



Better practice für das Datenmanagement

Dr. Jörn Hoffmann
D-GEO Sekretariat
DLR Raumfahrt-Agentur

D-GEO Workshop
DLR
26.06.2013, Bonn

Wissenschaftliche Geodateninfrastrukturen – warum?



- Transparenz
- Nachnutzung (GEOSS!)
- Datenmanagement – auch während des Projekts
- Kooperatives Arbeiten (ICSU!)

Wissenschaftliche Geodateninfrastrukturen – warum nicht?

- „Zu viel Arbeit!“ (Metadaten!)
- Wie geht das? Was ist das?
- Was habe *ich* davon?
- „Meine Daten!“
 - Die kann nur ich richtig interpretieren
 - Andere könnten schneller publizieren

Nebenbestimmungen

Vorschlag

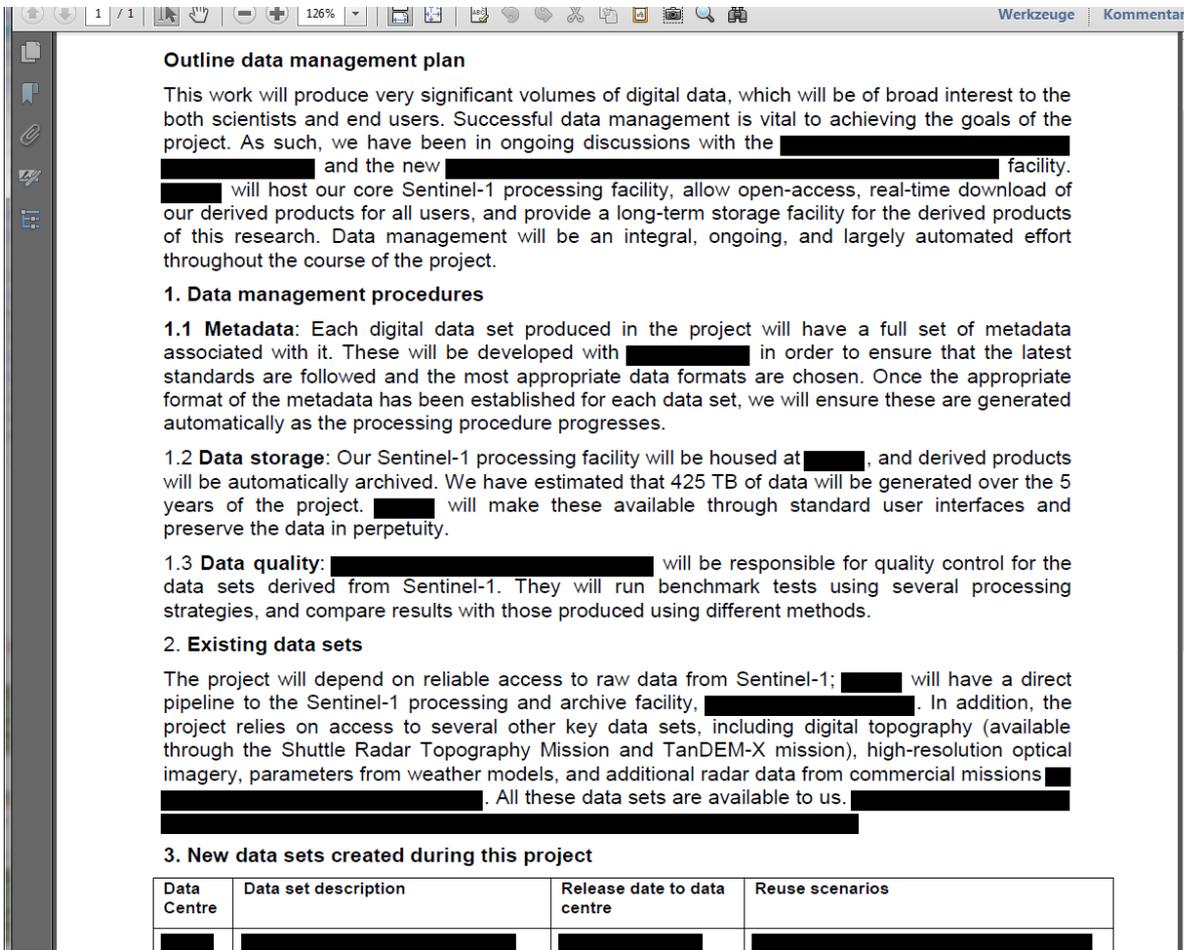
- Nebenbestimmungen im Förderbescheid

1. *Die Ergebnisse der Messungen bzw. der Proben-Analysen sind an relevante Projektdatenbanken zu übergeben.*
2. *Im Rahmen des Vorhabens erzeugte Geodaten, insbes. Erdbeobachtungsdaten, sind über das Portal der Group on Earth Observations („GEO Portal“) recherchierbar zu machen. Es wird empfohlen, dies durch einen Eintrag in geeignete Datenbanken oder dem Geodatenkatalog Deutschland (http://geoportal.bkg.bund.de/nn_127072/DE/Geodatensuche/Geodatensuche_node.html?_nnn=true) zu realisieren. Die Daten sollten über OGC-konforme Dienste bereitgestellt werden.*

Datenmanagementplan

Vorschlag

- Datenmanagementplan als Standard-Abschnitt im Antrag



Outline data management plan

This work will produce very significant volumes of digital data, which will be of broad interest to the both scientists and end users. Successful data management is vital to achieving the goals of the project. As such, we have been in ongoing discussions with the [REDACTED] and the new [REDACTED] facility. [REDACTED] will host our core Sentinel-1 processing facility, allow open-access, real-time download of our derived products for all users, and provide a long-term storage facility for the derived products of this research. Data management will be an integral, ongoing, and largely automated effort throughout the course of the project.

1. Data management procedures

1.1 Metadata: Each digital data set produced in the project will have a full set of metadata associated with it. These will be developed with [REDACTED] in order to ensure that the latest standards are followed and the most appropriate data formats are chosen. Once the appropriate format of the metadata has been established for each data set, we will ensure these are generated automatically as the processing procedure progresses.

1.2 Data storage: Our Sentinel-1 processing facility will be housed at [REDACTED], and derived products will be automatically archived. We have estimated that 425 TB of data will be generated over the 5 years of the project. [REDACTED] will make these available through standard user interfaces and preserve the data in perpetuity.

1.3 Data quality: [REDACTED] will be responsible for quality control for the data sets derived from Sentinel-1. They will run benchmark tests using several processing strategies, and compare results with those produced using different methods.

2. Existing data sets

The project will depend on reliable access to raw data from Sentinel-1; [REDACTED] will have a direct pipeline to the Sentinel-1 processing and archive facility, [REDACTED]. In addition, the project relies on access to several other key data sets, including digital topography (available through the Shuttle Radar Topography Mission and TanDEM-X mission), high-resolution optical imagery, parameters from weather models, and additional radar data from commercial missions [REDACTED]. All these data sets are available to us. [REDACTED]

3. New data sets created during this project

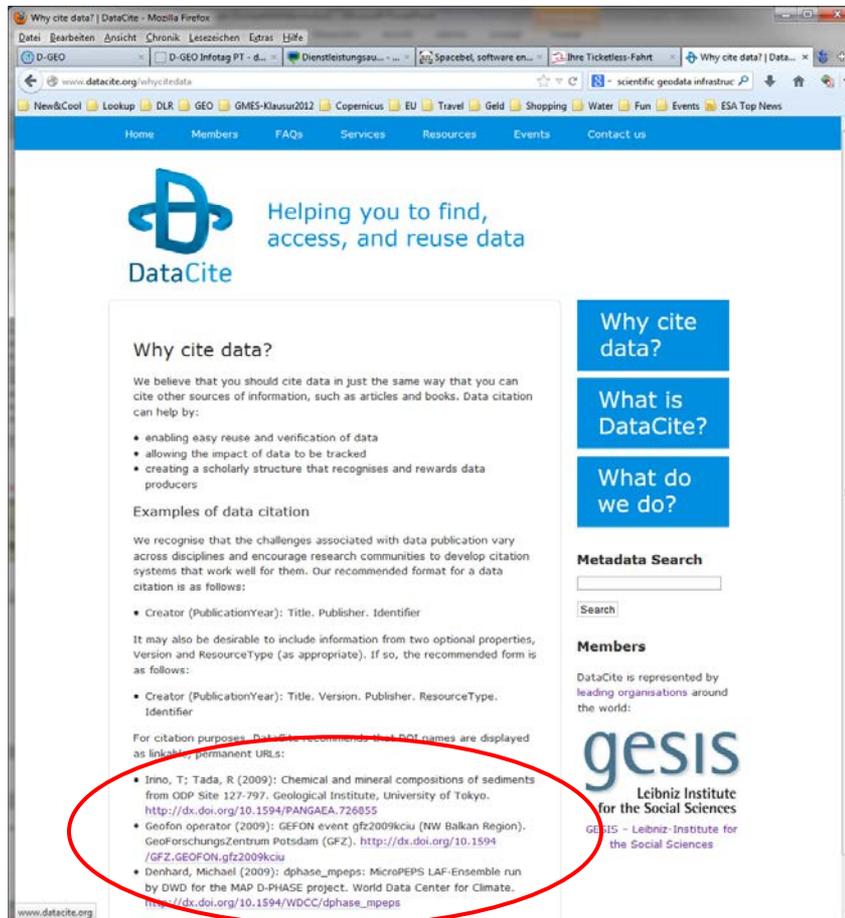
Data Centre	Data set description	Release date to data centre	Reuse scenarios
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]

- Welche Daten?
- Metadaten?
- Qualität?
- Speicherung
 - kurzfrist
 - langfrist
- Zugang?
- Nutzung?
- Publikation?

Datenreferenzen

Vorschlag

- Antrag sollte Referenzen zu publizierten Daten enthalten



Why cite data? | DataCite - Mozilla Firefox

www.datacite.org/whycitedata

Helping you to find, access, and reuse data

DataCite

Why cite data?

We believe that you should cite data in just the same way that you can cite other sources of information, such as articles and books. Data citation can help by:

- enabling easy reuse and verification of data
- allowing the impact of data to be tracked
- creating a scholarly structure that recognises and rewards data producers

Examples of data citation

We recognise that the challenges associated with data publication vary across disciplines and encourage research communities to develop citation systems that work well for them. Our recommended format for a data citation is as follows:

- Creator (PublicationYear): Title. Publisher. Identifier

It may also be desirable to include information from two optional properties, Version and ResourceType (as appropriate). If so, the recommended form is as follows:

- Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier

For citation purposes, DataCite recommends that DOI names are displayed as [linkable, permanent URLs](#):

- Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo. <http://dx.doi.org/10.1594/PANGAEA.726055>
- Geofon operator (2009): GEFON event gfz2009kciu (NW Balkan Region). GeoForschungsZentrum Potsdam (GFZ). <http://dx.doi.org/10.1594/GFZ.GEOFON.gfz2009kciu>
- Denhard, Michael (2009): dphase_mpeps: MicroPEPS LAF-Ensemble run by DWD for the MAP D-PHASE project. World Data Center for Climate. http://dx.doi.org/10.1594/WDC/dphase_mpeps

Why cite data?

What is DataCite?

What do we do?

Metadata Search

Members

DataCite is represented by leading organisations around the world:

gesis
Leibniz Institute for the Social Sciences
GESIS – Leibniz-Institute for the Social Sciences

www.datacite.org

Diskussion - Ansätze

- Nebenbestimmungen
- Datenmanagementplan
- Datenreferenzen
- ...



Vielen Dank!