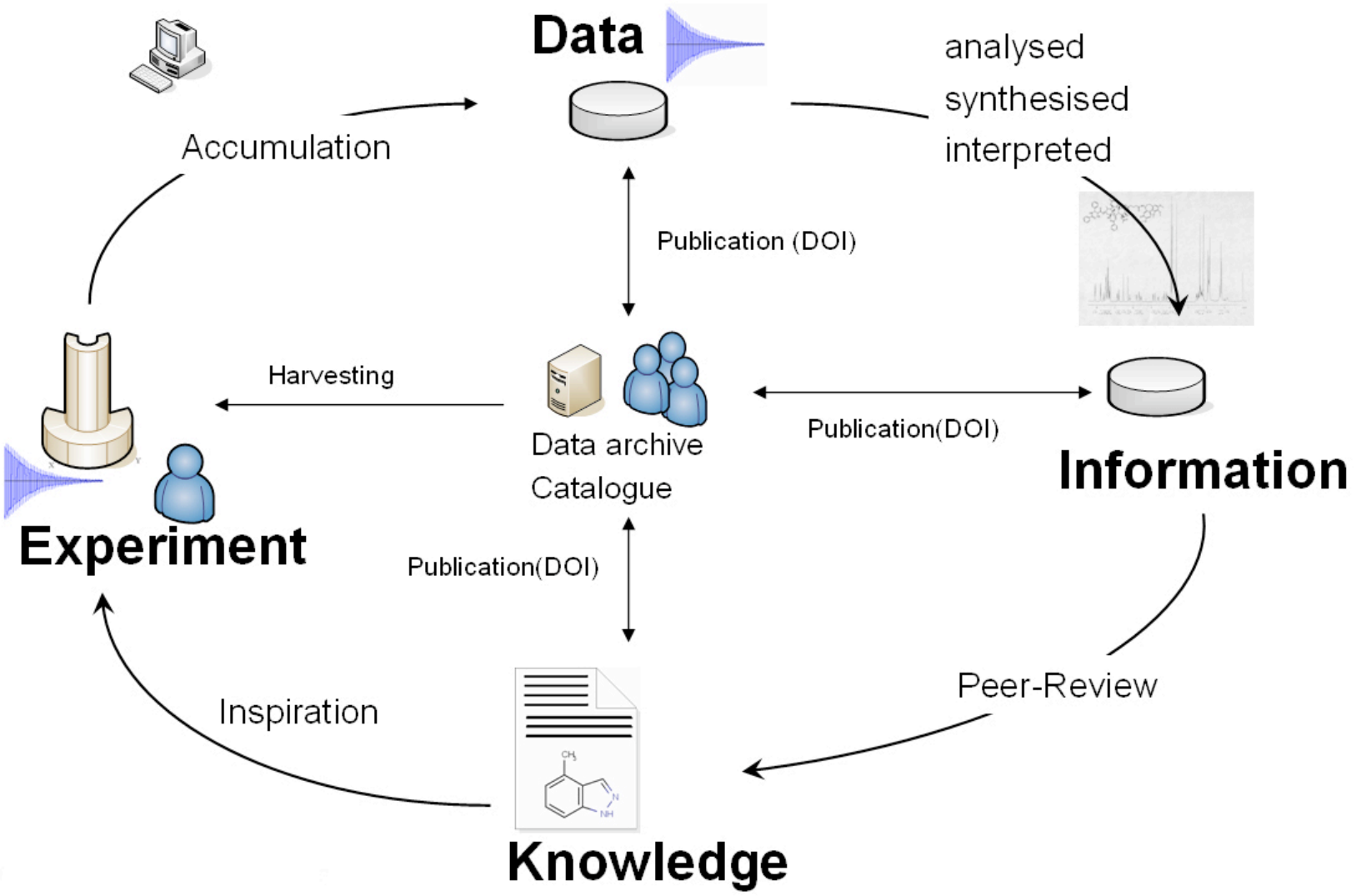
Publication and Citation of Scientific Primary Data at WDC Climate (WDCC)

Michael Lautenschlager (WDCC)
Heinke Höck (WDCC)
Jan Brase (TIB)
Susanne Waszkewitz (WDCC)

MPG eScience Seminar: Persistent Identifier

Garching, 27 + 28.03.08

The new Trajectory of Science



STD-DOI Project

- The *German Research Foundation* (DFG) has started the project *Publication and Citation of Scientific Primary Data* to increase the accessibility of scientific primary data, starting with the field of earth science.
- The *German National Library of Science and Technology* (TIB) is established as a “non-commercial” DOI-registration agency for scientific primary data as a member of the *International DOI Foundation* (IDF).
- The World Data Center Climate (WDCC) is awarding their data sets with a Digital Object Identifier (DOI) by the TIB

Partners in the Project

URL: www.std-doi.de

- World Data Center Climate (Hamburg)
- World Data Center Mare (Bremen)
- Alfred Wegener Institut (Bremerhaven)
- World Data Center RSAT (Oberpfaffenhofen)
- GeoForschungsZentrum Potsdam
- German National Library of Science and Technology (TIB)

Primary Data Publication (use case)

- During her research the scientist Mrs. Weather gains primary data about the temperature in Hannover in the year 2003.
- As usual the primary data is tested, evaluated, stored, documented and administrated at the

  World Data Center Climate (WDCC).

- In addition Mrs. Weather registers the primary data at the TIB.

(Data publication: data are quality proven, they are freely available and they are not any longer matter of change)

Registration of Primary Data (use case)

Mrs. Weather transmits to the TIB the URL where the data can be accessed, together with a XML-file containing all relevant bibliographic metadata

Including the information obligatory for citing of electronic media (ISO 690-2)

- author
- title
- size
- edition
- language
- publisher
- publishing date
- publishing place

(STD-DOI metadata profile)

Identifier (use case)

The TIB includes the metadata about the primary data in its catalogue, and awards the primary data with a unique identifier for registration: a DOI (and an URN) name

In her publications, Mrs. Weather is now citing this primary data together with its unique DOI name:

Weather, Hermione (2003): Temperature in Hannover for 2003. WDC Climate [doi: 10.1594/WDCC/W_Han_2003_MMB_2]

Data and Article

The DOI system offers an easy way to connect the article with the underlying data:

The dataset:

G.Yancheva, . R Nowaczyk et al (2007)

Rock magnetism and X-ray flourescence spectrometry analyses on sediment cores of the Lake Huguang Maar, Southeast China, PANGAEA

doi:10.1594/PANGAEA.587840

Is cited in the article:

G. Ycheva, N. R. Nowaczyk et al (2007)

Influence of the intertropical convergence zone on the East Asian monsoon

Nature 445, 74-77

doi:10.1038/nature05431

Status

So far 460 TIBORDER catalogue entries for scientific data have been registered at TIB and assigned a DOI / URN

The storage, maintainence and evaluation of the contents lies with the sustainable content providers (institutions, data centers, …) as so-called publication agents.

The system is always open for new partners also from other disciplines!

Please contact TIB: Jan.Brase@tib.uni-hannover.de

The STD-DOI scientific data publication concept fits into the rules of good scientific practice.

RULES OF GOOD SCIENTIFIC PRACTICE

- adopted by the Senate of the Max Planck Society on 24 November 2000 -

"Primary data as a basis for publications must, as far as possible, be stored for at least ten years on durable, secure carriers in the institutes or research establishments in which they arose. Access to the data has to be granted for persons with a justifiable interest.

Scientific examinations, experiments and numerical calculations can only be reproduced or reconstructed if all the important steps are comprehensible. For this reason, full and adequate reports are necessary, and these reports must be kept for a minimum period of ten years."

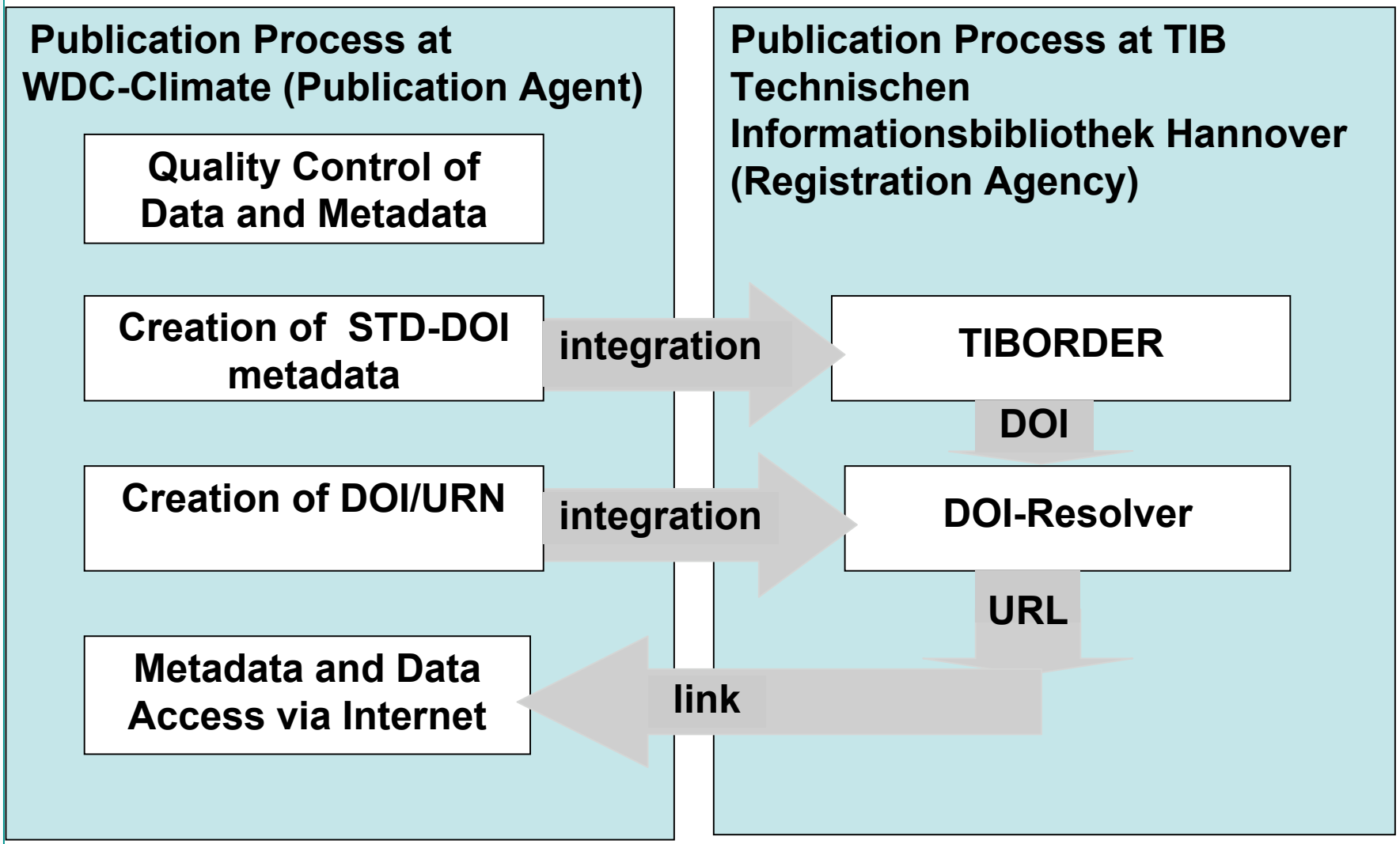
http://www.maxplanck.de/pdf/rulesScientificPract.pdf

Data Management Aspects related to the Rules of good Scientific Practice

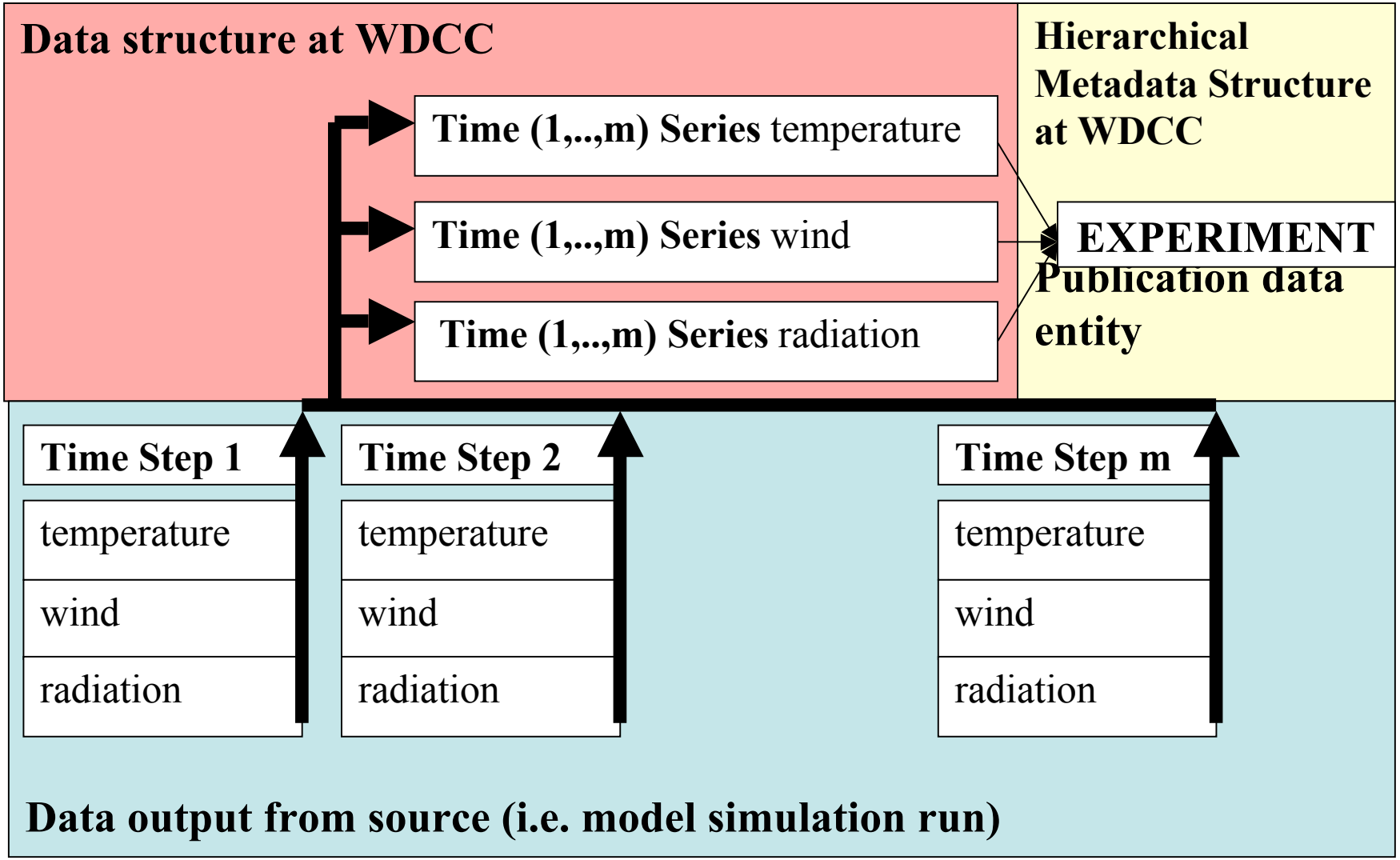
- Scientific data quality
- Metadata quality (reports …)
- Searchable data catalogues
- Persistent access to scientific data
- Persistent access to metadata
- Scientific data can be reproduced or reconstructed (if possible)
- Contact to data provider
- Citation convention
- Bit stream preservation

Publication of Scientific Primary Data at WDCC

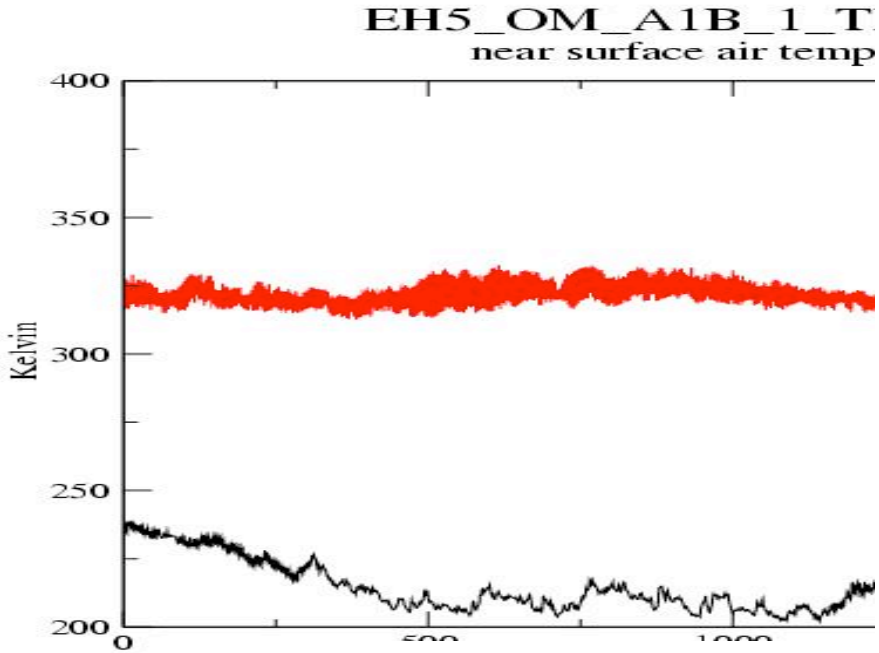
Precondition:
long term availability of Data and Metadata at WDC-Climate

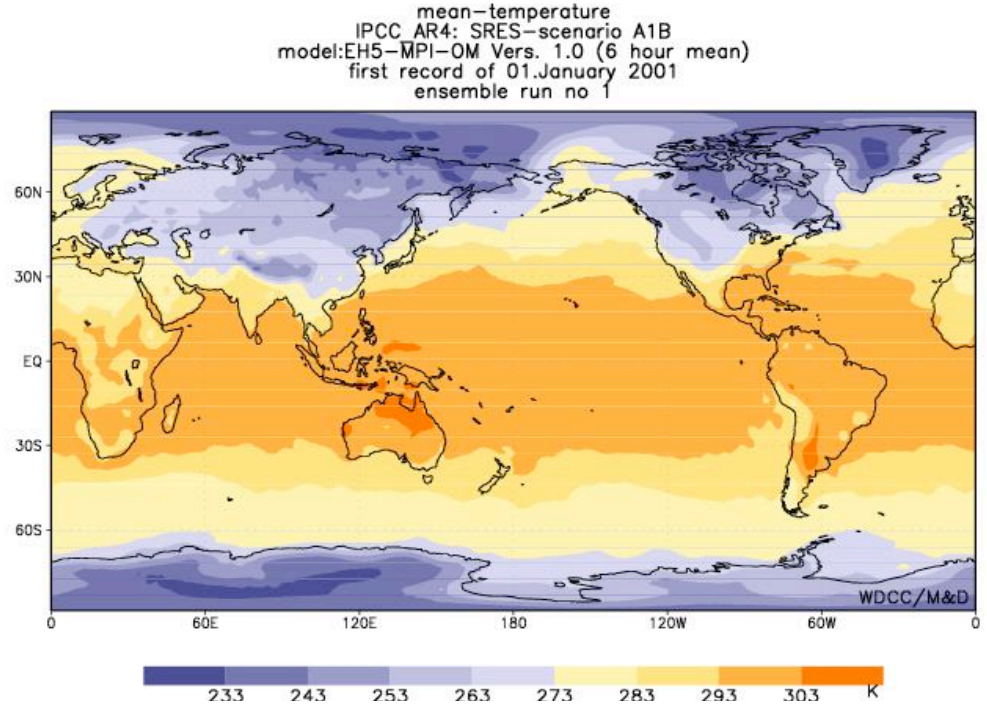


Granularity of Primary Data at WDCC



Syntactic Quality Control of Primary Data and Metadata





- Scientific data and metadata review together with data providers (Flag: “approved by author”)
- Completeness and consistency of scientific data + metadata
  - Time series min/max, statistics and samples of records
  - Permission
  - Size of primary data

(Semantic: scientific publications)

ne Wasz

Integration in TIBORDER and DOI/URN Resolver

After review process completion („approved by author“):

- Assignment of STD-DOI metadata (ISO 690-2) + persistent identifier (DOI/URN)
- Dissemination to TIB for inclusion in TIBORDER + DOI/URN Resolver

- Published data are part of the WDC Climate which has 340 Terabytes web accessible in its DB system.

  Data are split in $7 * 10^9$ individual table entries (BLOBS) which results in typical BLOB sizes of

  64 kBytes – 256 kBytes.

  48 Terabytes are published in 64 TIBORDER entries by WDCC

STD-DOI metadata

Atribute (example) *comment*	LOM (DC) mapping
1. DOI (10.1594/WDCC/EH5-T63L31_OM-GR1.5L40_A1B_1_6H)	dc:identifier
2. identifier (**Urn:tib:10.1594/WDCC/EH5-T63L31_OM-GR1.5L40_A1B_1_6H**)	dc:identifier
3. creator (Erich Roeckner, Michael Lautenschlager, Heiko Schneider)	dc:creator
4. Contributor *institution collected or created data apart from publisher*	dc:contributor
5. publisher (World Data Center for Climate (WDCC))	dc:publisher
6. title (IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 SRESA1B run no.1: atmosphere 6 HOUR values MPImet/MaD Germany)	dc:title
7. language (en)	dc:language
8. structuralType (digital)	dc:format

STD-DOI metadata

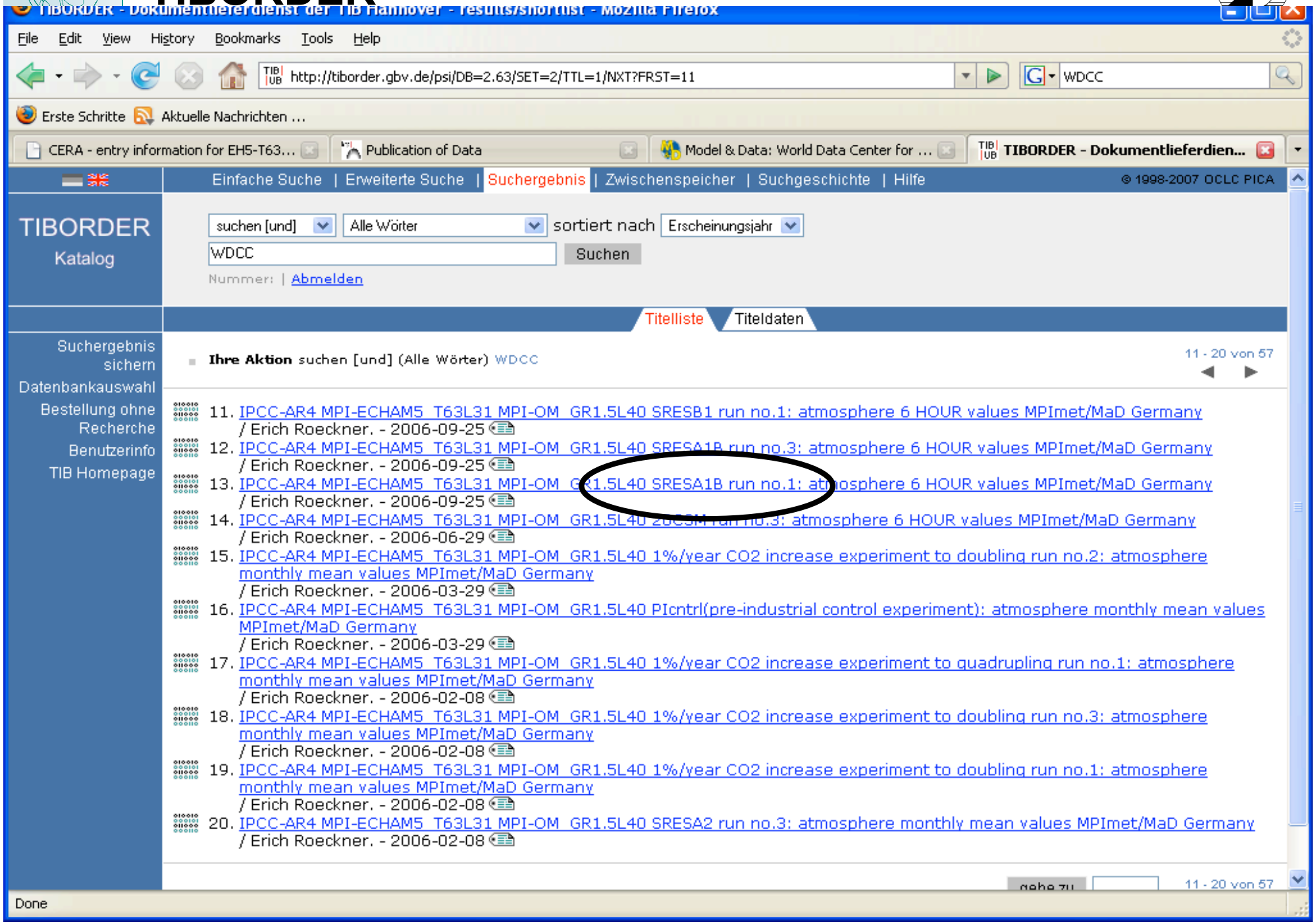
9. mode (abstract)	none
10. resourceType (dataset)	dc:format
11. registrationAgency (10.1594)	dc:contributor
12. issueDate ()	dcterms:issued
13. IssueNumber (1)	dc:identifier
14. creationDate (2004-07-12)	dc:source
15. publicationDate (2006-09-25)	dc:date
16. startPublicationDate ()	none
17. description (The data represent 6 hourly values of selected variables for the Data Distribution Centre(DDC) of the Intergovermental Panel on Climate Change (IPCC). (see also: http://ipcc-ddc.cru.uea.ac.uk) The A1B scenario is the part of the A1 family which describes a balance across all energy sources. The experiment has been initialized in year 2000 of the 20C_1 run and continues until year 2100 with anthropogenic forcings (CO2, CH4, N2O, CFCs, O3 and sulfate) *etc*)	dc:description

STD-DOI metadata

18. publicationPlace (Hamburg)	none
19. size (3308811833628 Bytes = 3.3 TB)	dcterms:extend
20. format (Grib)	dc:format
21. edition (1)	none
22. relatedDOIs ()	dc:source (and others)
23. relatedDOIs () *with type definition*	dc:source (and others)
22. discipline (earthScience)	dc:source (and others)

(STD-DOI metadata = DOI metadata + ISO 690-2)

TIBORDER



TIBORDER - Dokumentlieferdienst der TIB Hannover - results/shortlist - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://tiborder.gbv.de/psi/DB=2.63/SET=2/TTL=1/NXT?FRST=11

WDCC

Erste Schritte Aktuelle Nachrichten ...

CERA - entry information for EH5-T63... | Publication of Data | Model & Data: World Data Center for ... | TIBORDER - Dokumentlieferdien...

Einfache Suche | Erweiterte Suche | Suchergebnis | Zwischenspeicher | Suchgeschichte | Hilfe © 1998-2007 OCLC PICA

TIBORDER
Katalog

suchen [und] Alle Wörter sortiert nach Erscheinungsjahr

WDCC Suchen

Nummer: | Abmelden

Titelliste Titeldaten

Suchergebnis sichern
Datenbankauswahl
Bestellung ohne Recherche
Benutzerinfo
TIB Homepage

Ihre Aktion suchen [und] (Alle Wörter) WDCC

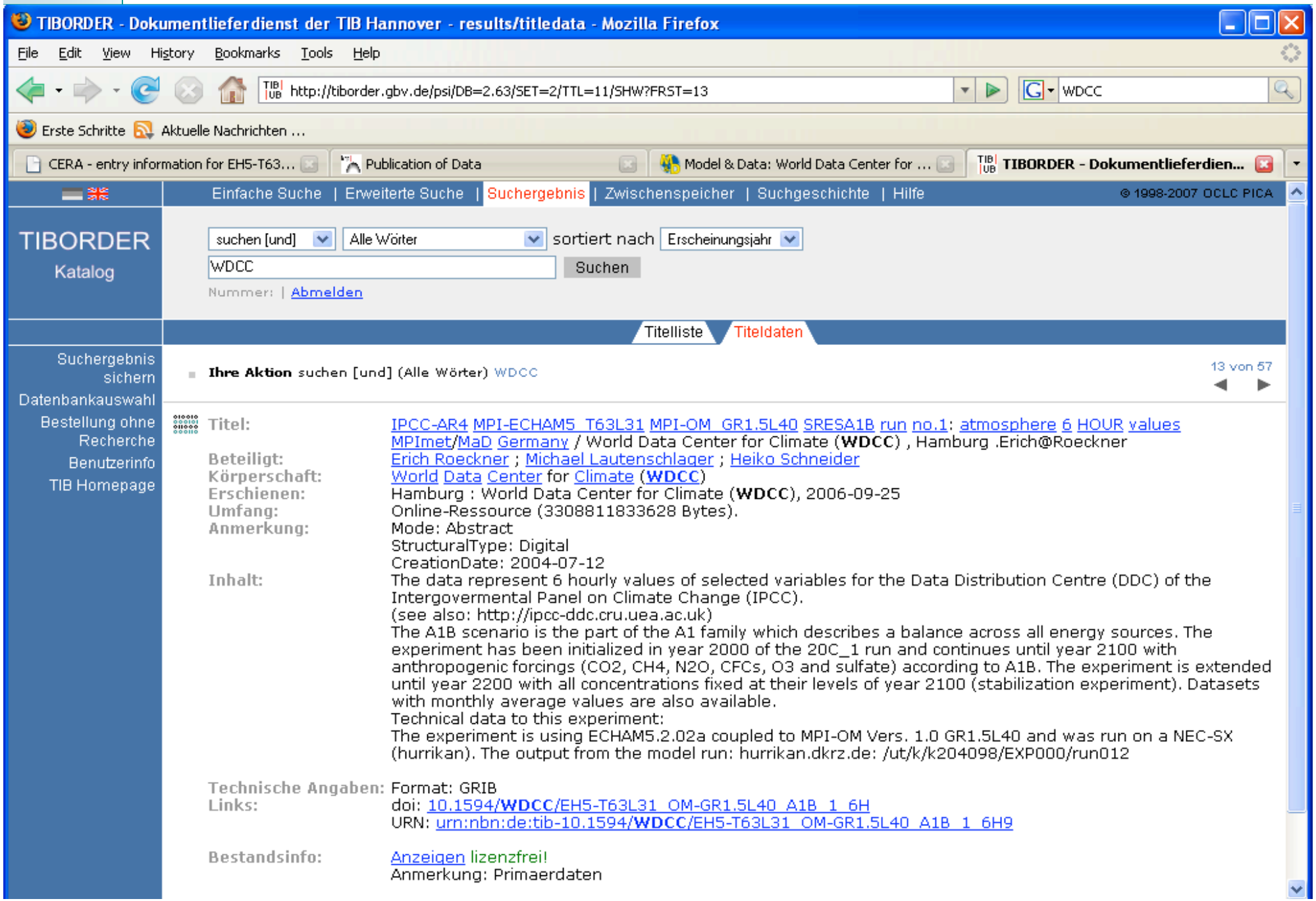
11 - 20 von 57

11. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 SRESB1 run no.1: atmosphere 6 HOUR values MPImet/MaD Germany
/ Erich Roeckner. - 2006-09-25
12. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 SRESA1B run no.3: atmosphere 6 HOUR values MPImet/MaD Germany
/ Erich Roeckner. - 2006-09-25
13. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 SRESA1B run no.1: atmosphere 6 HOUR values MPImet/MaD Germany
/ Erich Roeckner. - 2006-09-25
14. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 20C3M run no.3: atmosphere 6 HOUR values MPImet/MaD Germany
/ Erich Roeckner. - 2006-06-29
15. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 1%/year CO2 increase experiment to doubling run no.2: atmosphere monthly mean values MPImet/MaD Germany
/ Erich Roeckner. - 2006-03-29
16. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 PIcntrl(pre-industrial control experiment): atmosphere monthly mean values MPImet/MaD Germany
/ Erich Roeckner. - 2006-03-29
17. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 1%/year CO2 increase experiment to quadrupling run no.1: atmosphere monthly mean values MPImet/MaD Germany
/ Erich Roeckner. - 2006-02-08
18. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 1%/year CO2 increase experiment to doubling run no.3: atmosphere monthly mean values MPImet/MaD Germany
/ Erich Roeckner. - 2006-02-08
19. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 1%/year CO2 increase experiment to doubling run no.1: atmosphere monthly mean values MPImet/MaD Germany
/ Erich Roeckner. - 2006-02-08
20. IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 SRESA2 run no.3: atmosphere monthly mean values MPImet/MaD Germany
/ Erich Roeckner. - 2006-02-08

gehe zu 11 - 20 von 57

Done

TIBORDER


TIBORDER - Dokumentlieferdienst der TIB Hannover - results/titledata - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://tiborder.gbv.de/psi/DB=2.63/SET=2/TTL=11/SHW?FRST=13
WDCC
Erste Schritte
Aktuelle Nachrichten ...
CERA - entry information for EH5-T63...
Publication of Data
Model & Data: World Data Center for ...
TIBORDER - Dokumentlieferdien...
Einfache Suche | Erweiterte Suche | Suchergebnis | Zwischenspeicher | Suchgeschichte | Hilfe
© 1998-2007 OCLC PICA
TIBORDER
Katalog
suchen [und]
Alle Wörter
sortiert nach
Erscheinungsjahr
WDCC
Suchen
Nummer: | Abmelden
Titelliste
Titeldaten
Suchergebnis sichern
Datenbankauswahl
Bestellung ohne Recherche
Benutzerinfo
TIB Homepage
Ihre Aktion suchen [und] (Alle Wörter) WDCC
13 von 57
Titel: IPCC-AR4 MPI-ECHAM5_T63L31 MPI-OM_GR1.5L40 SRESA1B run no.1: atmosphere 6 HOUR values MPImet/MaD Germany / World Data Center for Climate (WDCC) , Hamburg .Erich@Roeckner
Beteiligt: Erich Roeckner ; Michael Lautenschlager ; Heiko Schneider
Körperschaft: World Data Center for Climate (WDCC)
Erschienen: Hamburg : World Data Center for Climate (WDCC), 2006-09-25
Umfang: Online-Ressource (3308811833628 Bytes).
Anmerkung: Mode: Abstract
StructuralType: Digital
CreationDate: 2004-07-12
Inhalt: The data represent 6 hourly values of selected variables for the Data Distribution Centre (DDC) of the Intergovermental Panel on Climate Change (IPCC).
(see also: http://ipcc-ddc.cru.uea.ac.uk)
The A1B scenario is the part of the A1 family which describes a balance across all energy sources. The experiment has been initialized in year 2000 of the 20C_1 run and continues until year 2100 with anthropogenic forcings (CO2, CH4, N2O, CFCs, O3 and sulfate) according to A1B. The experiment is extended until year 2200 with all concentrations fixed at their levels of year 2100 (stabilization experiment). Datasets with monthly average values are also available.
Technical data to this experiment:
The experiment is using ECHAM5.2.02a coupled to MPI-OM Vers. 1.0 GR1.5L40 and was run on a NEC-SX (hurrikan). The output from the model run: hurrikan.dkrz.de: /ut/k/k204098/EXP000/run012
Technische Angaben: Format: GRIB
Links: doi: 10.1594/WDCC/EH5-T63L31_OM-GR1.5L40_A1B_1_6H
URN: urn:nbn:de:tib-10.1594/WDCC/EH5-T63L31_OM-GR1.5L40_A1B_1_6H9
Bestandsinfo: Anzeigen lizenzfrei!
Anmerkung: Primaerdaten

Resolving the DOI

This DOI name can be resolved (and the data can be accessed) in every browser worldwide in different ways:

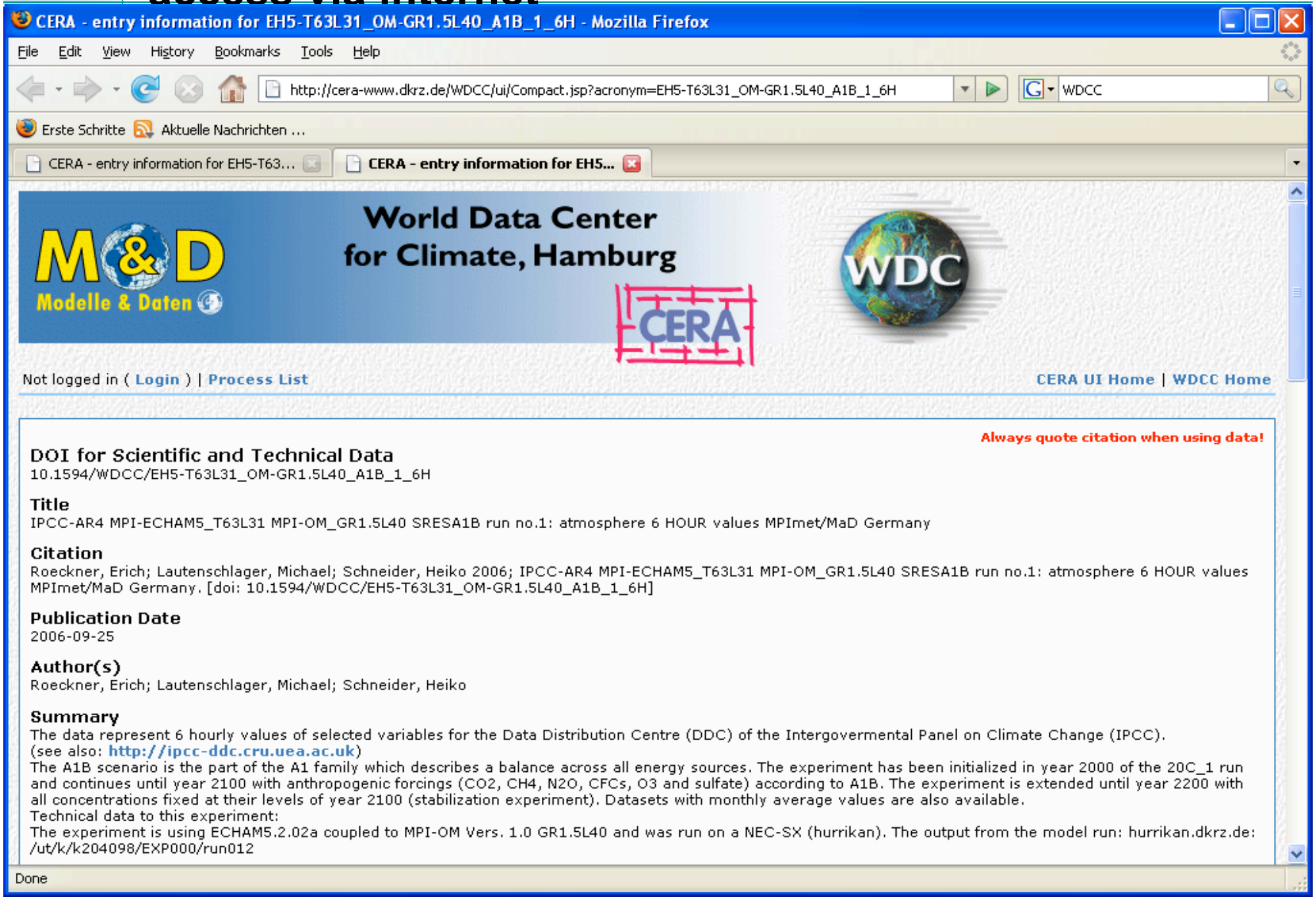
- http://dx.doi.org/10.1594/WDCC/W_Han_2003_MMB_2
- http://doi.tib-hannover.de:8000/10.1594/WDCC/W_Han_2003_MMB_2

Or by

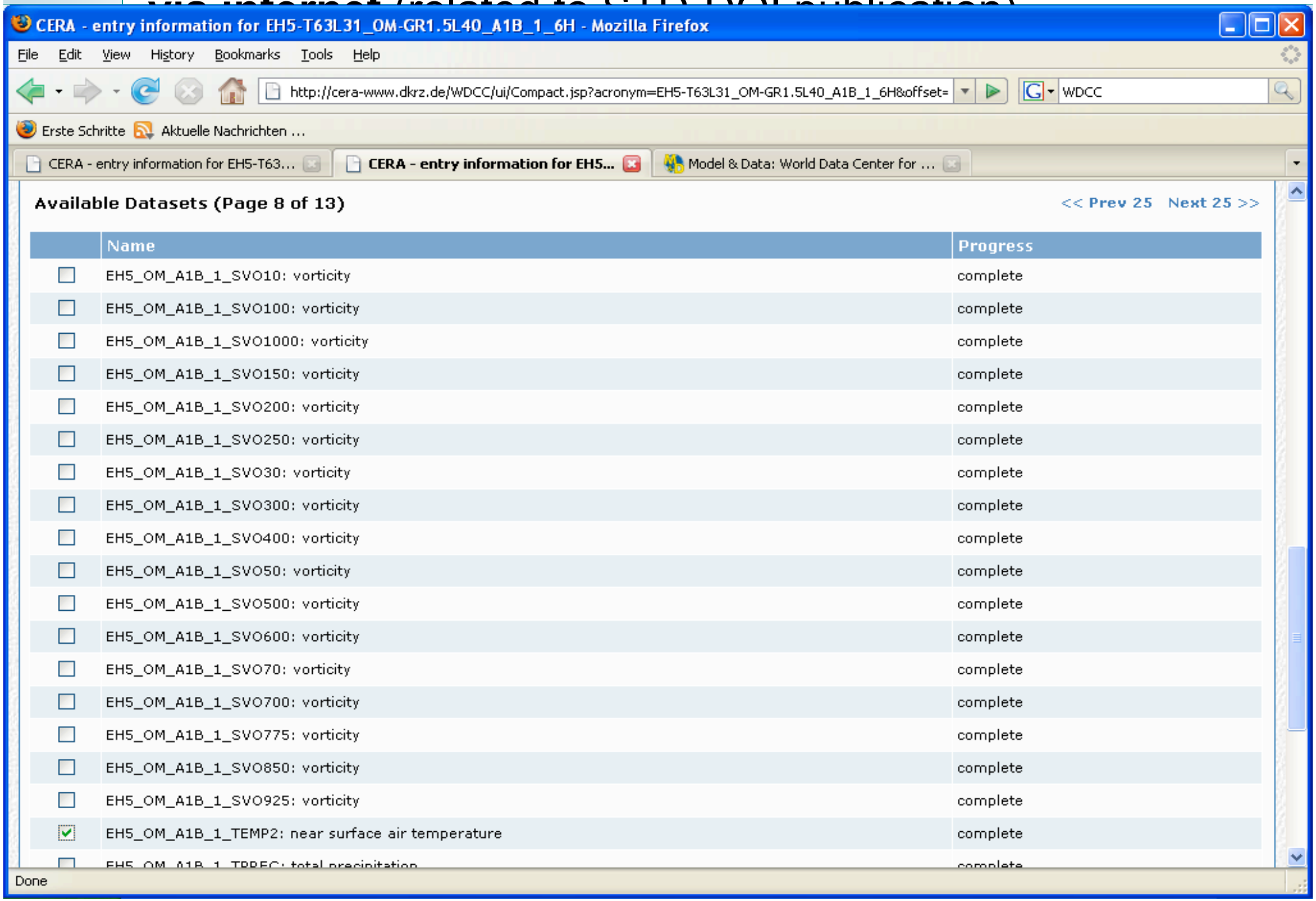
Doi://10.1594/WDCC/W_Han_2003_MMB_2

(after installing a browser plugin)

Example of WDCC Metadata (Compact View) access via internet



Example of Primary Data (Compact View) access via internet (related to STD-DOI publication)



Improvement of Scientific Work by the STD-DOI Publication Process

- Scientific data quality
  (quality flag “approved by author”)
- Metadata quality (quality control procedure)
- Searchable data catalogue
  (via WDCC GUI or TIBORDER)
- Persistent access to primary data and metadata
  (with DOI/URN via Compact View and WDCC GUI)
- Scientific data can be reproduced and reconstructed
  (link to metadata and contact to data provider)
- Contact to data provider (part of compact view)
- Citation convention for third party data usage which relates author and data entity
  (part of compact view, DOI)