verfügbar sind. Eine Variable, die beispielsweise mit vollständiger Information nur innerhalb von RemoteNEPS (und damit auch *on-site*) und im Download-SUF mit einer vergrößerten Information zur Verfügung steht, gibt es in der Variante D und R. Der Inhalt der Variable mit dem Suffix R wird in der *download*-Variante des Datensatzes geleert.

Tabelle 1: Durch das NEPS bereits registrierte und künftige DOIs für SUFs.

DOI	Startkohorte	Version	Anmerkung	registriert
doi:10.5157/NEPS:SC2:1.0.0	SC2 (Kindergarten)	1.0.0	1 Welle	nein
doi:10.5157/NEPS:SC3:1.0.0	SC3 (Klasse 5)	1.0.0	1 Welle	nein
doi:10.5157/NEPS:SC4:1.0.0	SC4 (Klasse 9)	1.0.0	2 Wellen	nein
doi:10.5157/NEPS:SC5:1.0.0	SC5 (Studierende)	1.0.0	noch unbestimmt	nein
doi:10.5157/NEPS:SC6:1.0.0	SC6 (Erwachsene)	1.0.0	2 Wellen	ja
doi:10.5157/NEPS:SC6:1.0.1	SC6 (Erwachsene)	1.0.1	2 Wellen	nein
doi:10.5157/NEPS:SC6:2.0.0	SC6 (Erwachsene)	2.0.0	3 Wellen	nein
doi:10.5157/NEPS:TH:1.0.0	Thüringen	1.0.0	2 Querschnitte	nein
doi:10.5157/NEPS:G8:1.0.0	G8-Reform	1.0.0	1 Querschnitt	nein

scheint es wieder aus der Replikationsperspektive angemessen, diese Varianten ebenfalls nicht zu unterscheiden.

Faktisch bedeutet dies, dass im NEPS die DOI allen genannten Varianten aus Sprache, Softwareplattform und Zugangsplattform gleichzeitig zugeordnet ist. Das durch die DOI referenzierte digitale Objekt ist damit nicht nur eine *collection* von einzelnen Datenfiles, sondern auch eine *collection* von Varianten dieser Datenfiles. Während *on-site* jede dieser Varianten zur Verfügung steht, wird im *download* nur ein Teil des durch die DOI referenzierten Materials zugänglich gemacht. Mit dem Blick auf das Zwiebelprinzip der Datenstruktur (vgl. Koberg 2011, S. 1–3) wird also die ganze Zwiebel (mit den Schichten O, R und D) referenziert und nicht nur der Teil, der mit dem jeweiligen Zugangsweg korrespondiert.

Im Ergebnis kann man einer vom NEPS für die SUFs verwendeten DOI, außer den Datenproduzenten auch die Startkohorte bzw. die Zusatzstudie, die Welle (zumindest indirekt) und die Version des SUF entnehmen. `10.5157/NEPS:SC6:1.0.0` lautet die erste vom NEPS registrierte DOI. Es war der erste Hauptrelease von Daten für die Startkohorte 6 (Erwachsene), der mit der ersten Haupterhebung des NEPS und den Informationen aus der Vorläuferstudie ALWA jedoch bereits Informationen aus zwei Panelwellen enthält (vgl. Tabelle 1).

2.2 Entwicklungsmöglichkeiten

Während die Gestalt der DOIs für SUFs im NEPS damit auf mittlere Sicht konzeptionell festgelegt ist, sind damit weitere zukünftige Verwendungsmöglichkeiten nicht ausgeschlossen:

- Identifikation für digitale Objekte im Rahmen von *enhanced publication*: Wenn Forschungsergebnisse mit NEPS-Daten veröffentlicht wurden, stellt sich zunehmend die Frage, wie begleitendes Material archiviert und identifiziert werden kann. Denkbar wäre die Zuordnung einer DOI zu *code* und Datenmaterial, die für eine Veröffentlichung benötigt und erstellt wurden.
- PID für Veröffentlichungen des NEPS: Mehr als ein Dutzend Dokumente begleiten den ersten SUF der Startkohorte 6 als Dokumentationsmaterial, teilweise in deutscher und englischer Fassung. Weiterhin entstehen im NEPS – insbesondere in der Reihe „Working Paper“ – Veröffentlichungen, die dauerhaft zur Verfügung gestellt werden sollen. Auch für diese Materialien wäre es möglich, DOIs zu vergeben und sie damit langfristig auffindbar zu machen.⁴

⁴ DOIs hierfür können ebenfalls bei der `da|ra` registriert werden. Alternativ hierzu bietet es sich an, die Dokumente in einem *open access repository*, wie z. B. SSOAR (`ssoar.info`) oder PEDOCs (`pedocs.de`), zu hinterlegen und die hier vergebene URN als PID zu nutzen.

3 Registrierung der DOIs bei der da|ra

3.1 Aktueller Stand der übermittelten Metainformationen

Mit der da|ra wurde vereinbart, zunächst nur die Pflichtinformationen (Hausstein, Zenk-Möltgen, Wilde und Schleinstein 2011, S. 9) bereitzustellen, die für die Registrierung einer DOI unabdingbar sind. Dies sind derzeit:

- Titel (de/en)
- Projektleiter inkl. *affiliation*
- Version
- Veröffentlichungsdatum
- URL der *landing page*
- Hinweise zur Verfügbarkeit (de/en)
- Publikationsagent inkl. Ansprechpartner
- DOI-Vorschlag
- Studien-ID (entspricht Suffix)

Ein aktuelles XML-Dokument findet sich im Anhang (S. 8).

Die derzeitigen Planungen des NEPS-Datenzentrums sehen vor, dass der bei der da|ra verfügbare Datenbestand die *single source of information* der Metainformation zu den DOIs darstellt. Angesichts der relativ kleinen Menge von zu erwartenden DOIs erscheint eine Datenbank gestützte Verarbeitung im NEPS selbst zunächst als wenig effizient.⁵

Die Pflege der Metadaten erfolgt derzeit durch per E-Mail vom NEPS an die da|ra übermittelte XML-Dateien. Künftig können diese Informationen über das Portal der da|ra mittels Webformularen gepflegt werden oder komplett als Upload übermittelt werden. Die da|ra will auch entsprechende Webservices anbieten.

3.2 Entwicklungsmöglichkeiten

In wie weit diese Metadaten angereichert werden, wird insbesondere davon abhängen, ob weitere Informationen auch in entsprechenden Portalen die Auffindbarkeit von NEPS-Datensätzen erleichtern. Denkbar sind in dieser Beziehung etwa die Ergänzung mit entsprechenden Begriffen aus *controlled vocabularies*, die den Inhalt der Datensätze und damit des Befragungsprogramms beschreiben, sowie Informationen über die zeitliche und räumliche Ausdehnung der Befragung. Eine mächtige Erweiterungsoption könnte die Verknüpfung mit gleichzeitig bereitgestellten *rich metadata* sein, die auf Datensatz- und Variablenebene Metainformationen liefern. Auch dies wird aber davon abhängen, ob aktuelle oder künftige Nutzerinnen und Nutzer über geeignete Suchfunktionen zuverlässiger auf Anwendbarkeit von NEPS-Daten zur Beantwortung ihrer Forschungsfragen hingewiesen werden.

Ein wichtiger Erfolgsmesser für einen Datenlieferanten wie das Nationale Bildungspanel ist die Akzeptanz der Daten in der *scientific community* und ihre Verwendung in wissenschaftlichen Veröffentlichungen. Nachdem wissenschaftliche Veröffentlichungen selbst zunehmend mit PIDs – etwa DOIs oder URNs – ausgestattet sind, bietet es sich an, die Verknüpfung zwischen Daten

⁵ Das schließt nicht aus, die an die da|ra gelieferten XML-Dateien automatisiert zu erzeugen.

und Veröffentlichungen auch tatsächlich herzustellen. Das Metadatenchema der da|ra sieht eine solche Möglichkeit in absehbarer Zeit vor und es erscheint aus heutiger Sicht zweckmäßig, davon auch Gebrauch zu machen.

Weiterhin erscheint es sinnvoll, die über die Versionsnummer sichtbare Verwandtschaft von SUFS in den Metadaten abzubilden. Wenn diese Beziehung über die Metadaten expliziert wird, ist es möglich, automatisch auf neuere Versionen aufmerksam machen zu können. Die da|ra plant, genau dies zu ermöglichen und in den Metadaten einer DOI hinterlegen zu können, unter welcher (anderen) DOI beispielsweise eine neue Version verfügbar ist.

Schließlich werden auf der *landing page* (s. u.) schon heute weitere Informationen angezeigt, die auch in die Metainformationen aufgenommen werden sollten:

- Feldzeit bzw. -zeiten
- Zitationshinweise
- Zusammenfassung
- Links zur Dokumentation

Die Metadateninfrastruktur der da|ra befindet sich derzeit auf dem Weg zur Version 2. Dann wird es auch erweiterte Möglichkeiten für Publikationsagenten geben, ihre Metadaten zu erfassen und anzuzeigen. Sobald diese zur Verfügung stehen, werden die skizzierten Ergänzungen der Metainformationen auf Machbarkeit überprüft und ggfs. schrittweise umgesetzt.

4 Die *landing page*

Mittels der DOI erreicht man eine sogenannte *landing page*, die in den NEPS-Internetauftritt integriert ist. Diese enthält derzeit zunächst die Metainformationen, die der da|ra übermittelt wurden. Grundlage für diese Darstellung sollten auf lange Sicht die von der da|ra per *webservice* etwa in Form einer XML-Datei bezogenen Informationen sein (vgl. Abbildung auf S. 9).

Weiterhin wird sämtliches Dokumentationsmaterial mittels einer komprimierten Übersicht zugänglich gemacht. Angemeldete Nutzer_innen können, abhängig von ihren Rechten, die Download-Variante des SUFS nach dem nächsten Klick direkt herunterladen oder RemotenePS starten.

Künftig könnte ein Link zum NEPSplorer⁶, in dem eine Suche in den, auf den entsprechenden SUF eingeschränkten, NEPS-Metadaten ermöglicht, genauso eine interessante Ergänzung sein, wie ein Hinweis auf eventuell verfügbare neuere Versionen.

5 Ausblick

Der unmittelbare Nutzen von PIDs aus der Perspektive einer wissenschaftlichen Organisation wie dem NEPS besteht zunächst einmal in der verbesserten Sichtbarkeit der Datenprodukte⁷

⁶ Der NEPSplorer ist ein Online-Angebot des Datenzentrums, das die vollständigen Metadaten (Fragen, Variablen etc.) der NEPS-SUFS erschließt: <https://www.neps-data.de/de-de/datenzentrum/studiendokumentation/nepsplorer.aspx>

⁷ Kürzlich hat Thomson Reuters den *Data Citation Index* für große Datensätze angekündigt. Der Index soll in das *Web of Knowledge* integriert sein und noch in diesem Jahr online gehen: "This new research resource from Thomson Reuters creates a single source of discovery for scientific, social sciences and arts and humanities information by connecting foundational research within data repositories around the world to related peer-reviewed literature in journals, books, and conference proceedings already indexed in the Web of Knowledge." (Pressemitteilung vom 22.06.2012)

und einer direkten Messung eines *impact*-Indikators⁸.

Die Verwendung von dauerhaften Identifikationsmerkmalen wie DOIs für Datensätze ist sicherlich eine notwendige Bedingung für eine gute wissenschaftliche Praxis. Diese Praxis kann weiter optimiert werden durch die Verknüpfung mit anderen Produkten, die diese Datensätze als Dokumentationsmaterial begleiten oder mit diesen Datensätzen erzeugt wurden. Aber nur die dauerhafte Archivierung dieser referenzierten digitalen Objekte sichert langfristig die Ergebnisse der Anstrengungen einer Wissenschaft, die auf Replikation ihrer Ergebnisse setzt, um sich selbst zu verbessern.

⁸ Die Ausbeute an Veröffentlichungen ist nicht das einzig denkbare Maß für den *impact* eines wissenschaftlichen Forschungsprojekts.

6 Anhang

6.1 XML zur Registrierung der DOI

```

<?xml version="1.0" encoding="UTF-8"?>
<study xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="
dara_v1_0_17082011.xsd">
  <titles>
    <title>
      <language>de</language>
      <title>NEPS-Startkohorte Erwachsene (SC6): Bildung im Erwachsenenalter und
      lebenslanges Lernen</title>
    </title>
    <title>
      <language>en</language>
      <title>NEPS-Starting Cohort Adults (SC6): Adult Education and Life-Long Learning</
      title>
    </title>
  </titles>
  <principalInvestigators>
    <principalInvestigator>
      <name>Hans-Peter Blossfeld</name>
      <affiliation>University of Bamberg, Germany</affiliation>
    </principalInvestigator>
    <principalInvestigator>
      <affiliation>
        NEPS, Nationales Bildungspanel/National Educational Panel Study, Bamberg (
        Germany)
      </affiliation>
    </principalInvestigator>
  </principalInvestigators>
  <currentVersion>1.0.0</currentVersion>
  <publicationDate>2011-12-15 00:00:00.0 MEZ</publicationDate>
  <urls>
    <dataURL>
      <dataURL>https://portal.neps-data.de/en-us/datacenter/researchdata/
      startingcohortadults/105157nepssc6100.aspx</dataURL>
    </dataURL>
  </urls>
  <availability>
    <availability>
      <language>de</language>
      <availability>
        Daten und Dokumente sind für die wissenschaftliche Nutzung bestimmt. Dieser
        Release wird durch eine DOI identifiziert. Alle Datensätze dieses Releases sind
        in unterschiedlichen Sprachen (englisch, deutsch), Formaten (Stata, SPSS) und
        Anonymisierungsgraden (Download, Remote Access und On-Site) verfügbar.
      </availability>
    </availability>
    <availability>
      <language>en</language>
      <availability>
        Data and documents are released for academic research. This release is uniquely
        identified by one DOI. All data sets of this release are available in different
        languages (English, German), formats (Stata, SPSS), and degrees of anonymization
        (download, remote access, and on-site).
      </availability>
    </availability>
  </availability>
  <publicationAgent>
    <name>Nationales Bildungspanel - National Educational Panel Study (NEPS) Otto-Friedrich-
    Universität Bamberg</name>
    <contactPerson>Knut Wenzig</contactPerson>
  </publicationAgent>

```

```
<email>Knut.wenzig@uni-bamberg.de</email>
<homepage>https://www.neps-data.de</homepage>
</publicationAgent>
<doiProposal>10.5157/NEPS:SC6:1.0.0</doiProposal>
<studyId>NEPS:SC6:1.0.0</studyId>
</study>
```

6.2 Landing page für doi:10.5157/NEPS:SC6:1.0.0

The screenshot shows a Firefox browser window displaying the NEPS Data Center website. The page title is "Starting Cohort 6 - Adult Education and Life-Long Learning". The page is sponsored by the Federal Ministry of Education and Research. The main content area includes a navigation menu on the left, a central text area with details about the study (Version 1.0.0, Date of Release 2011-12-22, Field time 2007-01-2008-04-30 and 2009-11-2010-08-03), Principal Investigator Hans-Peter Blossfeld, and a citation reference. A summary section is also present. On the right side, there are three panels: "Starting Cohort Adults" with a photo, "Download the Data" with a link to "SUF SC6 Version 1.0.0", and "Data Access via RemoteNEPS" with instructions for registered users. A "Documentation (Version 1.0.0)" panel lists various documents like the Data Manual, Questionnaire, and Codebook.

Abbildung 1: Die landing page aus der Sicht eines Datennutzers mit Datennutzungsvertrag ohne RemotenEPS-Ergänzung.

Literatur

Blossfeld, Hans-Peter, Hans-Günther Roßbach und Jutta von Maurice, Hrsg. (2011). *Education as a lifelong process*. Zeitschrift für Erziehungswissenschaft, Sonderheft 14. VS Verlag für Sozialwissenschaften.

Hausstein, Brigitte (2012). *Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten. Serviceleistungen der Registrierungsagentur da/ra*. RatSWD Working Paper Series 193.

Hausstein, Brigitte, Wolfgang Zenk-Möltgen, Anja Wilde und Natalija Schleinstein (2011). *da/ra Metadatenchema. Version 1.0*. Working Papers 2011|14. DOI: 10.4232/10.mdsdoc.1.0.

International DOI Foundation, Hrsg. (2012). *DOI Handbook*. DOI: 10.1000/182.

Koberg, Tobias (2011). *Starting Cohort 6: Adults (SC6), SUF-Version 1.0.0, Data Manual [Supplement E], Anonymization Procedures*. Nationales Bildungspanel (NEPS). Bamberg.

Paskin, Norman (2003). „On Making and Identifying a ‚Copy““. In: *D-Lib Magazine* 9.1. DOI: 10.1045/january2003-paskin.