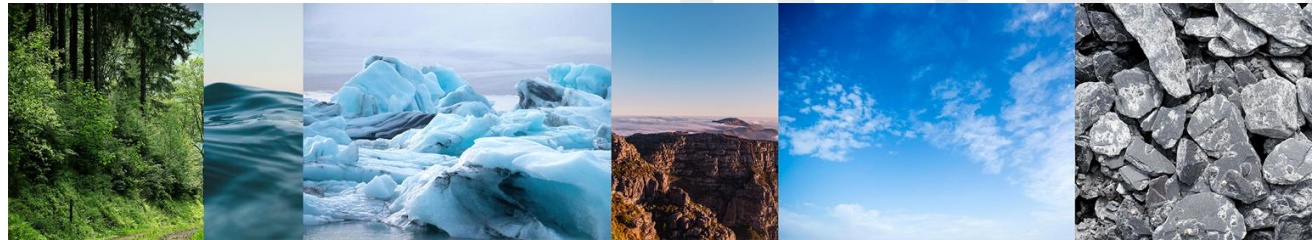


Claudia Martens [DKRZ]  
Aenne Löhden [DKRZ]  
Markus Stocker [TIB]

# Benefits of Terminologies for Interdisciplinary Research



BluePrints for the Integration of Terminology  
Services in Earth System Sciences



# What is it about?

Once upon a time at a Data Centre ...

Panel 1: A woman in a purple dress asks, "Do you have data for me?"

Panel 2: A woman in a red dress replies, "Data on declining rainfall, please."

Panel 3: The woman in the red dress says, "Alright. We have data on ...". The woman in the purple dress yells, "Ouch!". A large, messy CSV file is shown with columns like `station_id, rainfall, amount, convective_rainfall, time, convective_rainfall, time, rainfall, time, convective_rainfall, time, convective_rainfall, amount, station_id, rainfall, time, station_id, rainfall, time, station_id, convective_rainfall, amount, station_id, convective_rainfall, amount, station_id, convective_rainfall, amount, station_id, convective_rainfall, amount`.

Panel 4: The woman in the red dress asks, "Well, just choose the most proper variable". The woman in the purple dress asks, "What is convective rain?".

Panel 5: The woman in the red dress asks, "But what the heck do all those variables mean?". The woman in the purple dress asks, "What is convective rain?".

Panel 6: The woman in the red dress asks, "You should also think of the geo-coordinates you need!". The woman in the purple dress asks, "Madeira. It's just Madeira!".

Panel 7: The woman in the red dress asks, "Madeira. It's just Madeira!". The woman in the purple dress asks, "Madeira. It's just Madeira!".

# Introduction

What is it about?

A social scientist wants to investigate the correlation of climate change and the increase of social conflicts in peasant subsistence farming, using Levada irrigation systems on Madeira as an example.



Once upon a time at a Data Centre ...



# Introduction

What is it about?



# Introduction

What is it about?



# Introduction

What is it about?



# Introduction

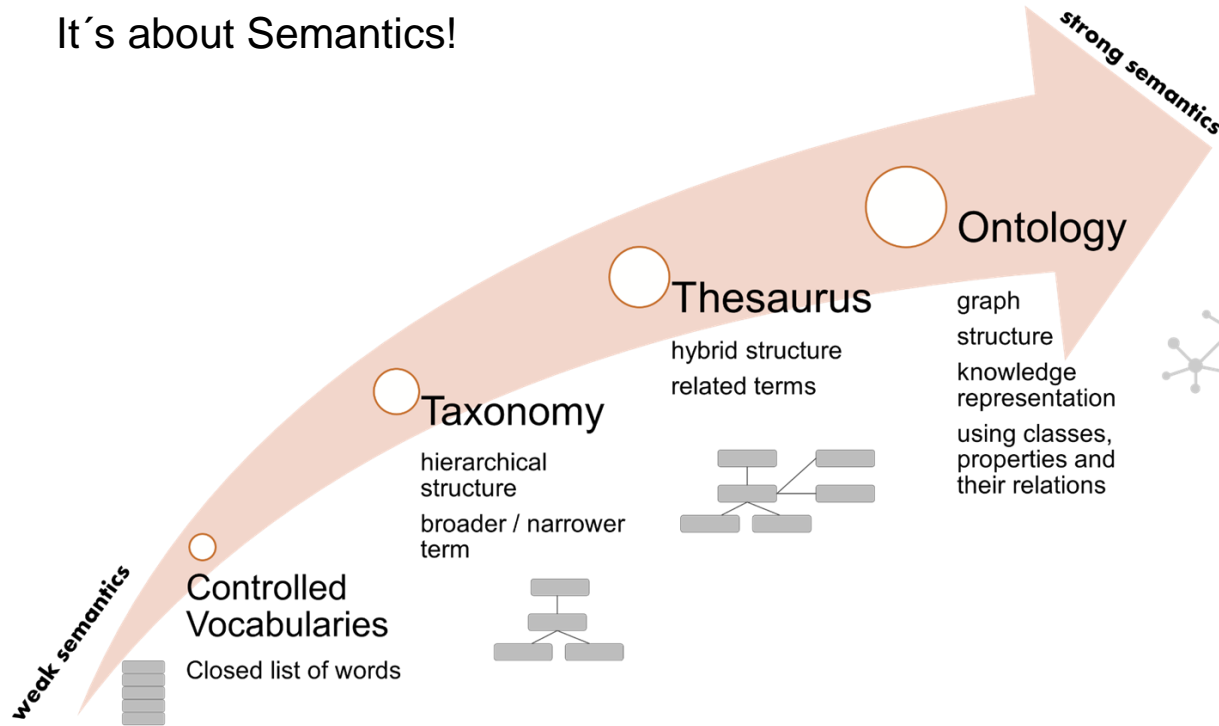
What is it about?

**Terminologies  
may help!**



# Introduction

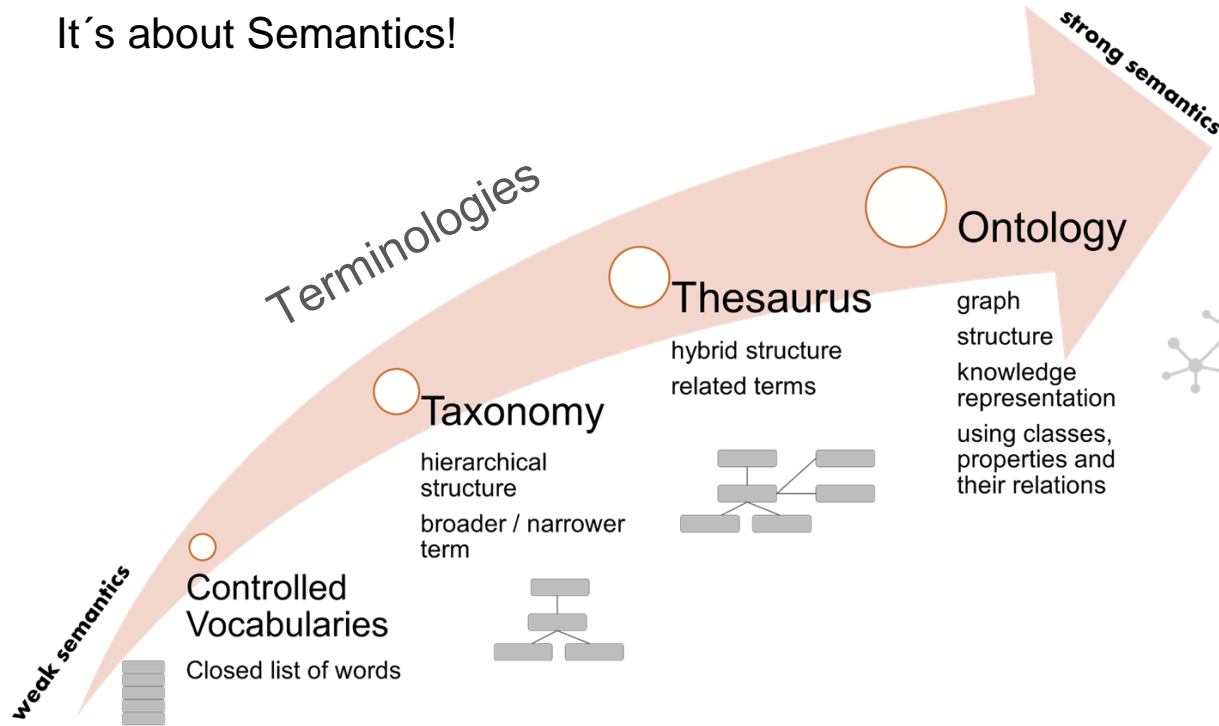
It's about Semantics!





# Introduction

It's about Semantics!



**Blueprint for the  
Integration of  
Terminology  
Services in ESS**

# The BITS Project

Terminologies in Earth System Sciences



BluePrints for the Integration of Terminology  
Services in Earth System Sciences



# Background

Data in Earth System Science - divergent, complex, huge!



# Background

Data in Earth System Science - divergent, complex, huge!

- **different data types**  
model output, observations, simulations, measurements
- **different formats**  
NetCDF (HDF5), GeoTIFF, GRIB, GeoJSON, ASCII/CSV or frameworks like GIS
- **different sizes / granularity level**  
from PB in radio astronomy and climate model output to MB in samples in Biology

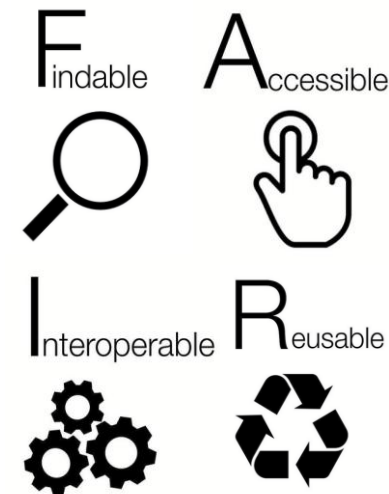




# From Chaos to FAIR

Using metadata standards and terminologies to make data FAIR

|                          |   |
|--------------------------|---|
| Atmospheric Science      | CF Conventions, ACDD, ISO 19115                                   |
| Oceanography             | SeaDataNet/SeaDataCloud, NODC/NCEI Profile, CF Conventions        |
| Hydrology                | OGC WaterML 2.0, WMO WIS, ISO 19156 (Observations & Measurements) |
| Geology / Geophysics     | CF Conventions, ISO 19115, ACDD                                   |
| Ecology & Biodiversity   | Darwin Core, EML (Ecological Metadata Language)                   |
| Remote Sensing / GIS     | - ISO 19115 / ISO 19139<br>- FGDC CSDGM                           |
| Environmental Monitoring | SensorML (OGC), INSPIRE (= ISO)                                   |



# Semantic Interoperability

## in Environmental Sciences

**Shared Meaning Across Systems:** Semantic interoperability ensures that data exchanged between different environmental systems (e.g., climate models, biodiversity databases) is understood consistently by using standardized vocabularies, ontologies, and metadata schemas.

**Use of Controlled Vocabularies and Ontologies:** It relies on community-agreed terms and structures (e.g., CF Standard Names, GEMET, ENVO) to describe data, enabling accurate interpretation, integration, and automated reasoning across disciplines.

**Supports FAIR and Cross-Domain Research:** Semantic interoperability enhances data reusability and integration across domains such as hydrology, ecology, and atmospheric science—supporting the FAIR principles and enabling interdisciplinary environmental research.



# Semantic Interoperability

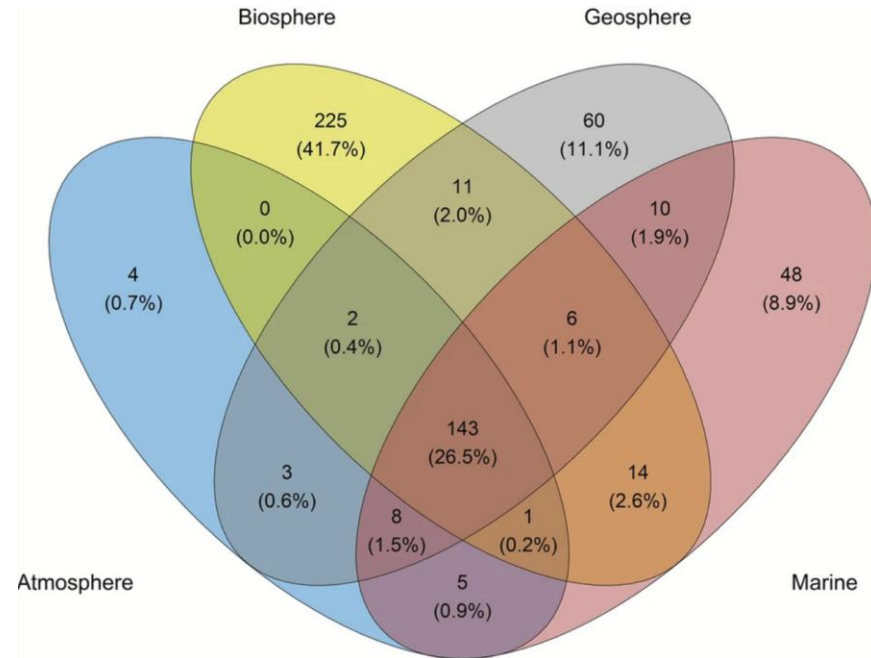
## in Environmental Sciences

**Shared Meaning Across Systems:** Semantic interoperability ensures that data exchanged between different environmental systems (e.g., climate models, biodiversity databases) is understood consistently by using standardized vocabularies, ontologies, and metadata schemas.

**Use of Controlled Vocabularies and Ontologies:** It relies on community-agreed terms and structures (e.g., CF Standard Names, GEMET, ENVO) to describe data, enabling accurate interpretation, integration, and automated reasoning across disciplines.

**Supports FAIR and Cross-Domain Research:** Semantic interoperability enhances data reusability and integration across domains such as hydrology, ecology, and atmospheric science—supporting the FAIR principles and enabling interdisciplinary environmental research.

Di Muri, C., Pulieri, M., Raho, D. *et al.* Assessing semantic interoperability in environmental sciences: variety of approaches and semantic artefacts. *Sci Data* **11**, 1055 (2024). <https://doi.org/10.1038/s41597-024-03669-3>



# Semantic Interoperability

in Environmental Sciences - burdens

Di Muri, C., Pulieri, M., Raho, D. *et al.* Assessing semantic interoperability in environmental sciences: variety of approaches and semantic artefacts. *Sci Data* **11**, 1055 (2024). <https://doi.org/10.1038/s41597-024-03669-3>

## Inconsistent Availability in Catalogues

Terminologies not listed or registered in semantic catalogues

## Non-Standardized Formats and Languages

hampers machine readability

## Absence of proper Licenses

## Lack of Version Information

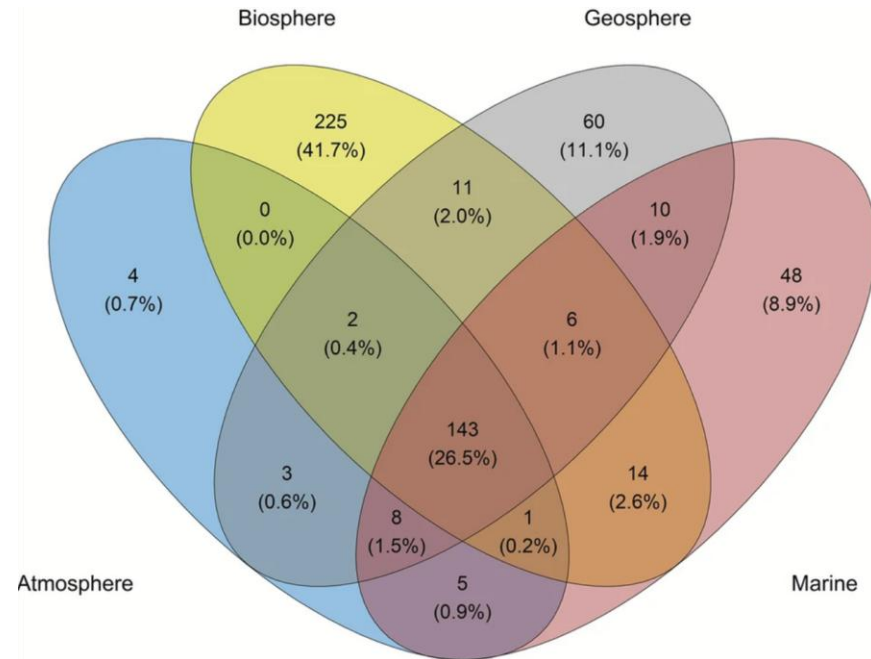
## No persistent identification

of Terminologies and Terms

→ difficult for long term accessibility

## Inadequate Description

## Ambiguous Maintenance Status





# BITS Project

Terminologies for Earth System Sciences



already existing Terminology  
Service offered by TIB

[Terminology Service](#)



# BITS Project

Terminologies for Earth System Sciences



+ SKOS

based on Ontology  
Lookup Service  
(EMBL-EBI) with a  
SKOS extension



already existing Terminology  
Service offered by TIB

[Terminology Service](#)



# BITS Project

Terminologies for Earth System Sciences



+ SKOS

based on Ontology  
Lookup Service  
(EMBL-EBI) with a  
SKOS extension



[Terminology Service](#)



maintained by ESS Community



# BITS Project

Terminologies for Earth System Sciences



+ SKOS

based on Ontology  
Lookup Service  
(EMBL-EBI) with a  
SKOS extension



## Terminology Service

### requirements

- license
- format/language

### features

- version control
- PIDs for terms
- long term maintenance



maintained by ESS Community

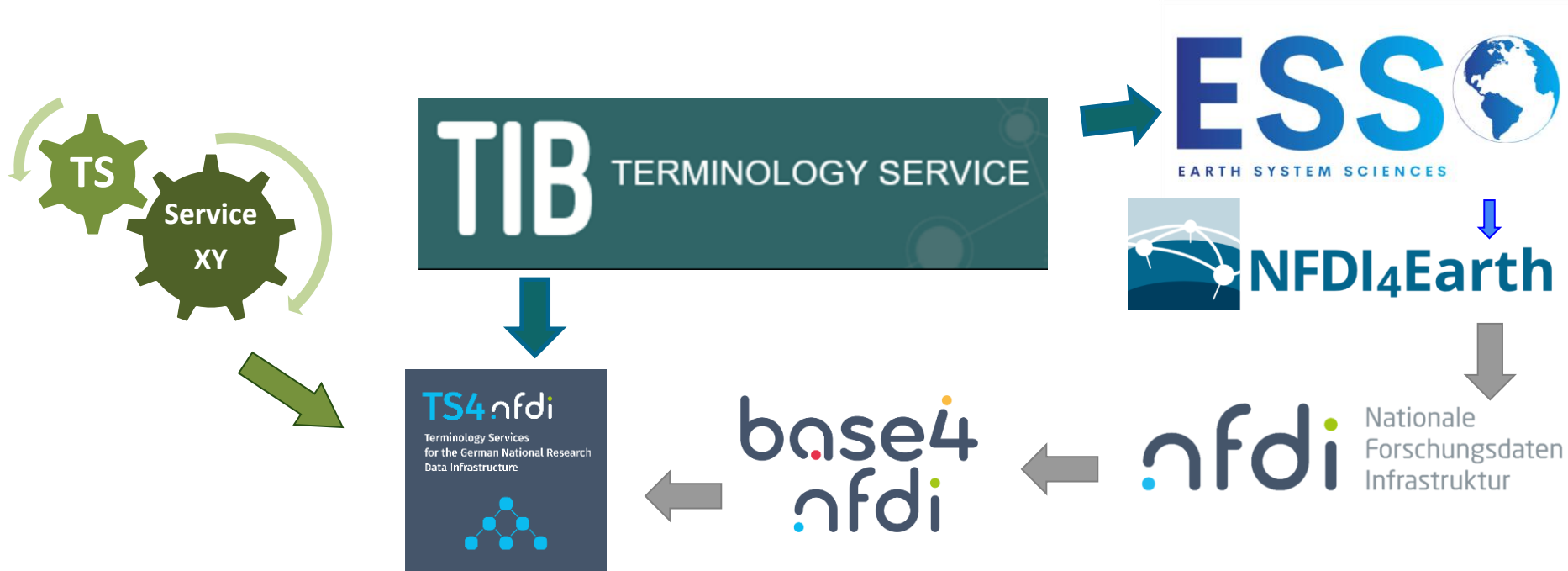
# From Project to Infrastructure

Sustainable service via NFDI



# From Project to Infrastructure

Sustainable service via NFDI



# Benefits

## Example: WDCC

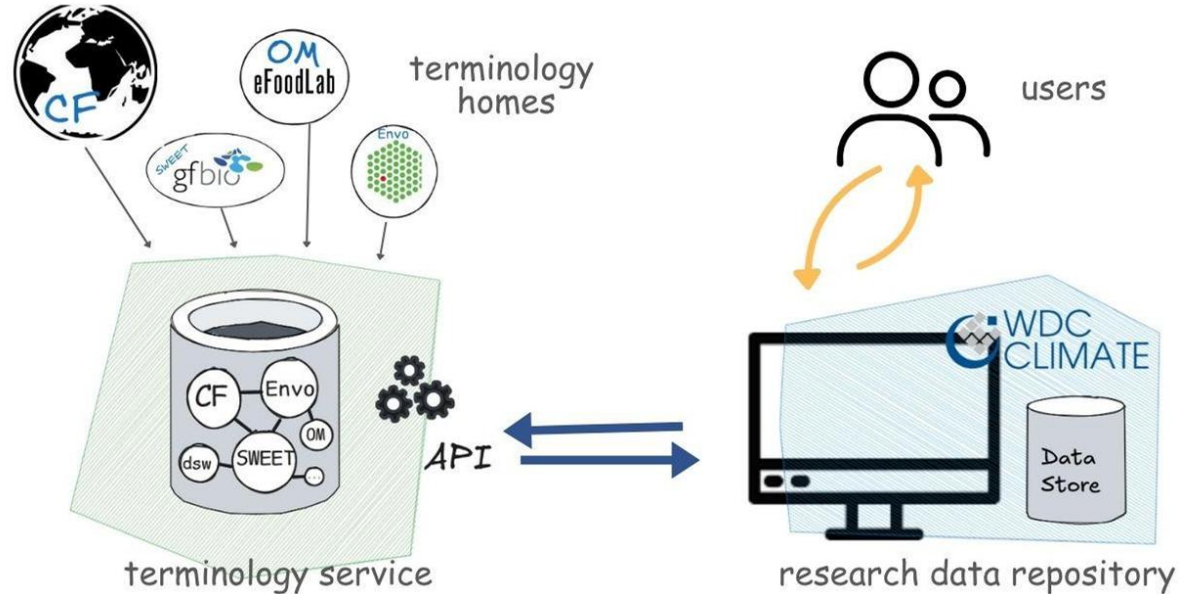
- WDCC = World Data Center for Climate, international data repository and archive for comprehensive climate data
- Improving interdisciplinary discovery by enhancing search terms using the ESS Collection



# Benefits

## Example: WDCC

- WDCC = World Data Center for Climate, international data repository and archive for comprehensive climate data
- Improving interdisciplinary discovery by enhancing search terms using the ESS Collection





# Benefits

Example: WDCC

## WDCC GUI

The screenshot shows the WDCC website homepage. At the top, there is a blue navigation bar with the WDCC logo on the left and links for 'About', 'Search', 'Submit data', 'FAIRness', 'Documentation', 'Register', and 'Login' on the right. The DKRZ logo is also present in the top right corner. The main header area features the title 'World Data Center for Climate' in a large, white font. Below the title is a search bar with the placeholder text 'Search ...' and a magnifying glass icon. Underneath the search bar are two buttons: 'Search interface' and 'Submit data'. The central content area contains a paragraph about the WDCC's mission to archive, disseminate, and publish Earth System model data. It mentions that the WDCC is certified by the Core Trust Seal (CTS) and is an accredited regular member of the World Data System. A 'CORE TRUST SEAL' logo is visible on the right side of this paragraph. Below the paragraph, it states that the WDCC follows the FAIR principles, all metadata is openly accessible, and the data is free of charge. It also mentions that the WDCC is hosted by the German Climate Computing Center (DKRZ) in Hamburg.

## Services

The 'Services' section is divided into three columns, each with a distinct icon and title. The first column, 'Data access', features a blue folder icon. The second column, 'Data archiving', features a blue database cylinder icon. The third column, 'Data publication', features a blue globe icon. Each column contains a paragraph of text describing the service. The 'Data access' section mentions the FAIR principles and provides a link to register. The 'Data archiving' section describes the long-term archiving service and mentions the DOKU service for DKRZ HPC users. The 'Data publication' section describes the DataCite data publication service and mentions the DOI system.

| Service          | Description   |
|------------------|---|
| Data access      | Regarding the <b>FAIR</b> principles all metadata is openly accessible. The data download is free for registered users. <a href="#">Please register here.</a><br><br>The <b>graphical web interface</b> of the WDCC offers a wide variety of search options based on the stored metadata. These allow a quick discovery of the data desired, sorted by project, topics, temporal and spatial coverage, or other facets.<br><br>Data from the long-term archive are downloadable |
| Data archiving   | The DKRZ provides two different tiers of long-term archiving:<br><br><b>WDCC</b> is the long-term archiving service in the WDCC. This service is available for DKRZ HPC (high performance computing) project data as well as data from external sources. Conditions for this need are to be negotiated. The full set of metadata facilitates the interdisciplinary reuse.<br><br><b>DOKU</b> is an additional service for DKRZ HPC users  |
| Data publication | The DKRZ offers a <b>DataCite</b> data publication service for WDCC long-term archived data.<br><br>The data publication process is similar to the publication of an article in a scientific journal. Data, published with a DOI at WDCC has to meet certain quality requirements. The permanent access to data is granted by the assigned unique persistent identifier DOI - data and metadata remain unchanged.<br><br>DataCite DOI data is formally citable in scientific    |

# Benefits

Example: WDCC



Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, IPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012

Search in all categories...

Citation-DOI 2200

Bounding box

Temporal coverage

Project 200

Variables 3576

Keywords 2932

Aggregation 174

Hierarchy Level 5

Type 12

Organization 513

Persons 2372

Format 51

Search ...



## Results

2732122 results found for ""

Sort by Score

Show rows: 15

1 of 182142

WINTER\_HAMBURG (Investigating climate change related impacts on the urban winter climate of Hamburg)

Project

LAMACLIMA (Output of two-phase Earth system model simulations for LAMACLIMA: idealized global land-use and high-ambition future mitigation land-use scenarios.)

Project

DKRZ\_Ita (Long-term Archiving of Climate Model Data at WDC Climate and DKRZ (DOKU))

Project

SWIFT-A (Shallow Water Ice Fracture Tracking and Acoustics)

Project

IPCC-AR6\_CMIP6 (Coupled Model Intercomparison Project Phase 6 (CMIP6) datasets)

Project

DataWave (DataWave: Collaborative Gravity Wave Research)

Project

WRF\_WIND\_EC (Mesoscale modeling of wind resources over Ecuador using WRF)

Project

HadCRU\_MLE (Maximum Likelihood Estimates of Temperatures using Data from the Hadley Centre and the Climate Research Unit)

Project

HISISP-QTP (Heat Island Intensity Prediction in an Intelligent Sponge Urban System in the Qinghai-Tibet Plateau)

Project

coastDat\_Marine-Scenarios (coastDat Water Level and Waves Future Scenarios)

Project

ClimAVA (Climate data for adaptation and vulnerability assessments)

Project

AqualNFRA

Project

DeIMod2

Project

# Benefits

Example: WDCC



Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, IPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012

Search in all categories...

☐ Citation-DOI 2200

☒ Bounding box

☒ Temporal coverage

☒ Project 200

☒ Variables 3576

Type to filter... Sort by

Show all

☐ air\_temperature 168103

☐ eastward\_wind 74118

☐ northward\_wind 71075

☐ specific\_humidity 57989

☐ area\_fraction 55321

☐ relative\_humidity 52733

☐ precipitation\_flux 51772

☐ geopotential\_height 51233

☐ cloud\_area\_fraction 37041

☐ surface\_downwelling\_shortwav... 36868

☐ surface\_temperature 36157

Search ...



## Results

2732122 results found for "wacc"

Sort by Score

Show rows: 15

< 1

of 182142 >

WINTER\_HAMBURG (Investigating climate change related impacts on the urban winter climate of Hamburg)

Project >

LAMACLIMA (Output of two-phase Earth system model simulations for LAMACLIMA: idealized global land-use and high-ambition future mitigation land-use scenarios.)

Project >

DKRZ\_Ita (Long-term Archiving of Climate Model Data at WDC Climate and DKRZ (DOKU))

Project >

SWIFT-A (Shallow Water Ice Fracture Tracking and Acoustics)

Project >

IPCC-AR6\_CMIP6 (Coupled Model Intercomparison Project Phase 6 (CMIP6) datasets)

Project >

DataWave (DataWave: Collaborative Gravity Wave Research)

Project >

WRF\_WIND\_EC (Mesoscale modeling of wind resources over Ecuador using WRF)

Project >

HadCRU\_MLE (Maximum Likelihood Estimates of Temperatures using Data from the Hadley Centre and the Climate Research Unit)

Project >

HISIP-QTP (Heat Island Intensity Prediction in an Intelligent Sponge Urban System in the Qinghai-Tibet Plateau)

Project >

coastDat\_Marine-Scenarios (coastDat Water Level and Waves Future Scenarios)

Project >

ClimAVA (Climate data for adaptation and vulnerability assessments)

Project >

AqualNFRA

Project >


PalMod2

Project >



# Benefits

Example: WDCC



Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, IPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012

Search in all categories...

☐ Citation-DOI 2200

☒ Bounding box

☒ Temporal coverage

☒ Project 200

☒ Variables 3576

Type to filter... Sort by

Show all

- ☐ air\_temperature 168103
- ☐ eastward\_wind 74118
- ☐ northward...
- ☐ specific\_hu...
- ☐ area\_fracti...
- ☐ relative\_hu...
- ☐ precipitation...
- ☐ geopotent...
- ☐ cloud\_area...
- ☐ surface\_da...
- ☐ surface\_temperature 36157

Search ...

Results

2732122 results found for ""

Sort by Score Show rows: 15

1 of 182142

WINTER\_HAMBURG (Investigating climate change related impacts on the urban winter climate of Hamburg) [Project](#)

LAMACLIMA (Output of two-phase Earth system model simulations for LAMACLIMA: idealized global land-use and high-ambition future mitigation land-use scenarios.) [Project](#)

DKRZ\_Ita (Long-term Archiving of Climate Model Data at WDC Climate and DKRZ (DOKU)) [Project](#)

SWIFT-A (Shallow Water Ice Fracture Tracking and Acoustics) [Project](#)

IPCC-AR6\_CMIP6 (Coupled Model Intercomparison Project Phase 6 (CMIP6) datasets) [Project](#)

DataWave (DataWave: Collaborative Gravity Wave Research) [Project](#)

WRF\_WIND\_EC (Mesoscale modeling of wind resources over Ecuador using WRF) [Project](#)

HadCRU\_MLE (Maximum Likelihood Estimates of Temperatures using Data from the Hadley Centre and the Climate Research Unit) [Project](#)

HISISP-QTP (Heat Island Intensity Prediction in an Intelligent Sponge Urban System in the Qinghai-Tibet Plateau) [Project](#)

scenarios (coastDat Water Level and Waves Future Scenarios) [Project](#)

Climate data for adaptation and vulnerability assessments) [Project](#)

PalModZ [Project](#)

'Eastward' indicates a vector component which is positive when directed eastward (negative westward). Wind is defined as a two-dimensional (horizontal) air velocity vector, with no vertical component. (Vertical motion in the atmosphere has the standard name upward\_air\_velocity.)

Variable information provided by The NERC Vocabulary Server (NVS), National Oceanography Centre - British Oceanographic Data Centre (BODC), <https://vocab.nerc.ac.uk>



cmip5 output1 CSIRO-QCCCE CSIRO-Mk3-6-0 rcp85 mon ocean Omon r3i1p1 v20110518  
hfbasinba

[hdl:21.14106/db3e8baad9a1512547c63a6f5b718d031a2093f8](https://hdl.handle.net/21.14106/db3e8baad9a1512547c63a6f5b718d031a2093f8) ?

Jeffrey, Stephen et al.

Dataset

[General Information](#) [Variables](#) [Data Hierarchy](#)

| Variable ↕  | Code ↕   | Aggregation ↕ | Unit ↕ |
|---|--|---------------|--------|
|  northward_ocean_heat_transport_due_to_bolus_advection |  hfbasinba (IPCC_DDC_AR5: 232) | mon           | W      |

### General information

#### Variable

##### Variable

[northward\\_ocean\\_heat\\_transport\\_due\\_to\\_bolus\\_advection](#) **DEPRECATED**

##### Description

The specification of a physical process by the phrase due\_to\_process means that the quantity named is a single term in a sum of terms which together compose the general quantity named by omitting the phrase. 'Northward' indicates a vector component which is positive when directed northward (negative southward). Northward transport by bolus advection in an ocean model means the part due to a scheme representing eddy-induced effects not included in the velocity field.

##### Replaced by

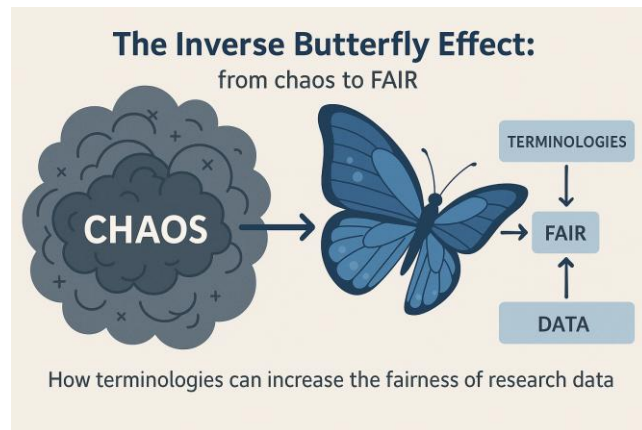
[northward\\_ocean\\_heat\\_transport\\_due\\_to\\_parameterized\\_eddy\\_advection](#)

Variable information provided by The NERC Vocabulary Server (NVS), National Oceanography Centre - British Oceanographic Data Centre (BODC), <https://vocab.nerc.ac.uk>



# Conclusion

Terminologies matter!



# Conclusion

Terminologies matter!

## Interdisciplinary Collaboration is Challenging

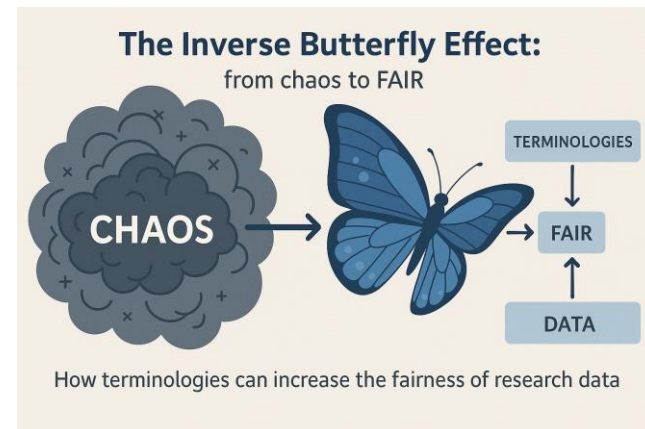
Different fields use varying vocabularies, taxonomies, and ontologies, making it hard to translate and interpret data effectively and in a comfortable way.

## The FAIR Principles Beyond Buzzwords

FAIR is not just a buzzword: actionable steps and solutions are required to let increased FAIRness of data become a reality. Using Semantic Artifacts in data ingestion, archiving, and distribution can help to increase FAIRness.

## The Role of Terminologies

All kinds of Terminologies (that is controlled Vocabularies, Taxonomies, Thesauri, or Ontologies) can help to bridge the gaps in interdisciplinary data interpretation. Tools to map and translate these terminologies help to reduce misunderstandings and improve data interoperability.



# THANK YOU!

<https://projects.tib.eu/bits/home>

**TIB** LEIBNIZ INFORMATION CENTRE  
FOR SCIENCE AND TECHNOLOGY  
UNIVERSITY LIBRARY

**SENCKENBERG**  
world of biodiversity



funded by:  
**DFG** Deutsche  
Forschungsgemeinschaft

Questions?  
Comments?  
Remarks?