

## GEOTRACES: SHARING MARINE GEOCHEMICAL DATA

Maite Maldonado, UBC Reiner Schlitzer, AWI







## ZOOM ROOM RULES

Muted (default) – use Chat to ask questions

No Video (default)

No Screen Sharing (default)

Closed captioning enabled

Presentations and the recording will be available at the IPO site: <u>https://GEOTRACES.org</u>

Also available through, the Ocean Decade Laboratory webpage.



#### AGENDA

- Welcome Elena Masferrer (<u>ipo@geotraces.org</u>)
- Introduction to GEOTRACES video
- Lessons learned in creating an international data product Maite Maldonado (<u>mmaldonado@eos.ubc.ca</u>)
- Accessing IDP2021 Reiner Schlitzer (<u>Reiner.Schlitzer@awi.de</u>)
- Potential uses of the data product: The case of genomics data Maite Maldonado
- Conclusion and Q&A Maite Maldonado

www.geotraces.org



@geotraces #OceanData #OceanScience

## GEOTRACES INTERNATIONAL PROGRAMME



Animation by Adrian Artis Directed by Catherine Jeandel and Elena Masferrer Voice by Thomas Boutilier / Thanks to Rogue Elephant Youtube link: <u>https://youtu.be/FoGnPTpOlCg</u>





## LESSONS LEARNED IN BUILDING AN INTERNATIONAL DATA PRODUCT

#### Maite Maldonado

University of British Columbia, Canada







## GEOTRACES MISSION

GEOTRACES mission is to identify processes that control the distributions of key trace elements and isotopes (TEIs) in the ocean, and to establish the sensitivity of these distributions to changing environmental conditions

Scientists from more than **35 nations** have been involved in the programme, which is designed to study all major ocean basins

2006 – Science Plan ..... 2010 – First cruises and IPO..... 2014, 2017 and 2021 – Release of data products







## KEY TRACE ELEMENTS -> RARE BUT PRECIOUS



DATA PRODUCT

Act as micronutrients to control ocean productivity and ecosystems (e.g. Fe, Zn, Cd, Cu, Ni, Co, Mn)

**Contaminants** in the present and future ocean (e.g. Pb, Hg)

**Trace modern processes** in the ocean (e.g. Al, REEs, Ra and Th isotopes, <sup>3</sup>He, <sup>15</sup>N, Nd isotopes)

Used as **proxies to reconstruct past climate** (e.g. <sup>231</sup>Pa, <sup>230</sup>Th, Cd, <sup>15</sup>N, <sup>30</sup>Si, Ba, ɛNd)



The sum is greater than the parts!

# GEOTRACES data document climate mediated changes in the Arctic...

Measurements of <sup>228</sup>Ra by US team in the Arctic surface waters has doubled due to an intensification of shelf-derived inputs

These climate mediated changes may have dramatic impacts on Arctic food webs and mammal populations





Kipp et al 2018

# GEOTRACES data also provide evidence of the efficacy of legislations to reduce contaminant emissions

The ability to measure natural lead again in seawater shows the decrease of anthropogenic lead contamination.

This demonstrates the efficacy of major environmental regulations.





Bridgestock et al, 2016

# GEOTRACES data provide a baseline to assess future anthropogenic inputs of REE into the ocean

Growing use of the rare earth elements in various industries.

Gadolinium (Gd) in MRI, now released in environment.

Anthropogenic Gd measurable in coastal waters, and may provide a new tracer of wastewater.



Pedreira et al, 2018



## DEVELOPMENT, IMPLEMENTATION AND SHARING OF CONTAMINATION-FREE METHODS







#### **Trace Elements**

Present in seawater in extremely low concentrations (orders of magnitude of 10<sup>-7</sup> mol/kg to 10<sup>-15</sup> mol/kg)









## CAPACITY BUILDING

## Trace metal - clean sampling technology and methods

- Many nations lack infrastructure and expertise
  for clean sampling
- Principal barrier is sampling at sea, not analyses
- GEOTRACES offers international assistance in design, construction and use of clean sampling systems
- It also offers expertise and training (through summer schools, workshops, etc.)
- Offers online resources (best practices, tutorials, etc.)





## IMPORTANT ASPECTS TO GEOTRACES

#### Collaboration and coordination

**SCIENCE PLAN 2006** was key for the coordination of such ambitious global research program

- -> International Project Office
- Good governance -> national representatives & scientific committees to oversight developments (EDI compliant)
- High quality data -> Standardization & Intercalibration
- Active data Management, including online data registration portal
- Release of intermediate data product





## GEOTRACES FIELD WORK



#### To date:

**35** nations

132 cruises completed

45 sections completed

GEOTRACES cruises must follow defined criteria and be approved by the GEOTRACES SSC.



# **Standardization and Intercalibration**: Assurance of high-quality, internally consistent global data set

A coherent global data set with contributions from scientists worldwide requires high standards of intercalibration

- Sharing resources, expertise and equipment when needed
- Recommended sampling protocols ("The Cookbook") is online and updated regularly: <u>https://www.geotraces.org/methods-cookbook/</u>
- All GEOTRACES cruises must follow intercalibration procedures: <u>https://www.geotraces.org/intercalibration-procedures/</u>
- Reference seawater samples provided to analysts as standards
- A Standards and Intercalibration Committee evaluates data to ensure accuracy and assists investigators who experience data quality problems





## HIGH-QUALITY DATA = INTERCALIBRATION

## How does the GEOTRACES S&I committee assess data quality for IDP?



Dialogue between S&I committee and analyst to obtain adequate quality control for IDP



>To submit data for the GEOTRACES Data Products: (1) submit data, AND in parallel, (2) submit an intercalibration report



# Active data management: Facilitate community access to a well designed database

- Ensures public access to a well-designed database
- Data Management Committee provides oversight and interfaces with data generators. Recommend data policies, standards and formats for data submission
- Lesson learned from 2017 Intermediate Data Product -> Decision of creating an online portal for data registration
- Data Assembly Center (lead by Dr. Mohamed Adjou) compiles data in an organized and accessible structure (<u>www.bodc.ac.uk/geotraces/</u>)
- Formalised parameter names
- FAIR Use Principles for data access and sharing by whole community





## GEOTRACES Data for Ocean Research Portal (DOoR)



**Project Manager:** Elena Masferrer, International Project Office, LEGOS

Development by SEDOO-OMP: François André

François André Guillaume Brissebrat Arnaud Mière

#### Welcome to the GEOTRACES Data for Oceanic Research (DOoR) Portal

The GEOTRACES DOoR is for:

- >> Scientists to register datasets for inclusion in GEOTRACES Data Products and track its status. The registration process should take about 15 minutes for first time users.
- >> Providing ORCIDs for other scientists to be associated with each dataset (graduate students, postdocs, etc.).
- >> Generating and downloading templates needed to submit intercalibration reports, submitting and/or resubmitting your intercalibration reports and tracing their progress.
- » Generating and downloading data templates to be used for data submission to the appropriate data centre and track the status of inclusion in IDP.
- >> Scientists to provide permission for the inclusion of your data in GEOTRACES Data Products.
- >> Providing DOIs of publications that include your data.

This is not a replacement for data submission to GDAC or the relevant US/Dutch/French/Chinese national data centre.

For further information please refer to the flow chart "How to Ensure that your data are in Intermediate Data Product (IDP)". **DOoR tutorials** - Detailed information on the DOoR functions is available on the Bow to document or on the Video guide available on Source Product (IDP).

#### Please use your ORCID to login

Register or Connect your ORCID iD



#### GEOTRACES Data for Ocean Research Portal (DOoR)

**On-line portal** to collect data and metadata from GEOTRACES scientists to build the data product - 880,000 data values from 77 cruises.

**Multilevel management tool (6 interfaces)** for scientists and technicians who review (intercalibration), process and assemble the data and metadata. One interface for function/group.







## ACCESSING THE GEOTRACES INTERMEDIATE DATA PRODUCT 2021

#### **Reiner Schlitzer**

Alfred Wegener Institute, Germany











By releasing and sharing more hydrographic and marine geochemical high-quality data from 77 cruises, we wish to strengthen and intensify the collaboration within the marine geochemical community itself, but also invite colleagues from other communities to join us.

#### IDP2021 consists on:

(1) a compilation of digital trace metal data 100,000 samples from 77 cruises (geotraces.org/dp)





Dissolved Fe in the Atlantic Ocean



(2) the eGEOTRACES Electronic Atlas (egeotraces.org)

## ACCESS THE IDP2021





## ACCESS THE IDP2021: Reiner Schlitzer (AWI) eGEOTRACES ATLAS

#### www.geotraces.org





www.geotraces.org

#### #GEOTRACESDataProduct #OceanData #OceanScience

## eGEOTRACES ATLAS

#### https://egeotraces.org/



# <page-header>

#### Description: Concentration of dissolved Fe.

Data creators: Tim Conway | Tomoharu Minami | Jun Nishioka | Hajime Obata | Matthias Sieber | Yoshiki Sohrin | Derek Vance | Linjie J

Other tracers along this section: Al dissolved | Cd dissolved | CTDSAL | Cu dissolved | delta.Cd\_114\_110 | delta.Fe\_56\_54 | Mn disse Potential.Temperature | Siticate | Zn dissolved

3D scenes with this tracer: Arctic | Atlantic | Indian Ocean | North Atlantic | Pacific | South Atlantic | South Pacific Publications for this tracer: GP19



© 2021 Reiner Schiltzer, Alfred Wegener Institute, Bremerhaven, G



1463 sections

269 3D scenes

## ACCESS THE IDP2021: DATA

#### www.geotraces.org





## 1 Bulk (full packages) download

2 Online subsetting and extraction

## 3 Online analysis and visualization

**New**: Agreement to the **IDP2021 Fair Data Use Statement** is required, but registration and login are no longer necessary.



## ACCESS THE IDP2021

#### www.geotraces.org





## IDP2021 BULK DOWNLOAD

#### www.bodc.ac.uk/geotraces/data/idp2021/

#### **Five Packages:**

Seawater Discrete Sample Data Seawater Sensor Data Aerosol Data Precipitation Data Cryosphere Data

#### Three Formats:

csv ASCII netCDF ODV collection



#### **GEOTRACES Intermediate Data Product 2021 (Version 1)**

The GEOTRACES Intermediate Data Product 2021 (IDP2021) is released on 17 November 2021. The IDP2021 represents a major new data release. Compared to IDP2017, IDP2021 improves data coverage significantly in all ocean basins, especially in the Arctic, the Indian and the Pacific Oceans. Overall, the new data product contains more than twice the number of cruises, stations and actual data values.

IDP2021 is, open access data sets, licensed under a Creative Commons Attribution 4.0 International license. Please, read the Fair Data Use Statement for GEOTRACES IDP2021<sup>(2)</sup>/<sub>2</sub> before using IDP2021.

The original versions of the IDP2021 digital datasets can be accessed from the links below - please use the <u>current version</u> of the digital datasets for any current research.

The GEOTRACES IDP2021 consists of two parts:

- the digital datasets available for download below or as a subset via the <u>online service</u> hosted by the Alfred Wegener Institute (AWI).
- 2. the <u>eGEOTRACES Electronic Atlas</u> is based on the digital data package and provides section plots (as illustrated below) and animated 3D scenes for many of the parameters. The 3D scenes provide geographical context, which is crucial for correctly assessing extent and origin of tracer plumes as well as for inferring processes acting on the tracers and shaping their distribution.

#### IDP2021 digital datasets (original version)

The IDP2021 digital datasets are made available to <u>DOWNLOAD</u> by the GEOTRACES International Data Assembly Centre hosted by the British Oceanographic Data Centre (BODC). They are provided in 3 data formats: ASCII, NetCDF and ODV collections. The latter is for use with the <u>Ocean Data View software</u>.

If the data sets are used in a presentation or publication then we ask that you acknowledge the source. This should be of the form: GEOTRACES Intermediate Data Product 2021 (IDP2021), NERC EDS British Oceanographic Data Centre NOC. doi:10.5285/cf2d9ba9-d51d-3b7c-e053-8486abc0f5fd.

Please help us to improve the Intermediate Data Product by sending us your feedback.





Distribution of dissolved iron (Fe) along GEOTRACES sections in the Pacific Ocean (IDP2021). () 2021 Reiner Schiltzer, Alfred Wegener Institute (AWI) Bremerhaven, Germany

## ACCESS THE IDP2021

#### www.geotraces.org





## IDP2021 SUBSETTING AND EXTRACTION

#### geotraces.webodv.awi.de/



#### Select:

Cruises Domain Time range



## IDP2021 SUBSETTING AND EXTRACTION

#### geotraces.webodv.awi.de/

#### Select:

Variables (entire groups or individual entries)



## IDP2021 SUBSETTING AND EXTRACTION

#### geotraces.webodv.awi.de/



#### Select: Outputformat

**Receive:** Data subset as zip file



## ACCESS THE IDP2021

#### www.geotraces.org





## IDP2021 online Analysis and visualization

#### geotraces.webodv.awi.de/ or explore.webodv.awi.de/



## IDP2021 online Analysis and visualization

geotraces.webodv.awi.de/ or explore.webodv.awi.de/

#### Features:

- ODV-like interface
- All *eGEOTRACES* section views
- No data download or software installation needed



S 16/3149: GP19 Fe D CONC BOTTLE

Day of Year

32

R



## Summary

- IDP2021 is an exciting new resource.
- Use it for your research, teaching or outreach activities.





Maite Maldonado (UBC, Canada)







To strengthen collaborations between the marine geochemical and the biological / omics' communities



#### 4 Types of omics data available in the IDP2021

- 480 Metagenomes
- 14 Single-Cell genomes
- 273 16S-18S-rRNA gene analyses

480 16S/18S amplicons (Functional & taxonomic analyses of metagenomes)



Major challenge: omics data are dynamic (changes depending on ever-growing genome reference databases) while the TEI data are static

#### Acknowledgements to:

Reiner Schiltzer (AWI) Paul Berube (MIT) Jesse McNichol (U. Southern California) 4 Types of Sample Descriptors for omics data linkages in the IDP2021:

To NCBI (National Center for Biotechnology Information database)

To EMBL-EBI (European Bioinformatic Institute database)

And many others who made this possible, specially members of Penny Chisholm lab



#### **OMICs linkages**

4 Sample Descriptors in the **Metadata variables**, (associated with **specific bottle numbers**):

- Accession numbers at NCBI or EMBL-EBI
- Or links to functional and taxonomic analyses webpages

#### OMICs linkages in Metadata VARIABLES #:

**11.** NCBI\_Metagenome\_BioSample accession #:

**12.** NCBI\_Single\_cell\_genome\_BioProject accession #: <u>@ https://www.ncbi.nlm.nih.gov/bioproject/</u>

**13.** NCBI\_16S-18S-rRNA-gene\_BioSample accession #: <u>@ https://www.ncbi.nlm.nih.gov/biosample/</u>

**14.** EMBL\_EBI\_Metagenome\_MGNIFY\_Analysis\_Accession #: <a href="mailto:@https://www.ebi.ac.uk/metagenomics/analyses/MGYA">@https://www.ebi.ac.uk/metagenomics/analyses/MGYA</a>



Metagenome Stations

#### **OMICs linkages**

#### 4 Sample Descriptors in the Metadata variables:

+ り 🤆 🏥 💾

- Accession numbers at
  NCBI or EMBL-EBI
- Or links to functional and taxonomic analyses webpages



	Station ID: 2539			
~	Cruise	GP13		
	Station	Station 32 (B)		
	Longitude	176.042°W		
	Latitude	32.5°S		
	Date	29 May 2011		
_	Time	16:59:29		
	CTDPRS_T_VALUE_SEN	[10 - 2401]		
	Bot. Depth [m]	5725		
	Operator's Cruise Name	SS2011		
	Ship Name	RV Southern Surveyo	pr	
	Period	14/05/2011 - 05/06/	2011	
	Chief Scientist	Bowie Andrew		
1	GEOTRACES Scientist	Hassler Christel		
			search	-
	1: CTDPRS_1_VALUE_SE		77	
	2: DEPTH [ff] 2: Departs Rattle Number		76	
	4. CEOTRACES Sample I		15	
	4: GEOTRACES Sample I	D	No problem reported (0)	
	5. Dottle Flag			
	7: Sampling Davica		33 CTD	
	8: BODC Bottle Number		1224218	
	9: BODC Event Number		1800307	
ſ	10: Single-Cell ID		1099597	
	11: NCBL Metagenome	BioSample Accession	SAMN07136753	
	12: NCBL Single-Cell-Ger	ome BioProject A	SAHNO/150/55	
1	13: NCBI 16S-18S-rRNA-	gene BioSample	SAMN15928721	
	14: FMBL_FBI_Metageno	me MGNIEY Analy	https://www.ebi.ac.uk/metagenomics/analy	15
_	15: CTDTMP_T_VALUE_S	SENSOR [deg C]	19.996	-
	16: CTDSAL D CONC SE	NSOR [pss-78]	35,487	
	17: SALINITY D CONC F	BOTTLE		
	18. CEC-11 D CONC BO	TTLE [pmol/ka]		
	19: CFC-12 D CONC BO	TTLE [pmol/kg]		
	20: CEC113 D CONC BC	TTLE [pmol/kg]		
	21: SE6 D CONC BOTTL	F [fmol/ka]		
	22: He D CONC BOTTLE	E [nmol/ka]		
		- [		_



## METAGENOME BIOSAMPLE ACCESSION NUMBERS AT NCBI

Example 1. Metagenome sample accession numbers at NCBI

www.ncbi.nlm.nih.gov





## METAGENOME BIOSAMPLE ACCESSION NUMBERS AT NCBI

Example 1. Variable #11 Metagenome sample accession numbers at NCBI

www.ncbi.nlm.nih.gov

				<u> </u>	in in to i
BioSample	BioSample v			Search	
	Advance	d			
	OVID-19 Information	formation (NIH)   <u>SARS-CoV-2 data (NCBI</u> )   <u>Prevention and trea</u>	tment information (HHS)	<u>Español</u>	E
Full <del>-</del>			Send to: -		
				Related information	
MIMS Envi	ronmental/Metagenome sample fr	om marine metagenome		BioProject	
Identifiers	BioSample: SAMN07136753 <mark>;</mark> Sample n	ame: S0286; SRA: SRS2329440		SRA	
Organism	marine metagenome			Taxonomy	
organism	unclassified entries; unclassified sequences; m	etagenomes; ecological metagenomes			
Package <u>MIMS: metagenome/environmental, wate</u>		r; version 5.0		Recent activity	
Attributes				Tu	urn Off
Attributes	collection date	2011-05-29		MIMS Environmental/Metagenon	ne sam
	depth	/om		from marine metagenome	DIOS
	local scale environmental context			Q PRJNA445865 (1)	Biof
	environmental medium	water		Single colls of gyapabastaria and	deump
	geographic location	Pacific Ocean		microorganisms	Biof
	latitude and longitude	<u>32.5 S 176 W</u>		Q SAMN07136495 (1)	
	geotraces_section	GP13			BioS
	cruise_id	SS2011		GP13-S0286	
	cruise_station	32			bios
	bottle_id	SS2011Station32Cast35Niskin13			See m
Description	Keywords: GSC:MIxS;MIMS:5.0				



BioProject <u>PRJNA385854</u> marine metagenome Retrieve <u>all samples</u> from this project

## FUNCTIONAL & TAXONOMIC ANALYSES EMBL-EBI

#### Example 2. Variable #14

https://www.ebi.ac.uk/meta genomics/analyses/MGYA0 0452481#overview



#### Analysis MGYA00452481

#### Other analyses

crivicia duality contra		T unclional analysis	Download
escription [-]			
Study:		MGYS00005294	
Sample:		SRS2329440	
Run:		SRR5788301	
Pipeline version:		4.1	
periment detail	s [-]		
	<u> </u>		
Experiment type:		metagenomic	
Instrument model:		NextSeq 550	
Instrument platform:		ILI UMINA	



## FUNCTIONAL & TAXONOMIC ANALYSES EMBL-EBI

#### Example 2.

Functional & taxonomic analyses of metagenomes at EMBL-EBI

 https://www.ebi.ac.uk/me tagenomics/analyses/MG YA00452481#overview

Browse data Sequence search Submit data Text search API About Help **P**Login Overview Analysis MGYA00452481 Other analyses Overview Quality control Taxonomic analysis F nctional analysis Download -rRNA  $\odot$ small subunit rRNA large subunit rRNA

These are the results from the taxonomic analysis steps of our pipeline. You can switch between different views of the data using the menu of icons below (pie, bar, stacked and interactive krona charts). The data used to build these charts can be found under the "Download" tab.





## FUNCTIONAL & TAXONOMIC ANALYSES EMBL-EBI

HOME / STUDY MGY800005294 / SAMPLE SR82329440 / RUN SRR5788301 / ANALYSIS MGYA00452481

#### Analysis MGYA00452481



These charts present the functional analysis of our pipeline, which focuses on matches to the <u>InterPro database</u> and <u>IGO</u> terms. These summarise the functional content of the sequences in the sample. The full set or results mes may be found under the "Download" tab.



#### Example 2.

Functional & Taxonomic analyses of metagenomes at EMBL-EBI

 https://www.ebi.ac.uk/me tagenomics/analyses/MG YA00452481#overview





# CONCLUSION & QUESTIONS

Maite Maldonado (UBC, Canada)







## CONCLUSION

**GEOTRACES data provide the clues** to tackling, understanding and predicting the future of the Ocean and the planet.

- Data generated following a Global Ocean Science Plan
- Through International collaboration and coordination for:
  - sample collection across all ocean basins
  - data processing and management
- This dataset contains **high-quality** marine geochemical data (1,721 publications to date)
- Open access to data in a timely manner (3 open source data products 2014, 2017 and 2021)
- And visuals within the data product for stakeholders, educators and the public.
- GEOTRACES provides and shares infraestructure, knowledge and technologies for capacity building worldwide



## Contact GEOTRACES

International Coordination:

**GEOTRACES International Project Office** (LEGOS-OMP, Toulouse, France)

Catherine Jeandel (Scientific Director) Elena Masferrer Dodas (Executive Officer) ipo@geotraces.org

Data Management:

Mohamed Adjou (Data Manager) GEOTRACES Data Assembly Centre (BODC, Liverpool) geotraces.dac@bodc.ac.uk



#### www.geotraces.org

Subscribe to our eNewsletter and mailing list!

