

# **Neue Wege des Publizierens: Die Zitierfähigkeit wissenschaftlicher Primärdaten**

**Jan Brase**

Forschungszentrum L3S, Universität Hannover

**Michal Diepenbroek**

World Data Center for Marine Environmental Sciences,  
MARUM Universität Bremen

**Hannes Grobe**

Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven

**Heinke Höck**

World Data Center Climate,  
Max-Planck-Institut für Meteorologie Hamburg

**Jens Klump**

GeoForschungsZentrum Potsdam

**Michael Lautenschlager**

World Data Center for Climate,  
Max-Planck-Institut für Meteorologie Hamburg

**Uwe Schindler**

World Data Center for Marine Environmental Sciences,  
MARUM Universität Bremen

**Irina Sens**

Technische Informationsbibliothek, Hannover

## ***Problematik***

Der Zugang zu wissenschaftlichen Primärdaten ist eine grundlegende Voraussetzung für die Forschungsarbeit vor allem in den Naturwissenschaften. Deshalb ist es notwendig, bestehende und zum Teil auch neu aufkommende Einschränkungen bei der Datenverfügbarkeit zu vermindern. Im wissenschaftlichen Bereich besteht zwar grundsätzlich Bereitschaft, Daten für eine interdisziplinäre Nutzung zur Verfügung zu stellen, aber es ist zur Zeit unüblich, dass die erforderliche Mehrarbeit für Aufbereitung, Kontextdokumentation und Qualitätssicherung im Wissenschaftsbetrieb anerkannt wird. Die klassische Form der Verbreitung wissenschaftlicher Ergebnisse ist ihre Veröffentlichung in Fachzeitschriften, normalerweise ohne Veröffentlichung der zugrunde liegenden Daten. Diese klassische Publikation wird im "Citation Index" erfasst. Dieser Index wird zur Leistungsbewertung von Wissenschaftlern herangezogen. Datenveröffentlichungen werden darin bisher nicht berücksichtigt.

Projektdateien sind breit über Forschungsinstitute verstreut und werden von Wissenschaftlern erhoben und meist selbst verwaltet. Aufgrund der fehlenden Anerkennung der mit der

Aufbereitung verbundener Arbeit sind Projektdaten häufig schlecht dokumentiert und somit schwer zugänglich sowie nicht langfristig gesichert. Große Datenbestände bleiben ungenutzt, da sie nur einen kleinen Kreis von Wissenschaftlern bekannt und zugänglich sind. Viele Primärdaten bleiben ungenutztes Rohmaterial und sind häufig nach wenigen Jahren verloren.

Die Diskussion um die Fälschung wissenschaftlicher Ergebnisse führte zu Empfehlungen in den ‚Regeln guter wissenschaftlicher Praxis‘ durch die DFG. Die Regeln beinhalten auch Richtlinien für den Datenzugang. Primärdaten einer Veröffentlichung müssen mindestens 10 Jahre gespeichert und zugänglich sein, um eine Prüfung der Ergebnisse zu ermöglichen. Zwar werden diese Vorschriften im Regelfall eingehalten, aufgrund der damit verbundenen Zeitbelastung werden Daten aber normalerweise nur in Rohform archiviert und nicht in ihrer Feinstruktur aufgearbeitet, dokumentiert und allgemein verfügbar erschlossen.

Die wichtigsten Ziele eines neuen Umgangs mit Primärdaten sind also langfristige und allgemein zugängliche Speicherung. Durchsetzbar ist dies am besten über eine persönliche Motivation der Wissenschaftler. Dies ließe sich durch zwei Faktoren fördern:

Daten sollen nicht mehr ausschließlich Teil einer wissenschaftlichen Veröffentlichung sein, sondern eine eigenständige Identität besitzen. Damit würden Primärdaten, ähnlich wie Zeitschriftenartikel zitierbar.

Mit der anerkannten Datenpublikation erhält ein Autor also eine zitierfähige Veröffentlichung. Zeitschriftenartikel, welche die Daten verwenden, verweisen auf die Datenpublikation. Umgekehrt kann auch von den publizierten Daten auf Artikel in Zeitschriften verwiesen werden, die den Datensatz verwenden. Die Publikation von Daten kann also sinngemäß in das bestehende System von wissenschaftlichen Veröffentlichungen und deren Zitierbarkeit eingebunden werden.

## **Ansatz**

Der DFG-Ausschuss „*Wissenschaftliche Literaturversorgungs- und Informationssysteme*“ hat 2004 ein Projekt gestartet, um den Zugang zu wissenschaftlichen Primärdaten zu verbessern. Die *Technische Informationsbibliothek (TIB)* soll im Rahmen dieses Projektes als zentrale Registrierungsagentur für Primärdaten fungieren.

Beispielhaft für den Bereich der Geowissenschaften werden seit Ende 2004 von der TIB Primärdatensätze registriert. Die Datensätze selber verbleiben bei den lokalen Datenzentren. Die TIB vergibt für jeden Datensatz einen DOI als dauerhaften Identifizierer, wodurch der Datensatz weltweit über jeden Webbrowser auf einfachste Weise zugänglich wird. Zusätzlich werden für Sammlungen oder Auswertungen von Primärdatensätzen die Metadatenbeschreibungen in den Katalog der TIB aufgenommen. Diese Beschreibungen enthalten alle Angaben, die nach ISO 690-2 zur Zitierung elektronischer Medien verlangt werden.

Der Datensatz wird somit eine eigene zitierfähige Einheit. Die Anzeige eines Primärdatensatzes im Katalog der TIB sehen sie in der folgenden Abbildung: (Das Zitat zeigt noch die klassischen Feldbezeichner des Bibliotheksverbundes, die jedoch zukünftig an die neuen Publikationsformen angepasst werden.)

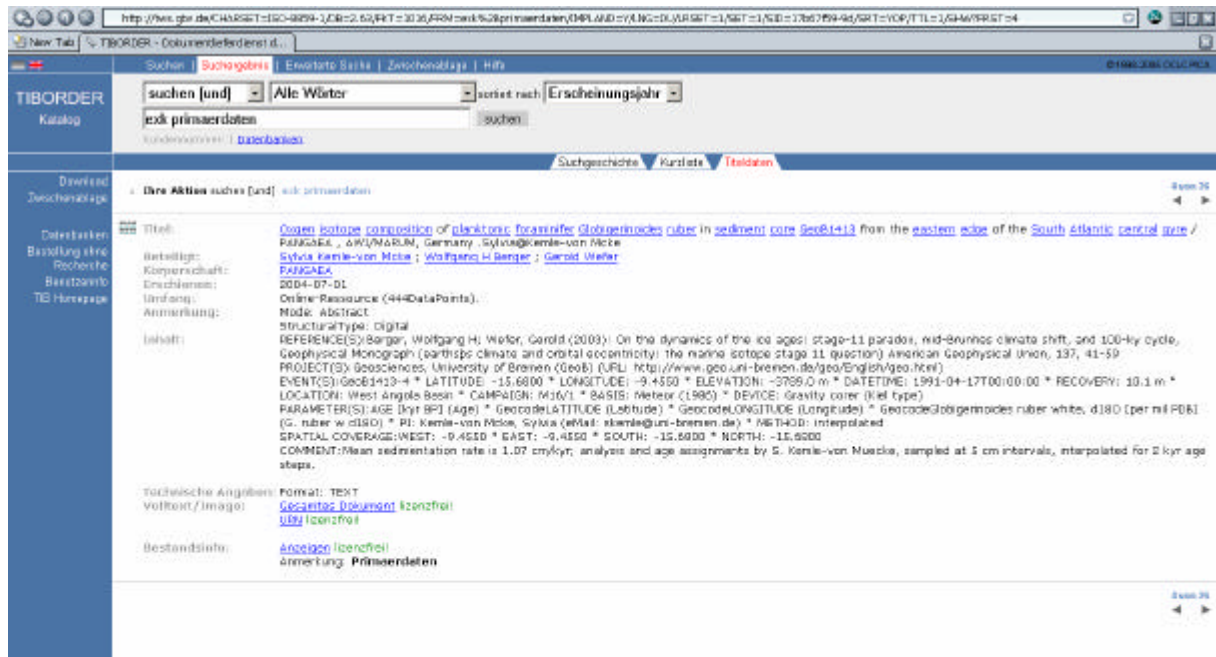


Abbildung: Anzeige eines Primärdatensatzes im Online-Katalog der TIB Hannover

## Identifizierer

Der Digital Object Identifier (DOI) wurde 1997 eingeführt, um Einheiten geistigen Eigentums in einer interoperativen digitalen Umgebung eindeutig zu identifizieren, zu beschreiben und zu verwalten. Verwaltet wird das DOI-System durch die 1998 gegründete International DOI Foundation (IDF, <http://www.doi.org/>).

Das IDF System besteht aus der „International DOI Foundation“ selbst, der eine Reihe von Registrationsagenturen ( „Registration Agencies “) zugeordnet sind . Für die Aufgaben einer RA können sich beliebige kommerzielle oder nicht kommerzielle Organisationen bewerben, die ein definiertes Interesse einer Gemeinschaft vorweisen können, digitale Objekte unter einem gemeinsamen Konzept zu verwalten.

Derzeit ist die TIB erweitertes Mitglied der IDF mit der Genehmigung DOIs für Primärdaten zu vergeben, Zum Stand 04/2005 wird die TIB volle Registrierungsagentur werden.

Durch eine Kooperation der TIB mit der *Deutschen Bibliothek* wird für jeden registrierten DOI für Primärdaten zusätzlich auch eine URN registriert. Die Strukturen von DOI und URN entsprechen sich, um eine höchstmögliche einfache Interoperabilität zu gewährleisten. Zum Auflösen der URNs wird die von der DDB zur Verfügung gestellte Infrastruktur verwendet.

## **Status**

Für den Bereich der Umwelt- und Geowissenschaften kooperiert die *TIB* mit

- World Data Center Climate (WDC Climate)
- World Data Center for Marine Environmental Sciences (WDC-MARE)
- GeoForschungsZentrum Potsdam
- Alfred-Wegener-Institut für Polar- und Meeresforschung

als weltweit erste zentrale Registrierungsagentur für Primärdaten

Zum Stand 02/2005 sind bereits über 100.000 Datensätze aus den beteiligten Datenzentren mit einem DOI versehen. 30 Datensätze sind im Katalog der *TIB* mit ihren Metadaten zitierfähig registriert und über das TIB-Portal zugänglich.