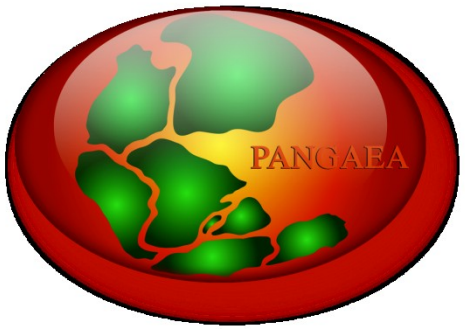
Integrated data-management in two EU-Projects

From ship-born data to scientific publications



HERMES / HERMIONE

Objectives:

- To investigate the dimensions, distribution and interconnection of deep-sea ecosystems;
- To understand changes in deep-sea ecosystems related to key factors including climate change, human impacts and the impact of large-scale episodic events;
- To understand the biological capacities and specific adaptations of deep-sea organisms, and investigate the importance of biodiversity in the functioning of deep-water ecosystems;
- To provide stakeholders and policymakers with scientific knowledge to support deep-sea governance aimed at the sustainable management of resources and the conservation of ecosystems.

Management of cruise-data

the first bottle-neck

large data sets are generated during cruises:

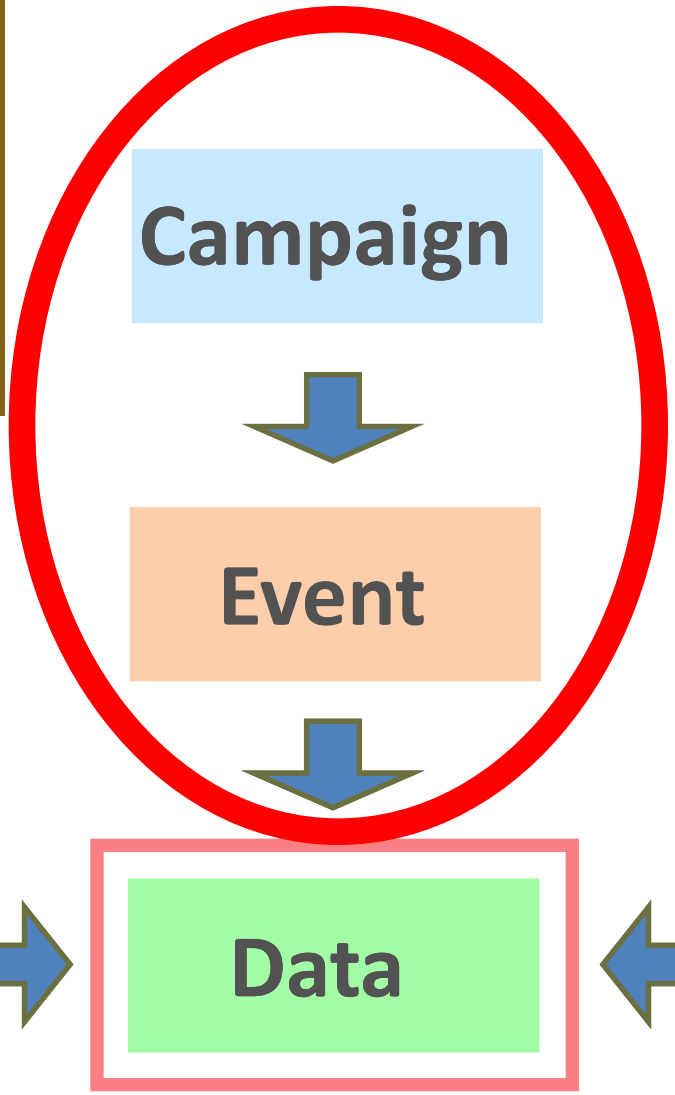
- ship station lists including sample positions and depths
- ROV, AUV and other tool's station lists
- some data (e.g. seismics) are processed during the cruises

access to cruise data:

- personal documentation (even best systems sometimes fail, does not include data of co-workers)
- preliminary data in cruise reports

How are metadata and data organized ?

Data Model
(simplified)



Author

Reference

Method

Parameter

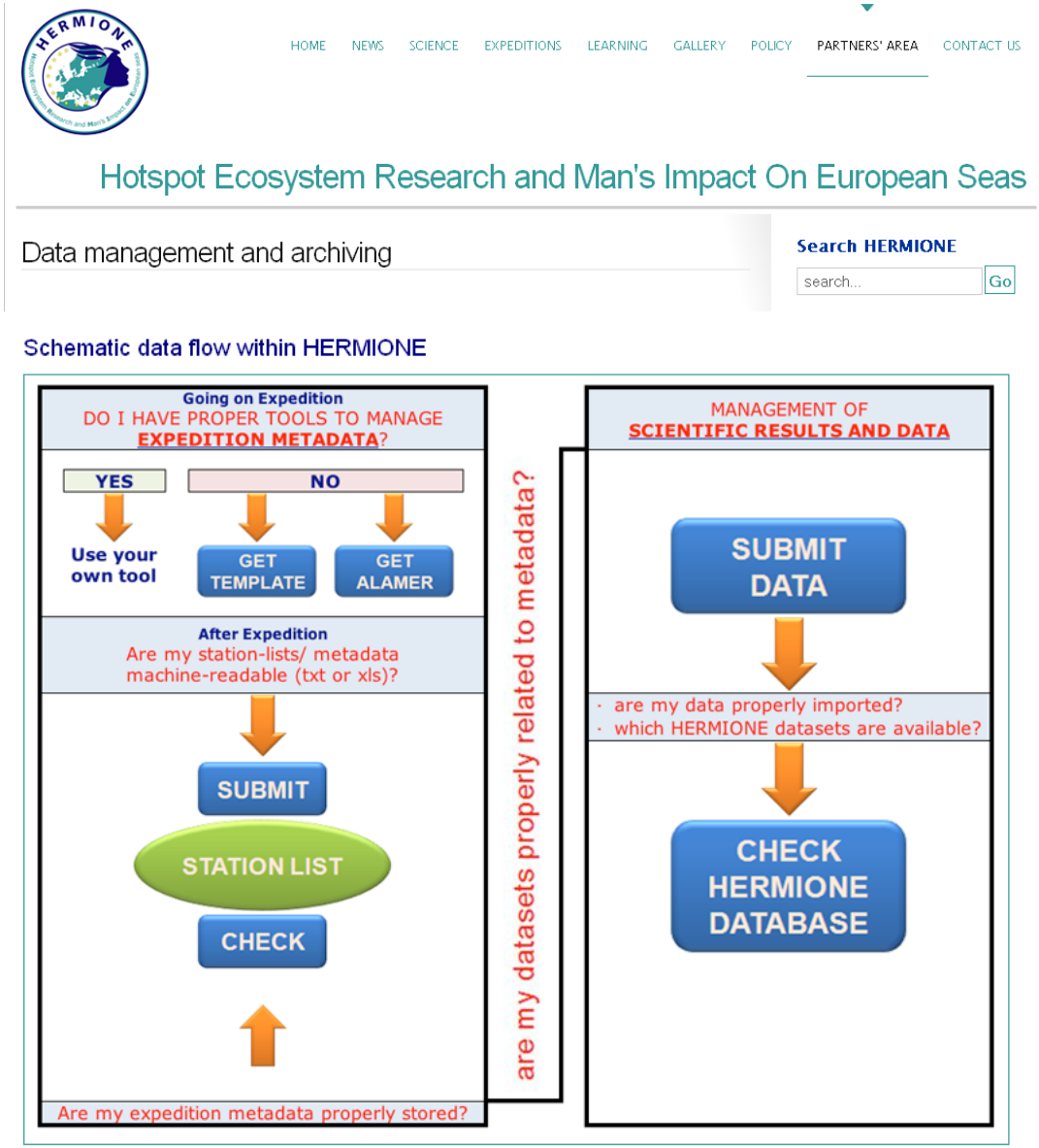
Cruise List

PANGAEA®
Data Publisher for Earth & Environmental Science

Station list of cruise MSM13/4

Event label	Event, optional label	Device	Date/Time	Latitude	Longitude	Elevation	Comment
MSM13/4-track		Underway cruise track measurements	2009-11-21T15:00	35.25623	31.60968		
MSM13/4_1000-1		Gravity corer	2009-11-27T07:41	35.33192	30.26795	-2024	SL max. 2031 m
MSM13/4_1001-1		Gravity corer	2009-11-27T08:55	35.33192	30.26791	-2024	SL max. 2034 m
MSM13/4_1002-1		Remote operated vehicle	2009-11-27T10:45	35.33483	30.27100	-2029	Beg. Aussetzen - Zugtest in Oberfläche
MSM13/4_1002_CMA4		Current meter, Aanderaa	2009-11-27T22:21	35.33196	30.26809	-2026	RCM
MSM13/4_1002_FLS1		Fish finder echolot, 20 kHz	2009-11-27T17:04			-2027	Sonar 1 recording on B: 17:04:00 - E: 17:05:14
MSM13/4_1002_FLS2		Fish finder echolot, 20 kHz	2009-11-27T18:30	35.33475	30.26865	-2029	screenshot and record of sonar 2 B: 18:30:11 - E: 18:31:00
MSM13/4_1002_MCAM1		MPI-Megacam	2009-11-27T14:31	35.33482	30.27161	-2029	MegaCam fotos of crustaceans
MSM13/4_1002_MCAM2		MPI-Megacam	2009-11-27T15:40	35.33492	30.27120	-2029	MegaCam beautiful tube worm images
MSM13/4_1002_VIDEO1		Video camera	2009-11-27T14:51	35.33490	30.27123	-2029	HDTV 1 Marker K living tubeworm B: 14:51:00 - E: 14:59:19
MSM13/4_1002_VIDEO2		Video camera	2009-11-27T17:21	35.33465	30.26869	-2029	HDTV2 at M6 /Homer 13 on black spot B: 17:21:17 - E: 17:25:01
MSM13/4_1002_VIDEO3		Video camera	2009-11-27T18:43	35.33480	30.26851	-2029	HDTV3 on bubbles M6 B: 18:43:11 - E: 18:47:34
MSM13/4_1002_VIDEO4		Video camera	2009-11-27T18:53	35.33482	30.26851	-2029	Hd 4 on on gas bubbles north of M6 B: 18:53:45 - E: 18:58:32
MSM13/4_1002_VIDEO5		Video camera	2009-11-27T19:30	35.33482	30.26845	-2029	HDTV 5 on on gas bubbles B: 19:30:11 - E: 19:34:12
MSM13/4_1002_VIDEO6		Video camera	2009-11-27T19:54	35.33482	30.26846	-2029	Hdtv 6 on on gas bubbles M6 with Mega Cam B: 19:54:30 - E: 20:02:37
MSM13/4_1002_VIDEO7		Video camera	2009-11-27T23:42	35.33217	30.26807	-2028	HD 7 on crab, reduced sed D255 B: 23:42:29 - E: 23:42:53
MSM13/4_1003-1		Autonomous underwater vehicle	2009-11-28T04:13	35.33267	30.26567	0	AUV in water
MSM13/4_1004-1		Large lift Colossos	2009-11-29T07:34	35.33300	30.26717	-2026	
MSM13/4_1005-1		Remote operated vehicle	2009-11-28T14:44	35.33267	30.26650	0	
MSM13/4_1006-1		Heat-Flow probe	2009-11-28T19:58	35.33525	30.26847	-2031	SL max. 2028 m
MSM13/4_1006-10		Heat-Flow probe	2009-11-29T00:36	35.33203	30.26865	-2037	Sl max. 2024m
MSM13/4_1006-11		Heat-Flow probe	2009-11-29T00:58	35.33188	30.26870	-2028	Sl max. 2024m
MSM13/4_1006-12		Heat-Flow probe	2009-11-29T01:30	35.33150	30.26867	-2025	keine Messung - verholen
MSM13/4_1006-13		Heat-Flow probe	2009-11-29T01:38	35.33122	30.26867	-2025	Sl max. 2026m
MSM13/4_1006-14		Heat-Flow probe	2009-11-29T02:03	35.33073	30.26870	-2024	Sl max. 2024m
MSM13/4_1006-15		Heat-Flow probe	2009-11-29T02:25	35.33038	30.26867	-2025	Sl max. 2025m
MSM13/4_1006-16		Heat-Flow probe	2009-11-29T02:53	35.32982	30.26870	-2025	Sl max. 2025m
MSM13/4_1006-2		Heat-Flow probe	2009-11-28T20:31	35.33463	30.26848	-2026	SL max. 2026 m
MSM13/4_1006-3		Heat-Flow probe	2009-11-28T21:20	35.33473	30.27162	-2025	SL max. 2026 m
MSM13/4_1006-4		Heat-Flow probe	2009-11-28T22:07	35.33373	30.26952	-2025	Sl max. 2026m
MSM13/4_1006-5		Heat-Flow probe	2009-11-28T22:39	35.33380	30.26855	-2025	Sl max. 2025m
MSM13/4_1006-6		Heat-Flow probe	2009-11-28T23:03	35.33322	30.26857	-2025	Sl max. 2025m
MSM13/4_1006-7		Heat-Flow probe	2009-11-28T23:31	35.33267	30.26862	-2023	Sl max. 2025m
MSM13/4_1006-8		Heat-Flow probe	2009-11-28T23:55	35.33245	30.26860	-2024	Sl max. 2026m
MSM13/4_1006-9		Heat-Flow probe	2009-11-29T00:16	35.33223	30.26860	-2025	Sl max. 2027m

HERMIONE Website



Ecosystem studies require interdisciplinary work

- bathymetry/seismics/video mosaicking (e.g. precise mapping of habitats)
- physics and geochemistry (e.g. temperature, pH, O_2, H_2S, SO_4^{2-}, CH_4)
- microbiology (e.g. identification, distribution, abundance)
- zoology (e.g. taxonomic identification, physiological studies)
- etc.

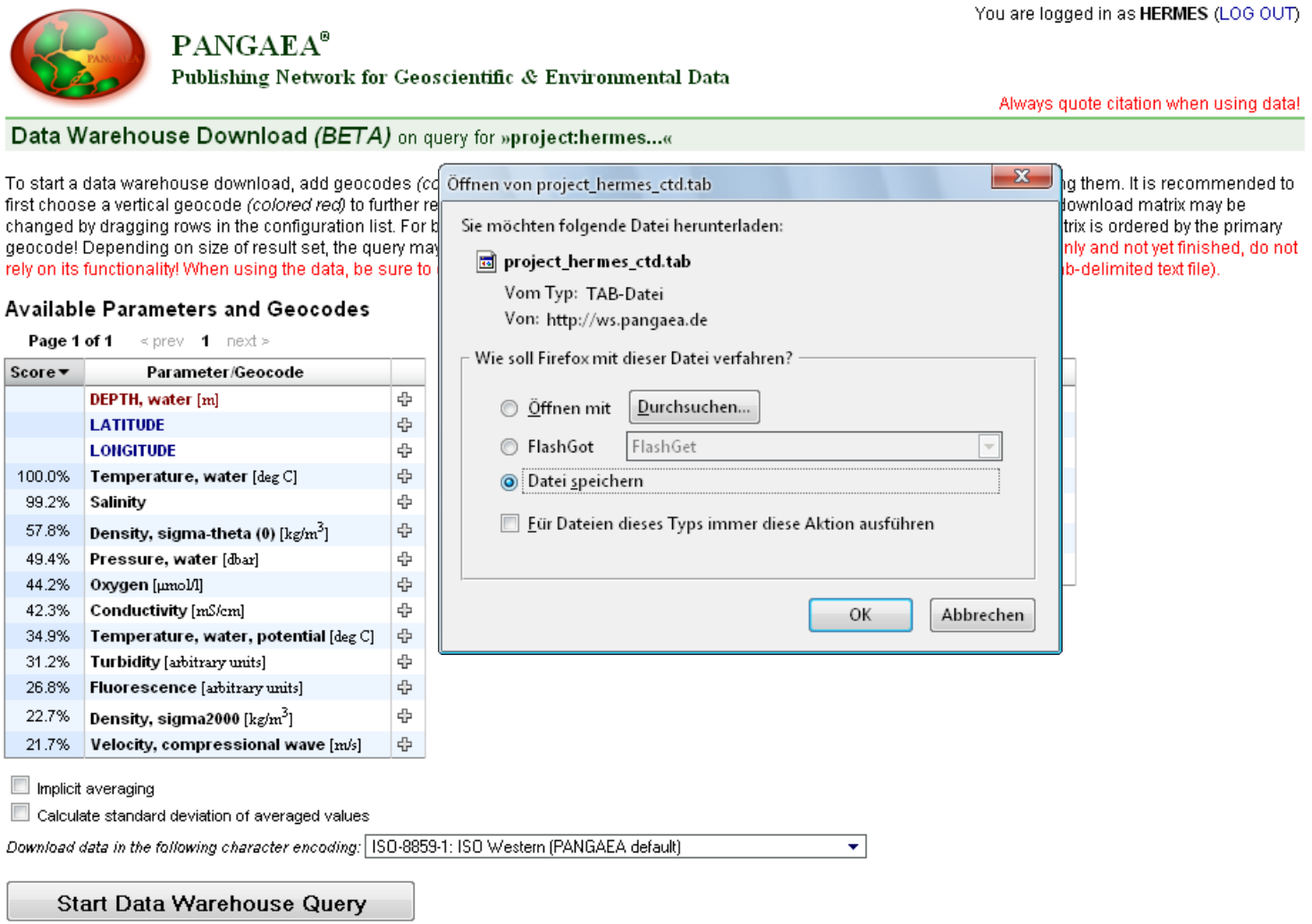
From where to take contextual data?

- cooperation
- publications:
  - take very long before data become available
  - do not always contain primary data; not all data sets of research cruises are published
  - old data set often not available electronically

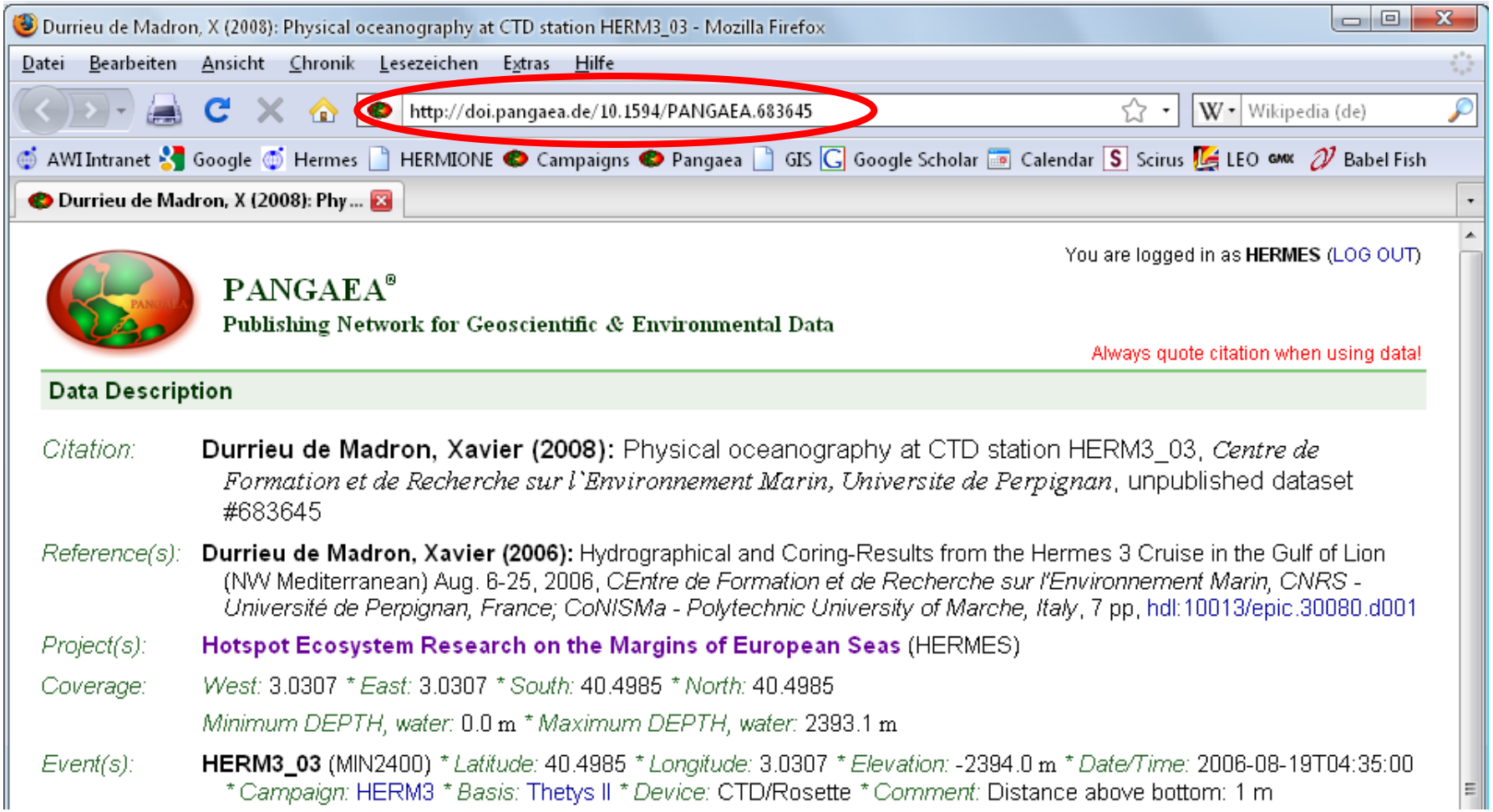
→ PANGAEA: useful tools to access rapidly contextual data from specific ecosystems (unpublished data are under moratorium)

PANGAEA®
Data Publisher for Earth & Environmental Science

Get an Overview



Digital Object Identifier doi>



second bottle-neck

SCIENTISTS RETICENCE TO PUBLISH THEIR DATA IN AN EARLY PHASE

Building a Culture of Data Citation

CREATE

Scientist creates a research dataset

3

doi is used in data citation

4

USE

Research community use the doi to access the dataset and carry out related research

Research community generate new publications using the doi to reference the dataset

5

Dataset is stored in a publicly accessible repository

Repository assigns to a service to mint a Digital Object Identifier(doi) for the dataset

1

2

Citation metric services (e.g Scopus,Web of Knowledge) accumulate citation references to the dataset

MEASURE

6

Funding agencies and scientists review publication and dataset citation metrics

REWARD

7

Researcher future funding and promotion influenced by dataset citation metrics

Author

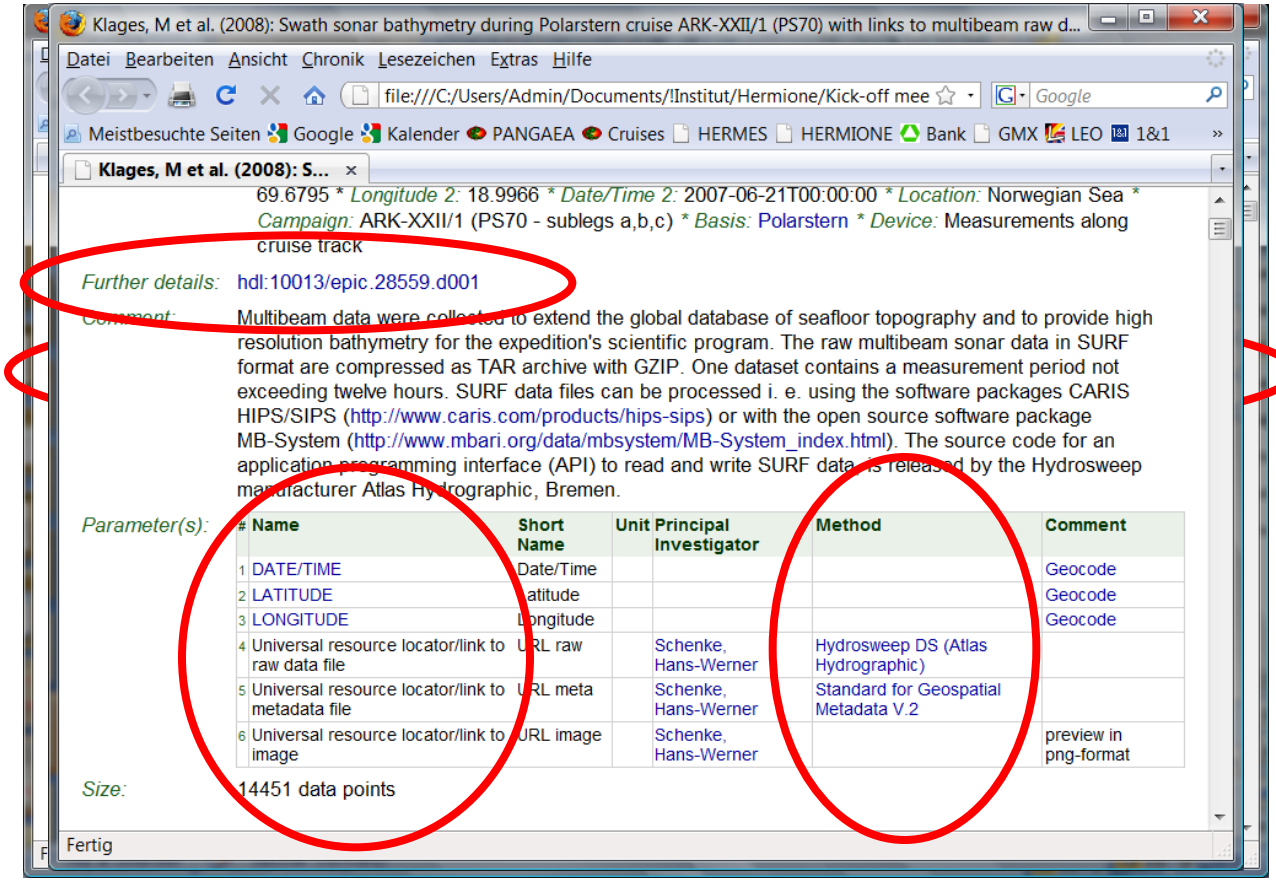
Reference

Data

Method

Parameter

doi:10.1594/PANGAEA.681690



Author

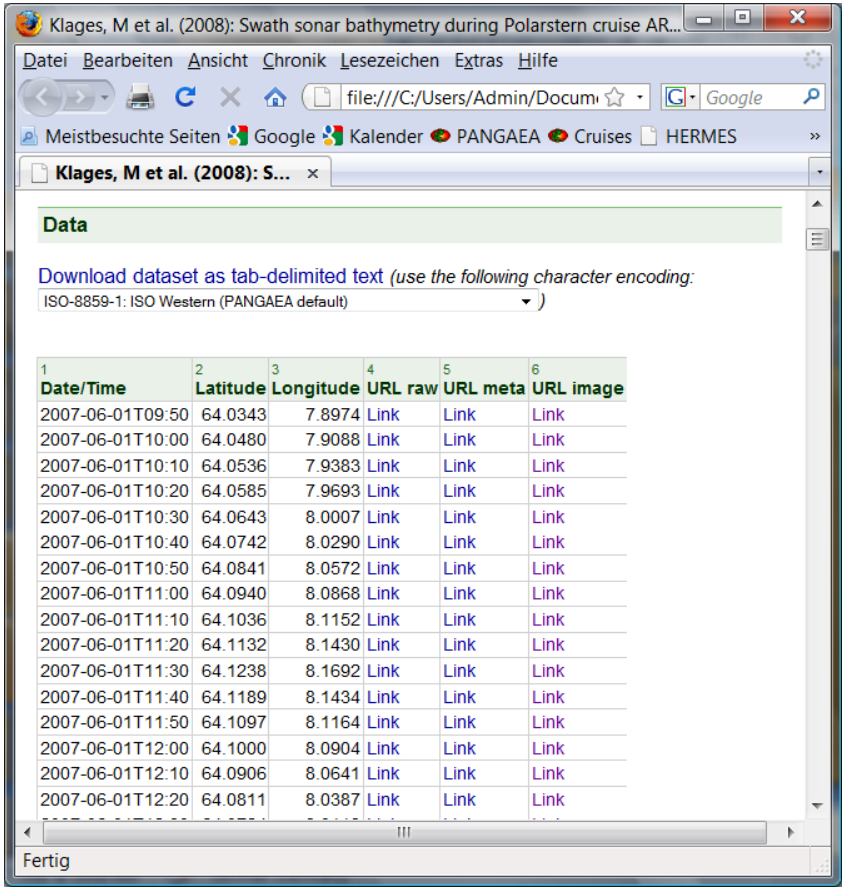
Reference

Data

Method

Parameter

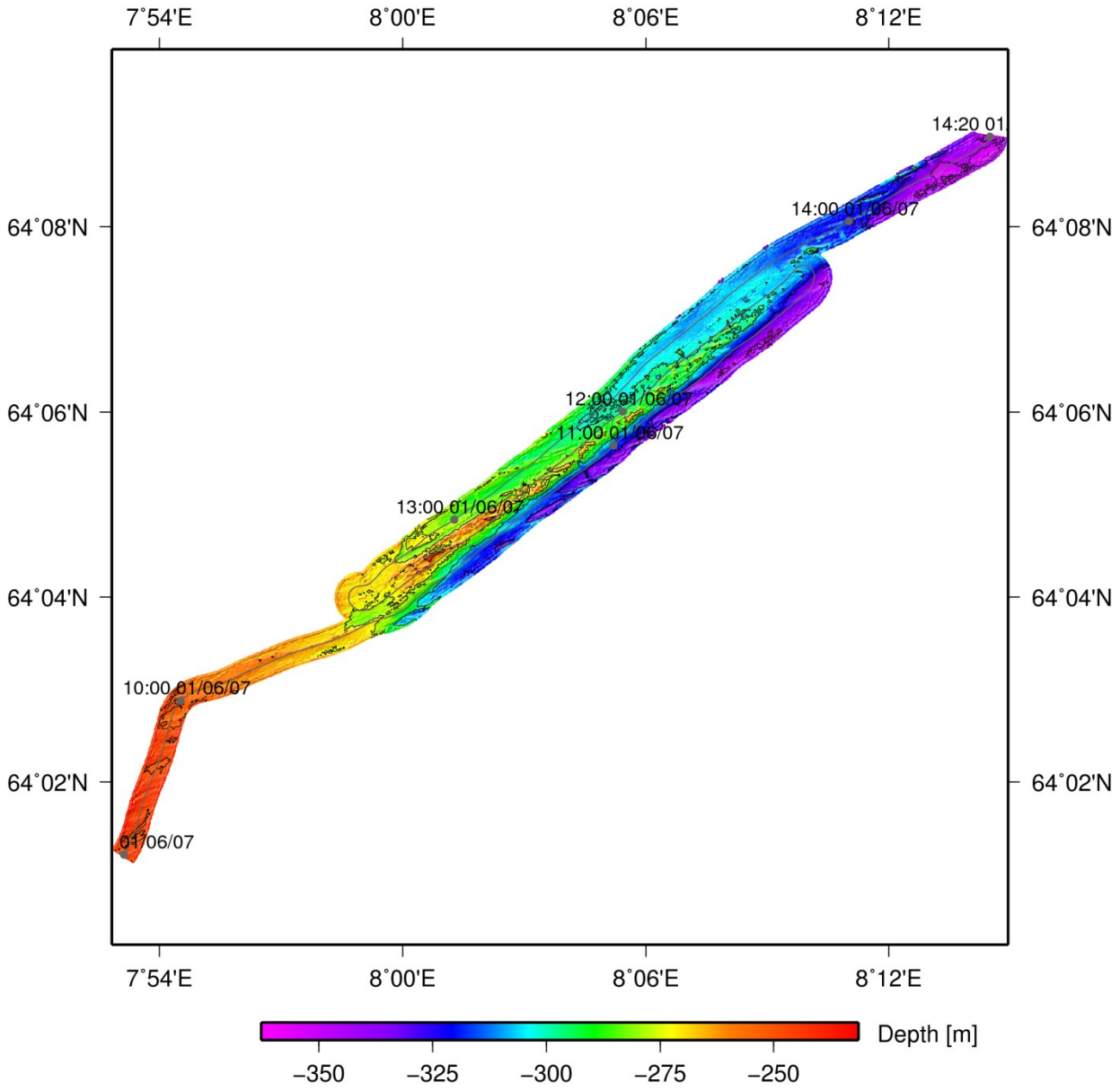
doi:10.1594/PANGAEA.681690



1 Date/Time	2 Latitude	3 Longitude	4 URL raw	5 URL meta	6 URL image
2007-06-01T09:50	64.0343	7.8974	Link	Link	Link
2007-06-01T10:00	64.0480	7.9088	Link	Link	Link
2007-06-01T10:10	64.0536	7.9383	Link	Link	Link
2007-06-01T10:20	64.0585	7.9693	Link	Link	Link
2007-06-01T10:30	64.0643	8.0007	Link	Link	Link
2007-06-01T10:40	64.0742	8.0290	Link	Link	Link
2007-06-01T10:50	64.0841	8.0572	Link	Link	Link
2007-06-01T11:00	64.0940	8.0868	Link	Link	Link
2007-06-01T11:10	64.1036	8.1152	Link	Link	Link
2007-06-01T11:20	64.1132	8.1430	Link	Link	Link
2007-06-01T11:30	64.1238	8.1692	Link	Link	Link
2007-06-01T11:40	64.1189	8.1434	Link	Link	Link
2007-06-01T11:50	64.1097	8.1164	Link	Link	Link
2007-06-01T12:00	64.1000	8.0904	Link	Link	Link
2007-06-01T12:10	64.0906	8.0641	Link	Link	Link
2007-06-01T12:20	64.0811	8.0387	Link	Link	Link

data sources

survey platform:	R. V. POLARSTERN
sonar system:	Atlas Hydrographics Hydrosweep DS2
data set name:	152_ark22-1_60F30706010940.xyz
data set info:	length: 25.3 nm, time: 4.7 h, speed: 5.4 kn



Sum up….

- not all data will be published
  - but in PANGAEA they are still available for new approaches or questions
- information about available data for the different ecosystems
  - establishing of new co-operations or co-authorships
- time record (e.g. climate change)
- model and budget estimations

Sum up….

- Bottleneck: scientific data-provision
- Scientists fear that:
  - it is very time consuming to enter data
  - wrong data cannot be corrected
  - it is difficult to retrieve data
  - data are not protected and will be abused by colleagues