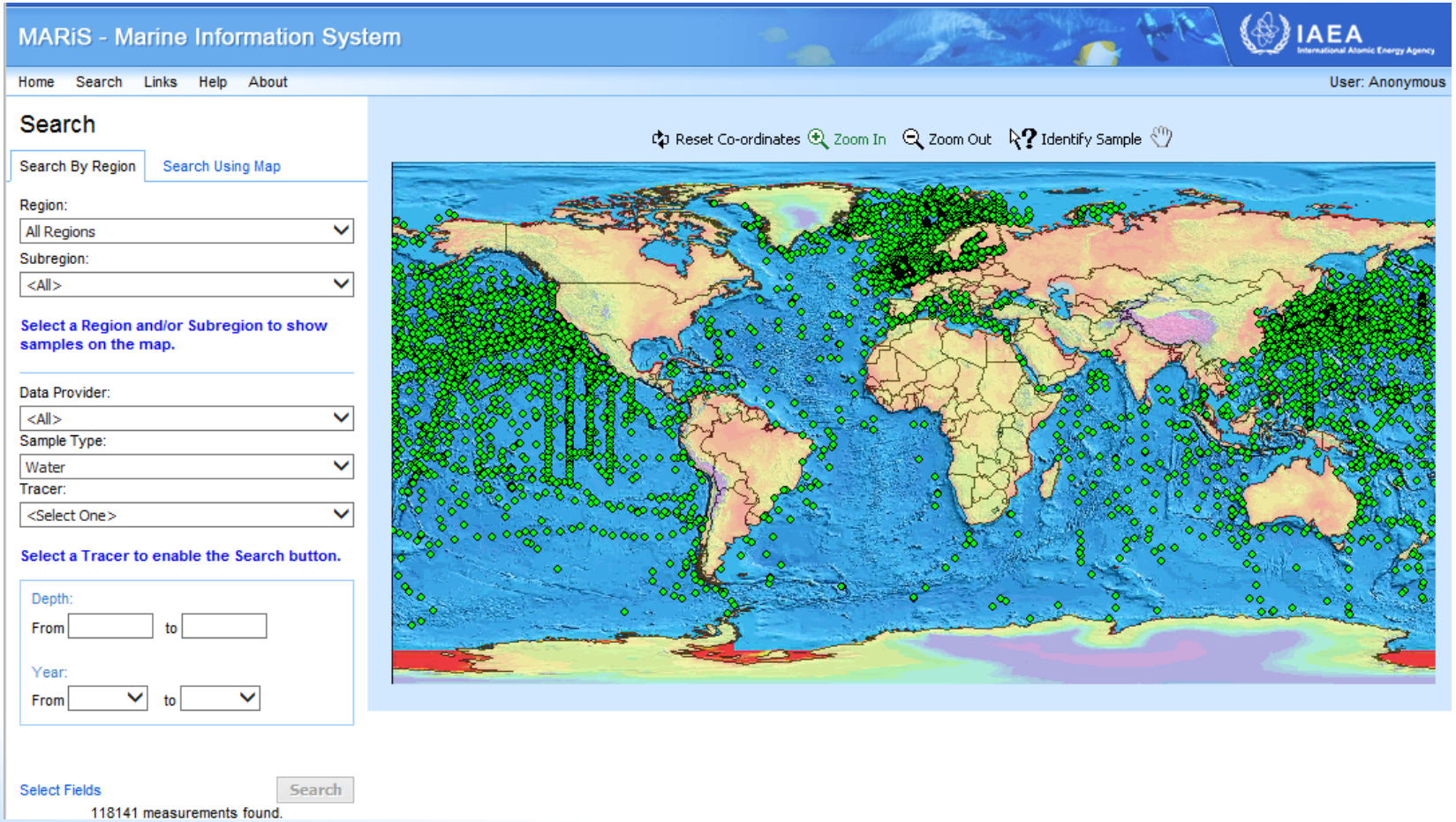
IAEA 60 Years
Atoms for Peace and Development

Overview of a marine radioactivity database and data management

Paul Morris
Department of Nuclear Sciences and Applications
Environment Laboratories, Monaco

MARiS – Marine Information System

- This presentation is orientated towards the IAEA's open access database of marine radionuclides (MARiS): maris.iaea.org



Presentation objectives

- To define database jargon:
  - data vs. datum
  - data and metadata
  - fields and records
  - datasets and databases
- To shed some light on databases and take you behind the scenes
- To distil data management to its component parts
- To demonstrate why we should care about archiving data?

Database talk outline

- Part 1 – What is data and why is metadata so important?
- Part 2 – From datasets to databases
- Part 3 – The MARiS database and data network

What is/are datum/data?

- Singular **datum** vs. plural **data**
  - Traditionally **data** is the plural of the singular **datum**
  - However, **data** is now commonly accepted as the singular and plural version of the noun
  - The internet is full of linguistic discussions about the word datum and data
  - The best approach is to be consistent in your own usage use the words **data** and **datum**

So, what is data?

- Is it:
  - Measurement?
  - Information?
  - Pattern?
  - Knowledge?

What is data?

- A single data point is a measurement of a parameter, for example:
  - The temperature of seawater: 20 °C
  - The length of a fish: 30 cm
  - The salinity of seawater: 35
  - The concentration of oxygen: 220 µmol L^{-1}
  - The mass of sediment: 3 kg
  - The activity of a radionuclide: 2 Bq m^{-3}

What is metadata?

IAEA 60 Years Atoms for Peace and Development

In data processing, metadata is definitional data that provides information about or documentation of other data managed within an application or environment.

The information contained in the metadata should be sufficient to allow another researcher to make use of your data, and, in a sense, to be able to recreate it

- So what does this mean in simple terms?

What is metadata?

- **Metadata is a description of the data**
- 20 °C – Data (measurement)
- Where was it measured? depth, lat, long
- When was it measured? Date and time
- How was it measured? Hg in glass, sensor
- Who measured it? Name, lab, contact info
- Citations to methods, standards used, ancillary measurements, general notes, cruise & station ID, etc…..

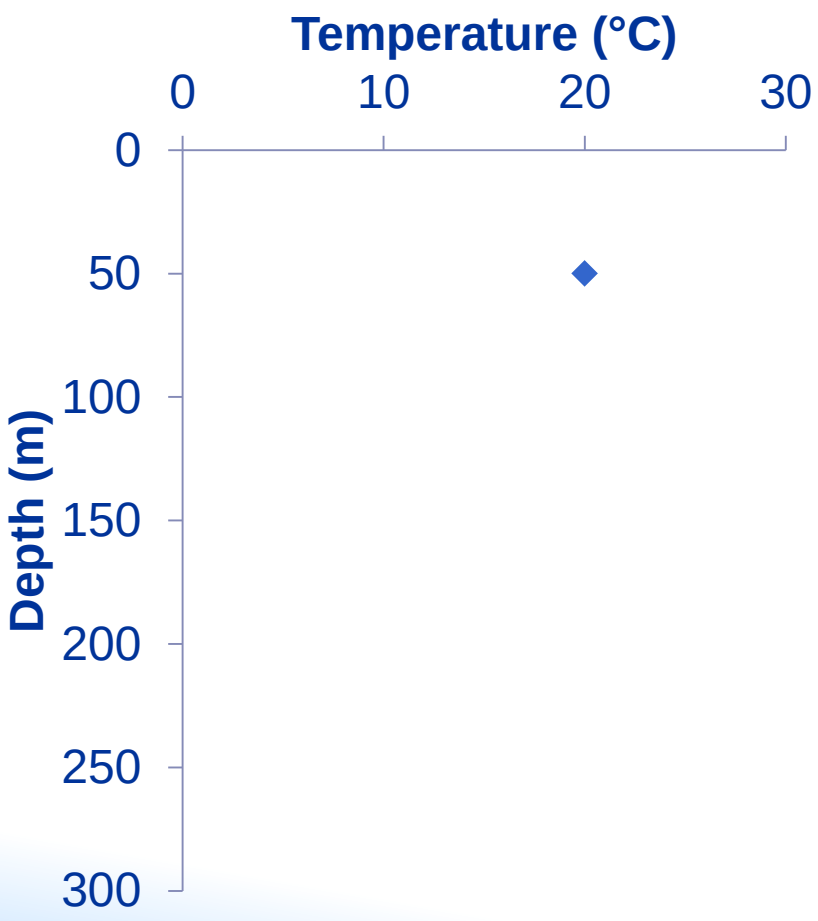
What is metadata?

- Metadata informs you about the data to create **context**
- **Without metadata**, the data is of no value because **there is no context**
  - Data originator - traceability
  - Depth
  - Position
  - Date
  - Method
  - ….

Data in the context of metadata

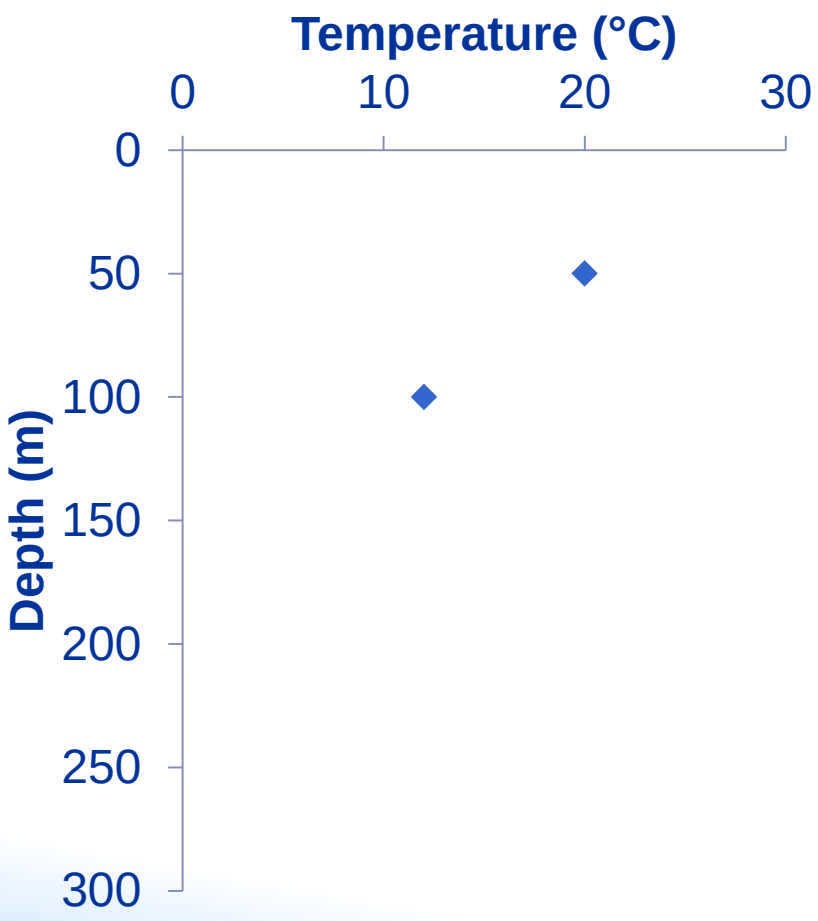
- A measurement is made: temperature is 20 °C
- So we have one data point of temperature
- **Measurement** without context

Data in the context of metadata



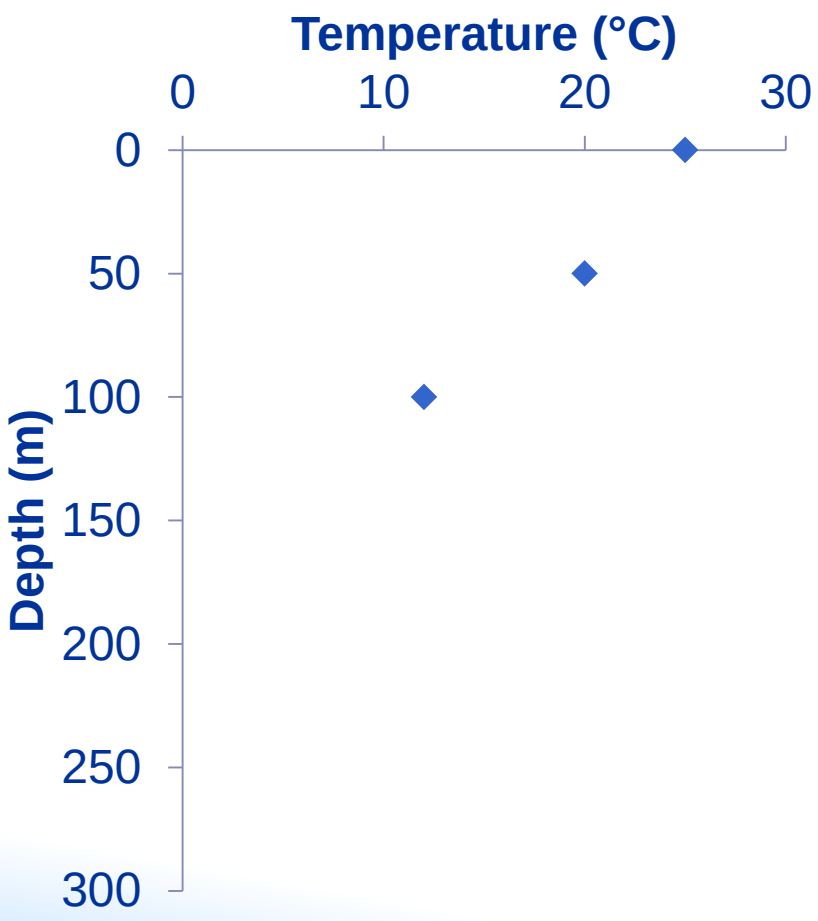
- The measurement depth is 50 m
- Metadata gives context
- The temperature at 50 m is 20 °C
- **Information** requires context (metadata)

Data in the context of metadata



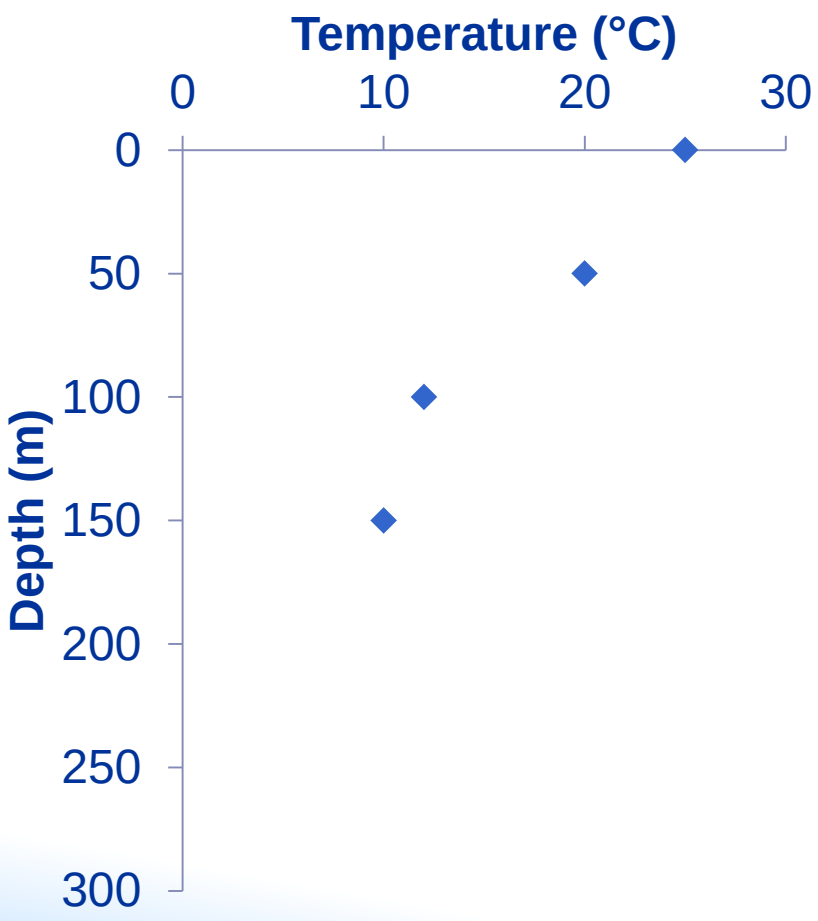
- Add more data in the context of depth to create more information

Data in the context of metadata



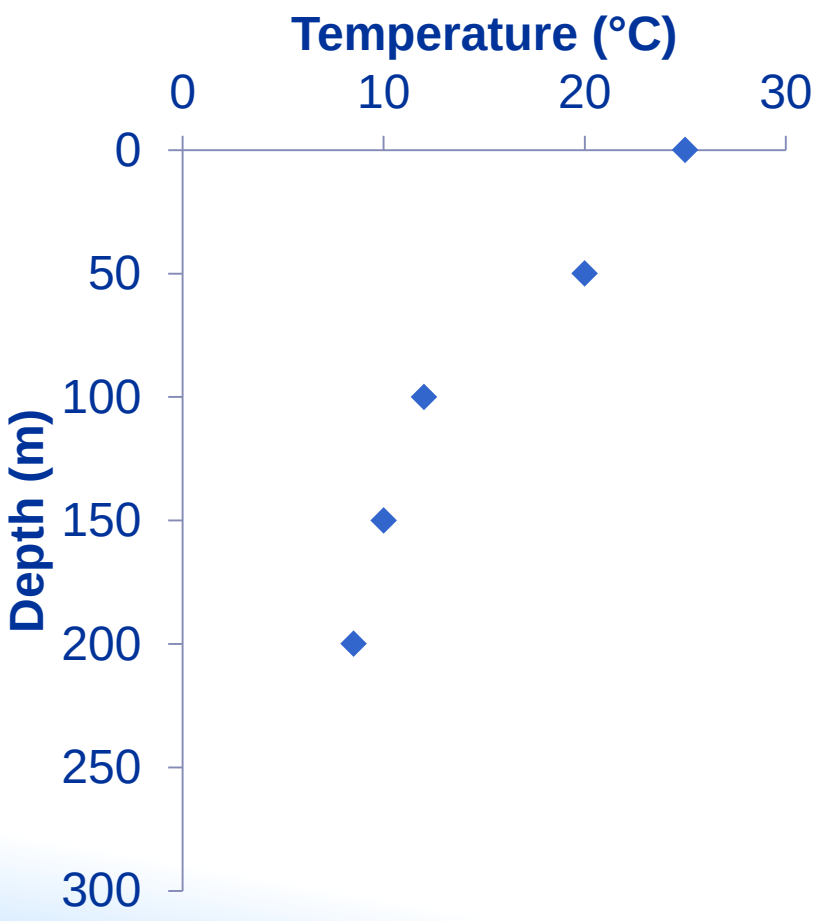
- Add more data in the context of depth to create more information

Data in the context of metadata



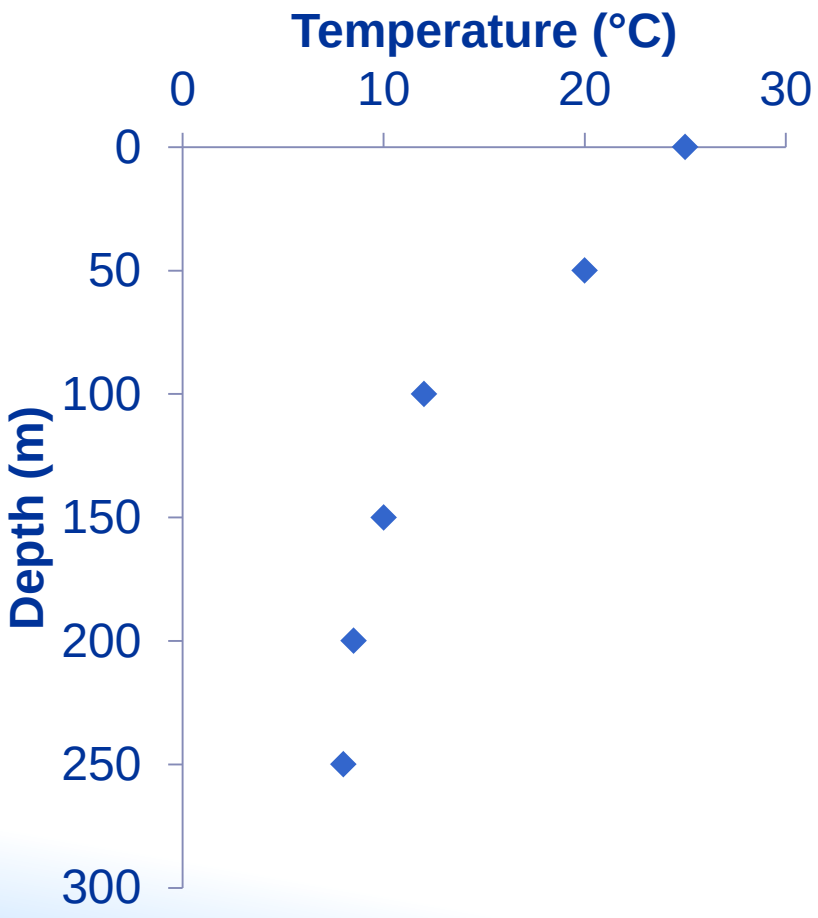
- Add more data in the context of depth to create more information
- **Patterns** emerge as more data is added

Data in the context of metadata



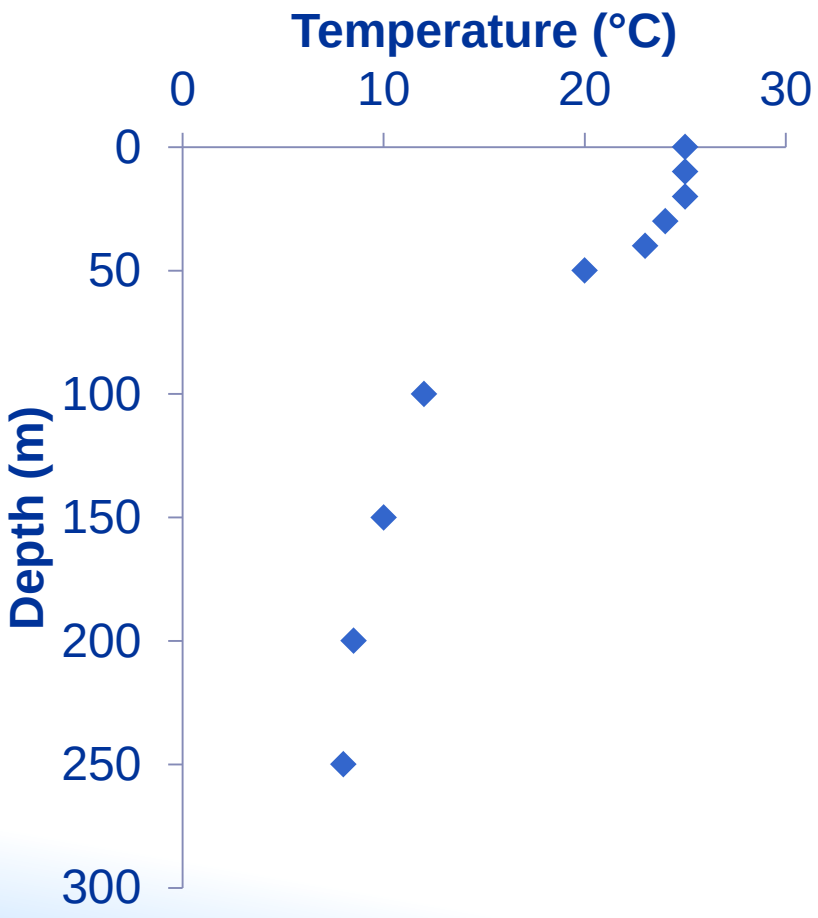
- Add more data in the context of depth to create more information
- **Patterns** emerge as more data is added

Data in the context of metadata



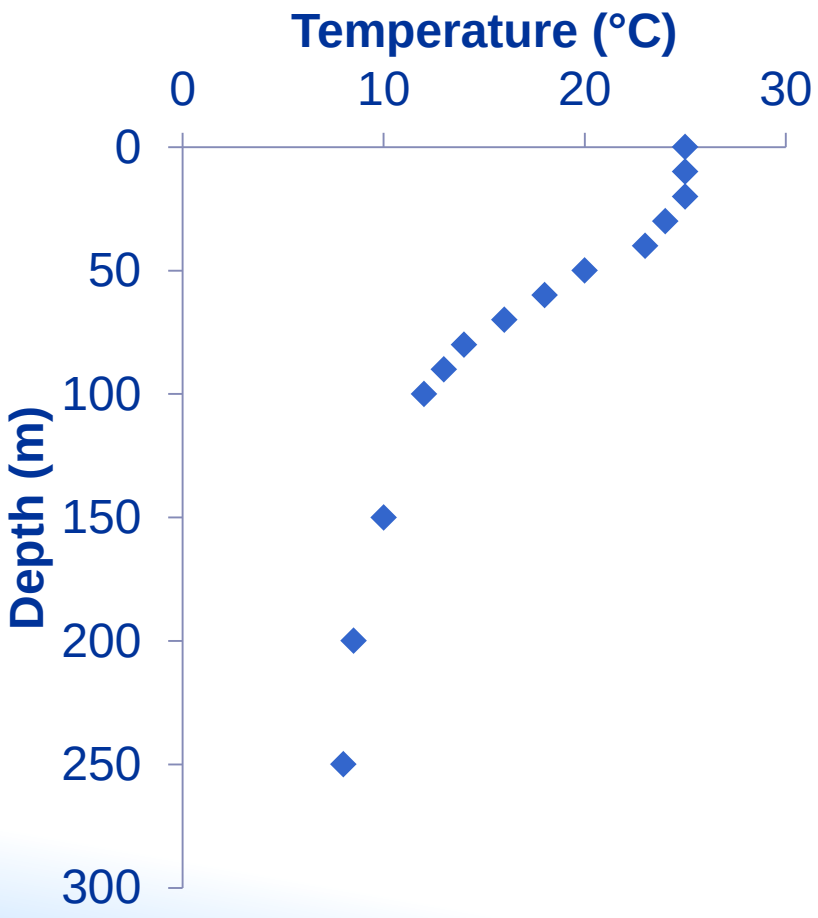
- As the resolution of the data increases, the pattern becomes more detailed

Data in the context of metadata



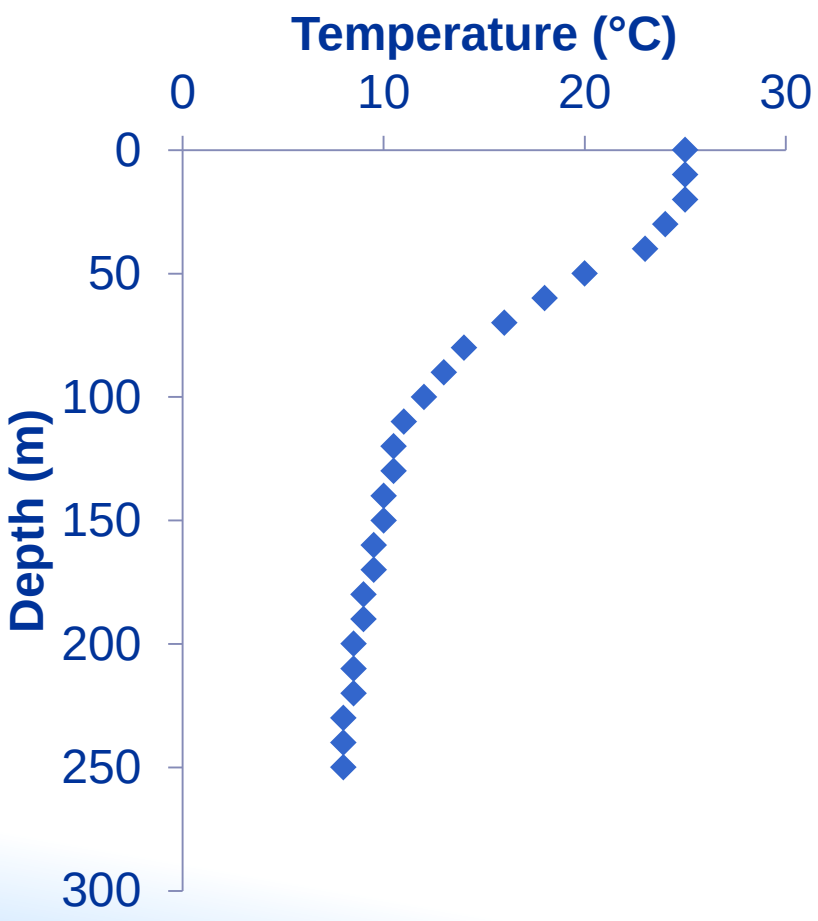
- As the resolution of the data increases, the pattern becomes more detailed

Data in the context of metadata



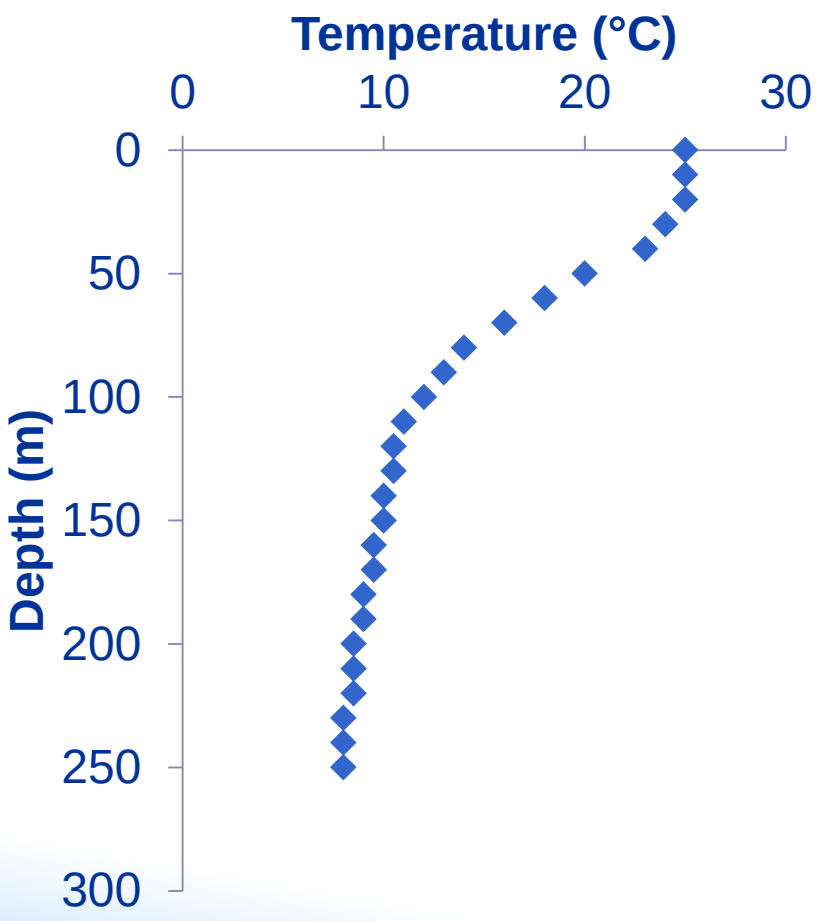
- As the resolution of the data increases, the pattern becomes more detailed

Data in the context of metadata



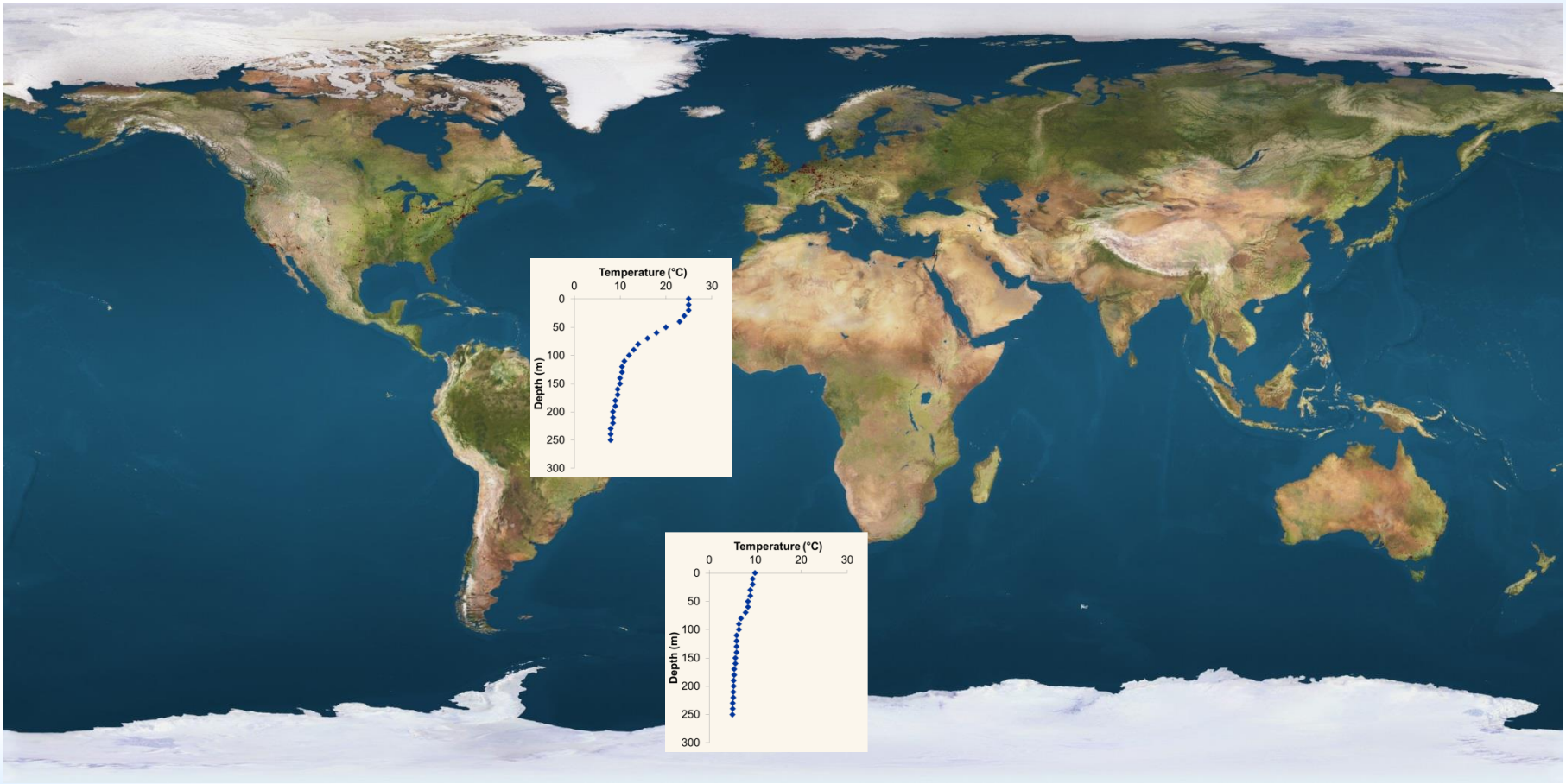
- As the resolution of the data increases, the pattern becomes more detailed
- Patterns in data provide **knowledge**
- 1 metadata parameter of depth

Data in the context of spatial metadata

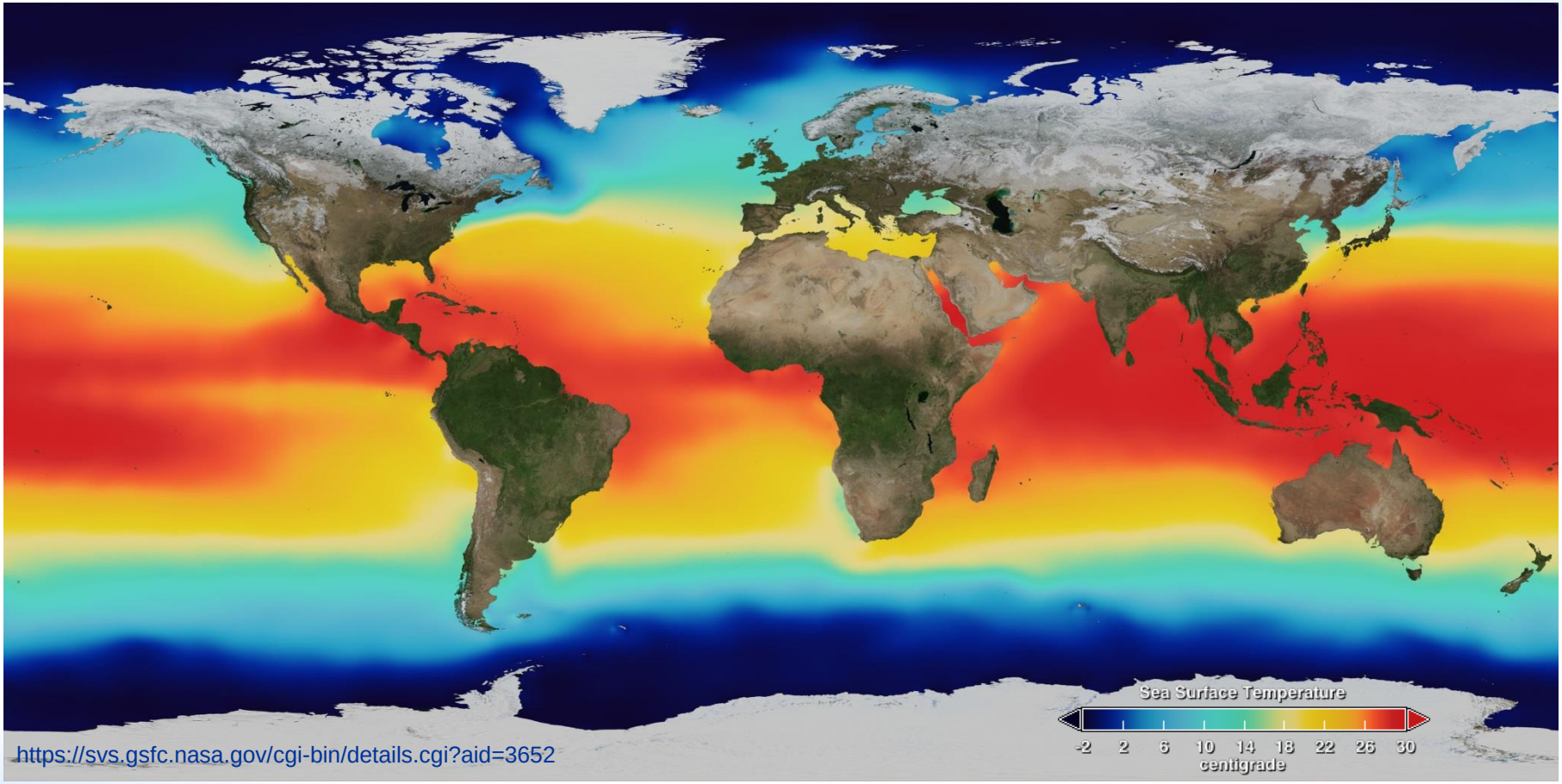


- Add spatial metadata
  - Latitude and Longitude

Data in the context of spatial metadata

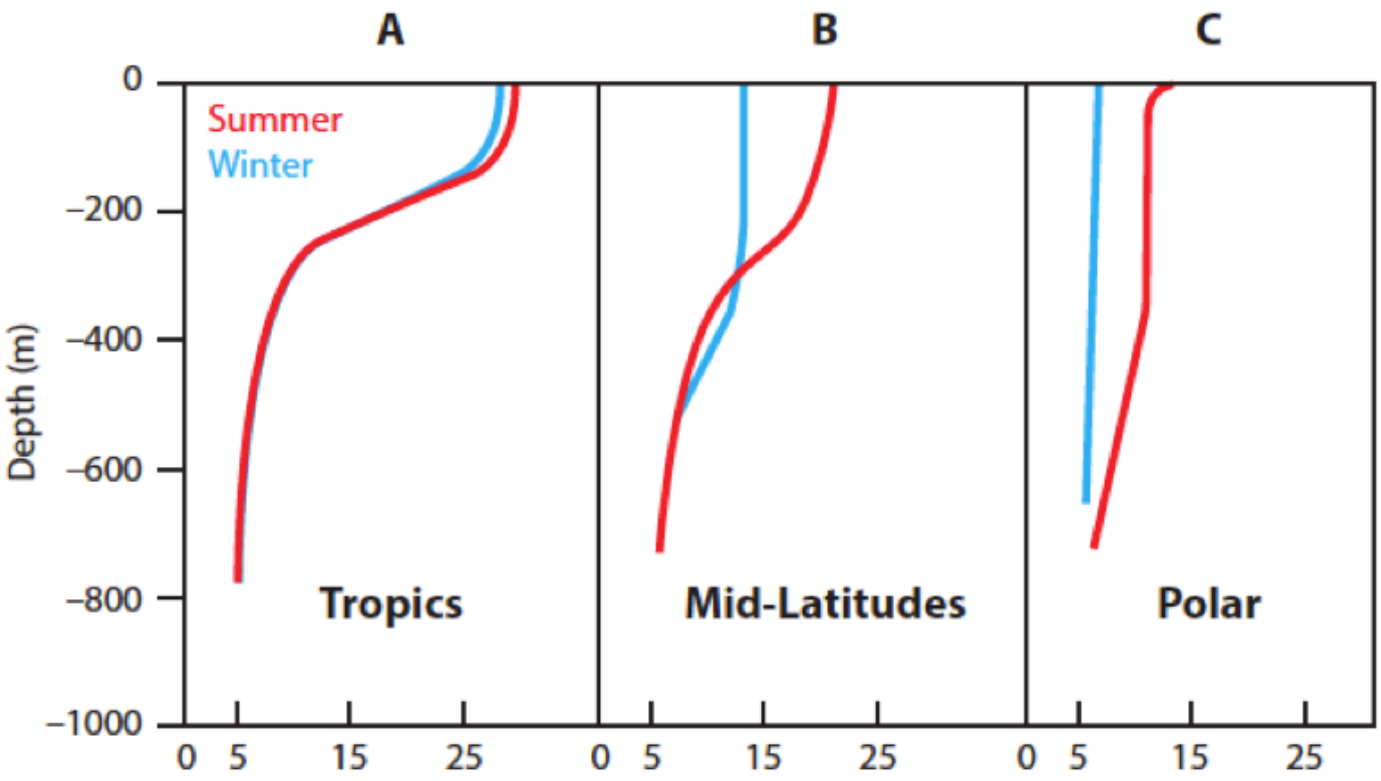


Data in the context of spatial metadata



Data in the context of temporal metadata

- Add temporal metadata: date and time



https://manoa.hawaii.edu/exploringourfluidearth
Image by Byron Inouye

Data and metadata

- Treat data & metadata with **equal importance**
- Submitted data often lacks basic metadata e.g.:
  - Latitude and longitude
  - Sampling depth
  - Sampling date
    - Dates are easily confused: 09/06/16
      - 9th June 2016 or September 6th 2016?
      - The best format for dates has lettered months: dd-mmm-yyyy
      - e.g. 09-Jun-2016

Metadata requirements

Seawater	Notes
Measurement ± Uncertainty	Data ± Uncertainty
Reference/citation	Fundamental metadata Without any one of these the data **cannot** be included in MARiS
Measurement type (=,<,ND)	
Unit	
Nuclide	
Date Sampled	
Latitude	
Longitude	
Sampling depth	
Laboratory	Highly informative metadata to include with the data
Collection method	
Sample prep method	
Counting method	
Salinity	
Temperature	
Sample lab code	Good to have and provides further useful context for data search and interpretation
Total water depth	
Station	
Cruise	
Sample volume	
Filtered	Other ancillary metadata fields exist in MARiS but are less used
Acidified	
Other fields	

Metadata requirements

Seawater	Sediment	Biota	Suspended matter	Notes
Measurement ± Uncertainty	Measurement ± Uncertainty	Measurement ± Uncertainty	Measurement ± Uncertainty	Data ± Uncertainty
Reference/citation	Reference/citation	Reference/citation	Reference/citation	Fundamental metadata Without any one of these the data **cannot** be included in MARiS
Measurement type (=,<,ND)	Measurement type (=,<,ND)	Measurement type (=,<,ND)	Measurement type (=,<,ND)	
Unit	Unit	Unit	Unit	
Nuclide	Nuclide	Nuclide	Nuclide	
Date Sampled	Date Sampled	Date Sampled	Date Sampled	
Latitude	Latitude	Latitude	Latitude	
Longitude	Longitude	Longitude	Longitude	
Sampling Depth	Upper and lower slice	Species	Sampling Depth	
Laboratory	Laboratory	Body part	Laboratory	Highly informative metadata to include with the data
Collection method	Collection method	Laboratory	Collection method	
Sample prep method	Sample prep method	Collection method	Sample prep method	
Counting method	Counting method	Sample prep method	Counting method	
Salinity	Sediment type	Counting method	Sample lab code	Good to have and provides further useful context for data search and interpretation
Temperature	Weights (wet/dry/ash)	Sampling Depth	Total water depth	
Sample lab code	Sample lab code	Sample lab code	Station	
Total water depth	Total water depth	Total water depth	Cruise	
Station	Station	Station	Drying method	
Cruise	Cruise	Cruise	Organic content	
Sample volume	Drying method			Other ancillary metadata fields exist in MARiS but are less used
Filtered	Organic content			
Acidified	Oxic/Anoxic			

From a dataset to a database

Casacuberta et al. (2013) Biogeosciences, 10, 3649–3659

Station	Sample code	Collection date	Depth(m)	Longtitude	Latitude	90Sr(Bq/m3)	90Sr Error	89Sr(Bq/m3)	89Sr Error
19	Sr-20	12.06.2011	200	142.9980	37.4940	0.8	0.2		
19	Sr-21	12.06.2011	100	142.9980	37.4940	1.7	0.2		
19	Sr-22	12.06.2011	50	142.9980	37.4940	4.7	0.4		
19	Sr-23	12.06.2011	20	142.9980	37.4940	8.1	0.4	25	4
20	Sr-24	13.06.2011	2	143.0016	37.6000	1.7	0.2		
21	Sr-25	13.06.2011	2	142.2947	37.3008	7.4	0.4	22	9
23	Sr-32	14.06.2011	100	142.0050	37.4810	4.8	0.7		
23	Sr-33	14.06.2011	50	142.0050	37.4810	34	2	137	43
23	Sr-34	14.06.2011	20	142.0050	37.4810	37	3	132	34
23	Sr-35	14.06.2011	10	142.0050	37.4810	28	1	65	21
24	Sr-36	14.06.2011	2	141.5984	38.0004	2.5	0.2		
25	Sr-37	15.06.2011	2	141.2673	37.3187	17	1	29	13

- Example dataset – flat single spreadsheet (Table in paper or supp. info)
- Contains all the necessary fundamental metadata
- The text of the paper contains much of the highly informative metadata
- Focus on a small number of samples

From a dataset to a database

Casacuberta et al. (2013) Biogeosciences, 10, 3649–3659

Station	Sample code	Collection date	Depth(m)	Longtitude	Latitude	90Sr(Bq/m3)	90Sr Error	89Sr(Bq/m3)	89Sr Error
23	Sr-33	14.06.2011	50	142.0050	37.4810	34	2	137	43
23	Sr-34	14.06.2011	20	142.0050	37.4810	37	3	132	34
23	Sr-35	14.06.2011	10	142.0050	37.4810	28	1	65	21

- The data and metadata can be broken into **unique** and **repeating** components
  - Red highlights unique entries
  - Green highlights repeated entries
- The repeated entries can be separated out from the dataset, placed in a new table and entered only once

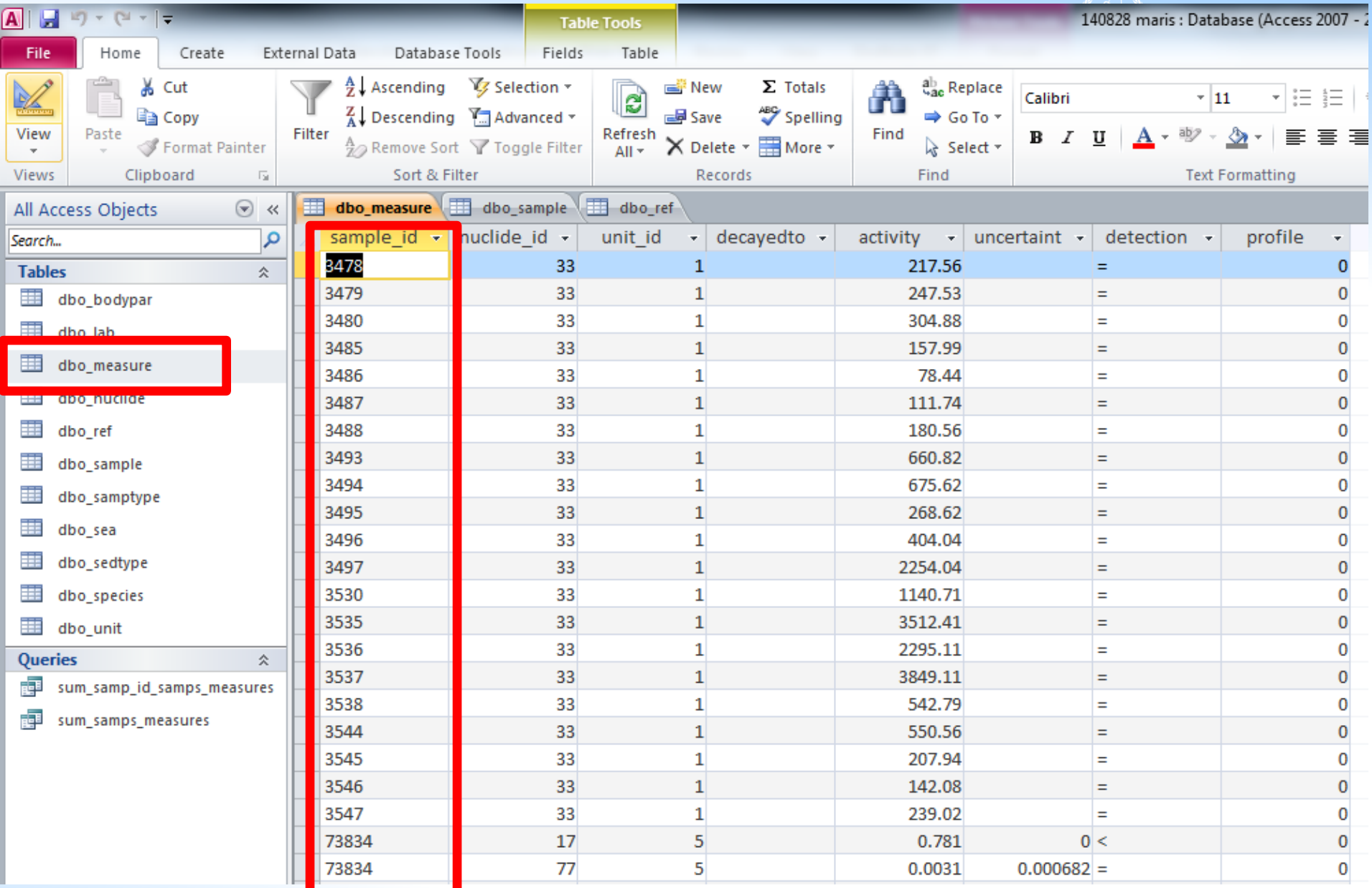
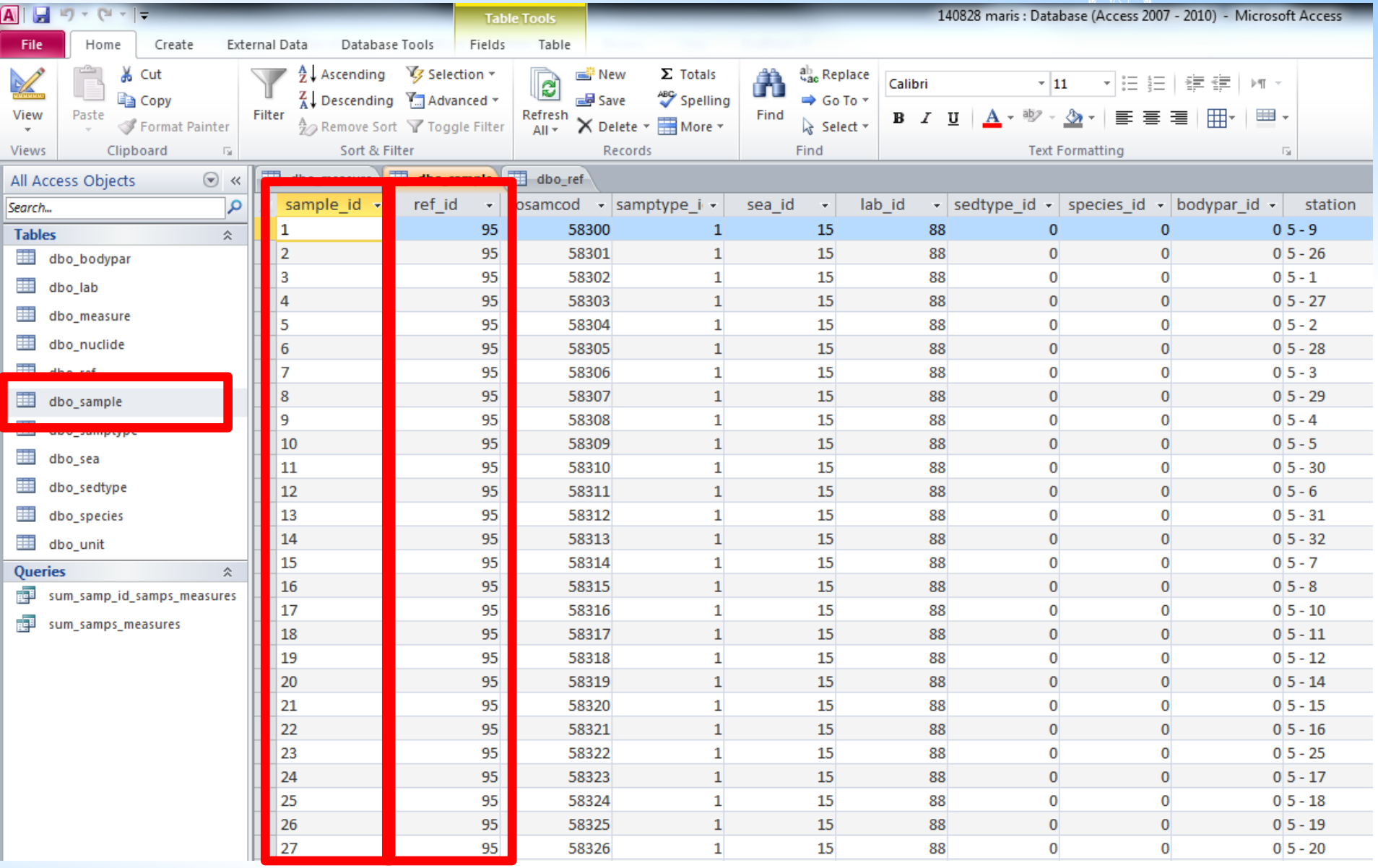


Table Tools
140828 maris : Database (Access 2007 -
File
Home
Create
External Data
Database Tools
Fields
Table
View
Views
Paste
Cut
Copy
Format Painter
Clipboard
Filter
Ascending
Descending
Remove Sort
Selection
Advanced
Toggle Filter
Sort & Filter
Refresh All
New
Save
Delete
Totals
Spelling
More
Records
Find
Replace
Go To
Select
Find
Calibri
11
Text Formatting
All Access Objects
Search...
Tables
dbo_bodypar
dbo_lab
dbo_measure
dbo_nuclide
dbo_ref
dbo_sample
dbo_samptype
dbo_sea
dbo_sedtype
dbo_species
dbo_unit
Queries
sum_samp_id_samps_measures
sum_samps_measures
dbo_measure
dbo_sample
dbo_ref
sample_id
nuclide_id
unit_id
decayedto
activity
uncertaint
detection
profile
3478 33 1 217.56 = 0
3479 33 1 247.53 = 0
3480 33 1 304.88 = 0
3485 33 1 157.99 = 0
3486 33 1 78.44 = 0
3487 33 1 111.74 = 0
3488 33 1 180.56 = 0
3493 33 1 660.82 = 0
3494 33 1 675.62 = 0
3495 33 1 268.62 = 0
3496 33 1 404.04 = 0
3497 33 1 2254.04 = 0
3530 33 1 1140.71 = 0
3535 33 1 3512.41 = 0
3536 33 1 2295.11 = 0
3537 33 1 3849.11 = 0
3538 33 1 542.79 = 0
3544 33 1 550.56 = 0
3545 33 1 207.94 = 0
3546 33 1 142.08 = 0
3547 33 1 239.02 = 0
73834 17 5 0.781 0 < 0
73834 77 5 0.0031 0.000682 = 0


140828 maris : Database (Access 2007 - 2010) - Microsoft Access
Table Tools
File
Home
Create
External Data
Database Tools
Fields
Table
View
Paste
Cut
Copy
Format Painter
Filter
Ascending
Descending
Remove Sort
Selection
Advanced
Toggle Filter
Refresh All
New
Save
Delete
Totals
Spelling
More
Find
Replace
Go To
Select
Calibri
11
Views
Clipboard
Sort & Filter
Records
Find
Text Formatting
All Access Objects
Search...
Tables
dbo_bodypar
dbo_lab
dbo_measure
dbo_nuclide
dbo_sample
dbo_sea
dbo_sedtype
dbo_species
dbo_unit
Queries
sum_samp_id_samps_measures
sum_samps_measures
dbo_ref
sample_id
ref_id
osamcod
samptype_i
sea_id
lab_id
sedtype_id
species_id
bodypar_id
station
1 95 58300 1 15 88 0 0 0 5 - 9
2 95 58301 1 15 88 0 0 0 5 - 26
3 95 58302 1 15 88 0 0 0 5 - 1
4 95 58303 1 15 88 0 0 0 5 - 27
5 95 58304 1 15 88 0 0 0 5 - 2
6 95 58305 1 15 88 0 0 0 5 - 28
7 95 58306 1 15 88 0 0 0 5 - 3
8 95 58307 1 15 88 0 0 0 5 - 29
9 95 58308 1 15 88 0 0 0 5 - 4
10 95 58309 1 15 88 0 0 0 5 - 5
11 95 58310 1 15 88 0 0 0 5 - 30
12 95 58311 1 15 88 0 0 0 5 - 6
13 95 58312 1 15 88 0 0 0 5 - 31
14 95 58313 1 15 88 0 0 0 5 - 32
15 95 58314 1 15 88 0 0 0 5 - 7
16 95 58315 1 15 88 0 0 0 5 - 8
17 95 58316 1 15 88 0 0 0 5 - 10
18 95 58317 1 15 88 0 0 0 5 - 11
19 95 58318 1 15 88 0 0 0 5 - 12
20 95 58319 1 15 88 0 0 0 5 - 14
21 95 58320 1 15 88 0 0 0 5 - 15
22 95 58321 1 15 88 0 0 0 5 - 16
23 95 58322 1 15 88 0 0 0 5 - 25
24 95 58323 1 15 88 0 0 0 5 - 17
25 95 58324 1 15 88 0 0 0 5 - 18
26 95 58325 1 15 88 0 0 0 5 - 19
27 95 58326 1 15 88 0 0 0 5 - 20

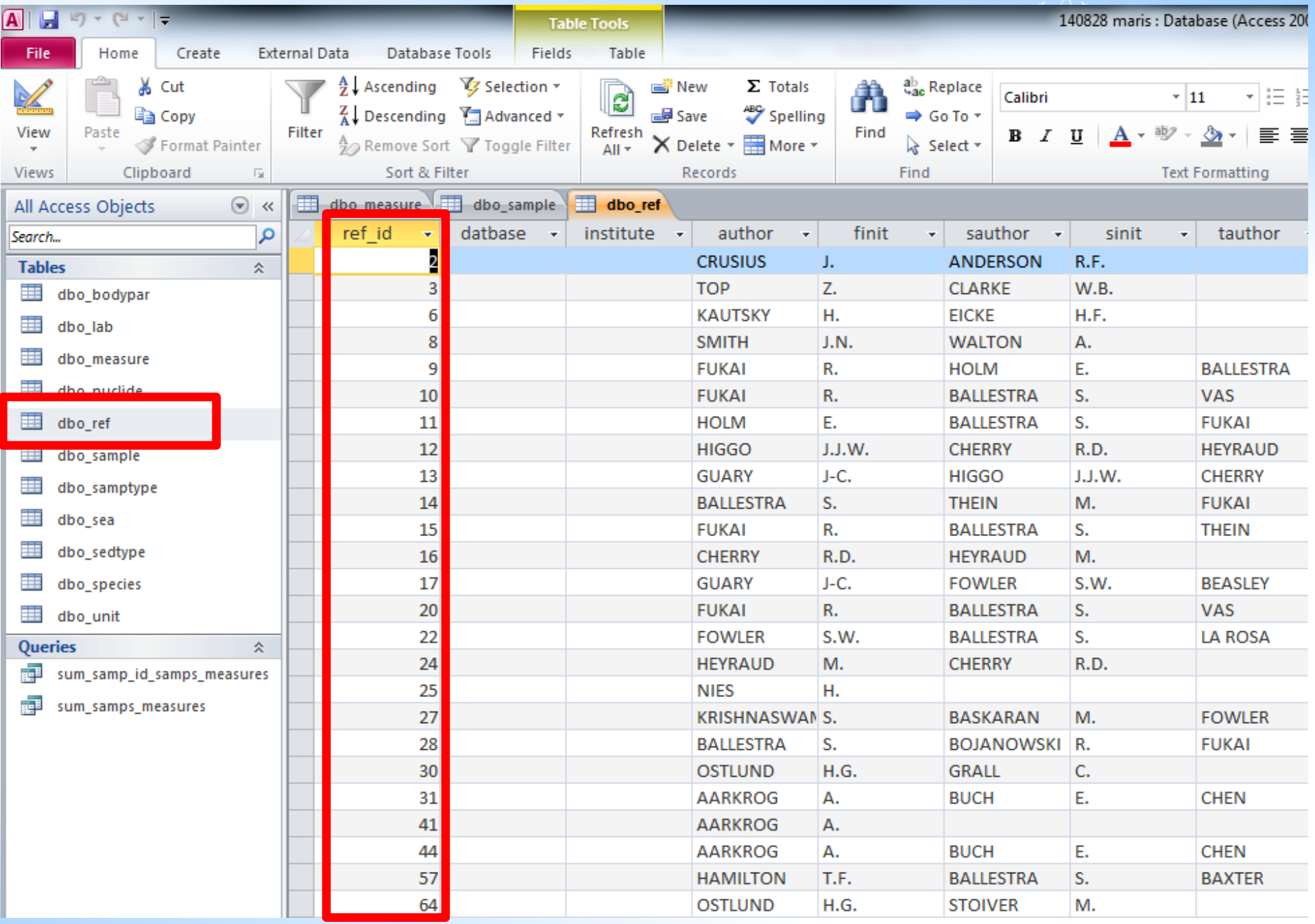


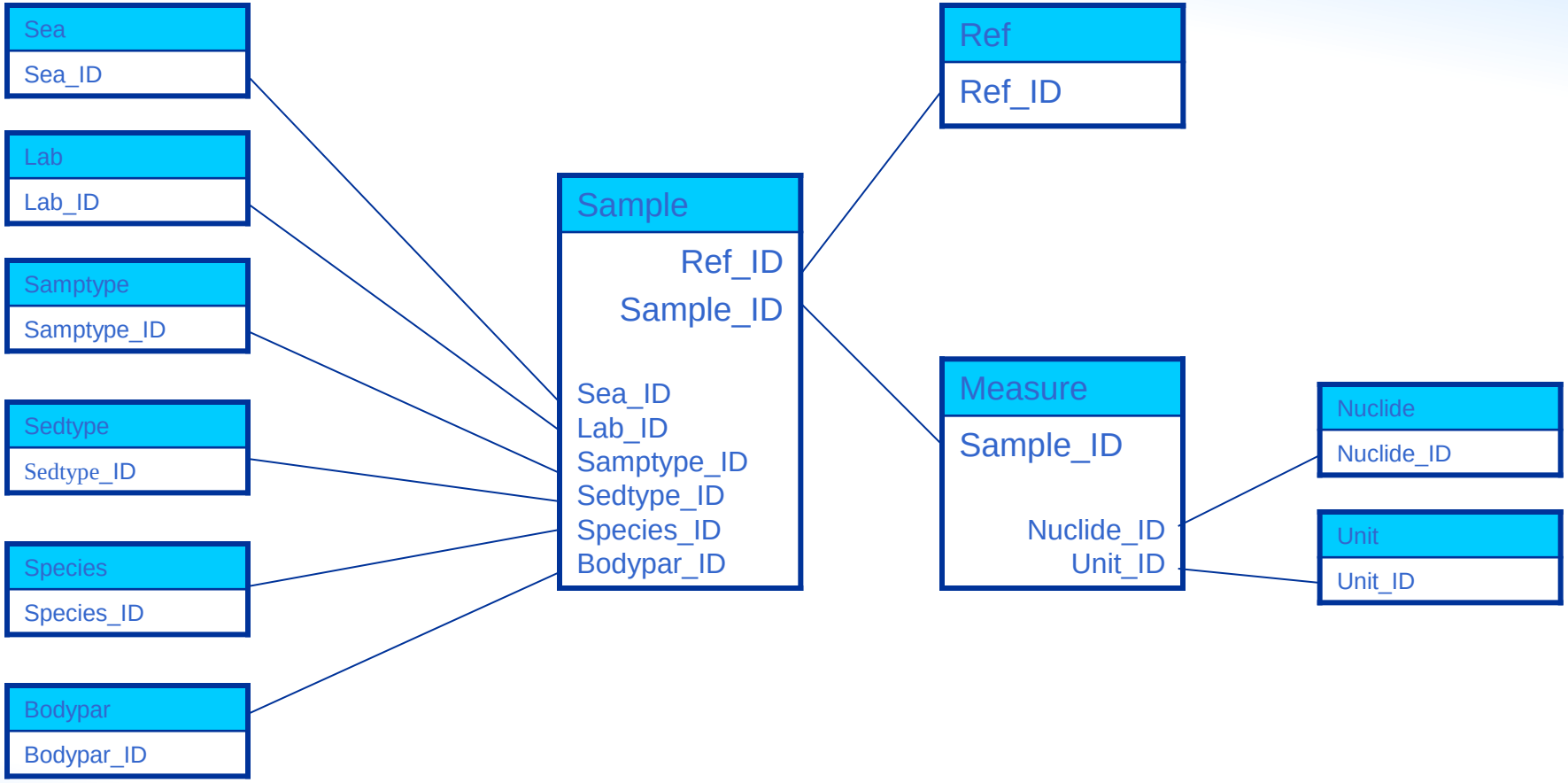
Table Tools
140828 maris : Database (Access 200
File
Home
Create
External Data
Database Tools
Fields
Table
View
Views
Paste
Cut
Copy
Format Painter
Clipboard
Filter
Ascending
Descending
Remove Sort
Selection
Advanced
Toggle Filter
Sort & Filter
Refresh All
New
Save
Delete
Totals
Spelling
More
Records
Find
Replace
Go To
Select
Find
Calibri
11
Text Formatting
All Access Objects
Search...
Tables
dbo_bodypar
dbo_lab
dbo_measure
dbo_nuclide
dbo_ref
dbo_sample
dbo_samptype
dbo_sea
dbo_sedtype
dbo_species
dbo_unit
Queries
sum_samp_id_samps_measures
sum_samps_measures
dbo_measure
dbo_sample
dbo_ref
ref_id
datbase
institute
author
finit
sauthor
sinit
tauthor
2 CRUSIUS J. ANDERSON R.F.
3 TOP Z. CLARKE W.B.
6 KAUTSKY H. EICKE H.F.
8 SMITH J.N. WALTON A.
9 FUKAI R. HOLM E. BALLESTRA
10 FUKAI R. BALLESTRA S. VAS
11 HOLM E. BALLESTRA S. FUKAI
12 HIGGO J.J.W. CHERRY R.D. HEYRAUD
13 GUARY J-C. HIGGO J.J.W. CHERRY
14 BALLESTRA S. THEIN M. FUKAI
15 FUKAI R. BALLESTRA S. THEIN
16 CHERRY R.D. HEYRAUD M.
17 GUARY J-C. FOWLER S.W. BEASLEY
20 FUKAI R. BALLESTRA S. VAS
22 FOWLER S.W. BALLESTRA S. LA ROSA
24 HEYRAUD M. CHERRY R.D.
25 NIES H.
27 KRISHNASWAI S. BASKARAN M. FOWLER
28 BALLESTRA S. BOJANOWSKI R. FUKAI
30 OSTLUND H.G. GRALL C.
31 AARKROG A. BUCH E. CHEN
41 AARKROG A.
44 AARKROG A. BUCH E. CHEN
57 HAMILTON T.F. BALLESTRA S. BAXTER
64 OSTLUND H.G. STOIVER M.

From a dataset to a database

- Why make the effort reorganise the dataset?
  - Removes repetition – more efficient
  - Easier to manage – update one record instead of many
  - Easier to search for data & metadata

- But, the links between the separated records have to be re-established or related
  - This forms the basis of a **relational database**

MARiS – schema



MARiS – Main tables

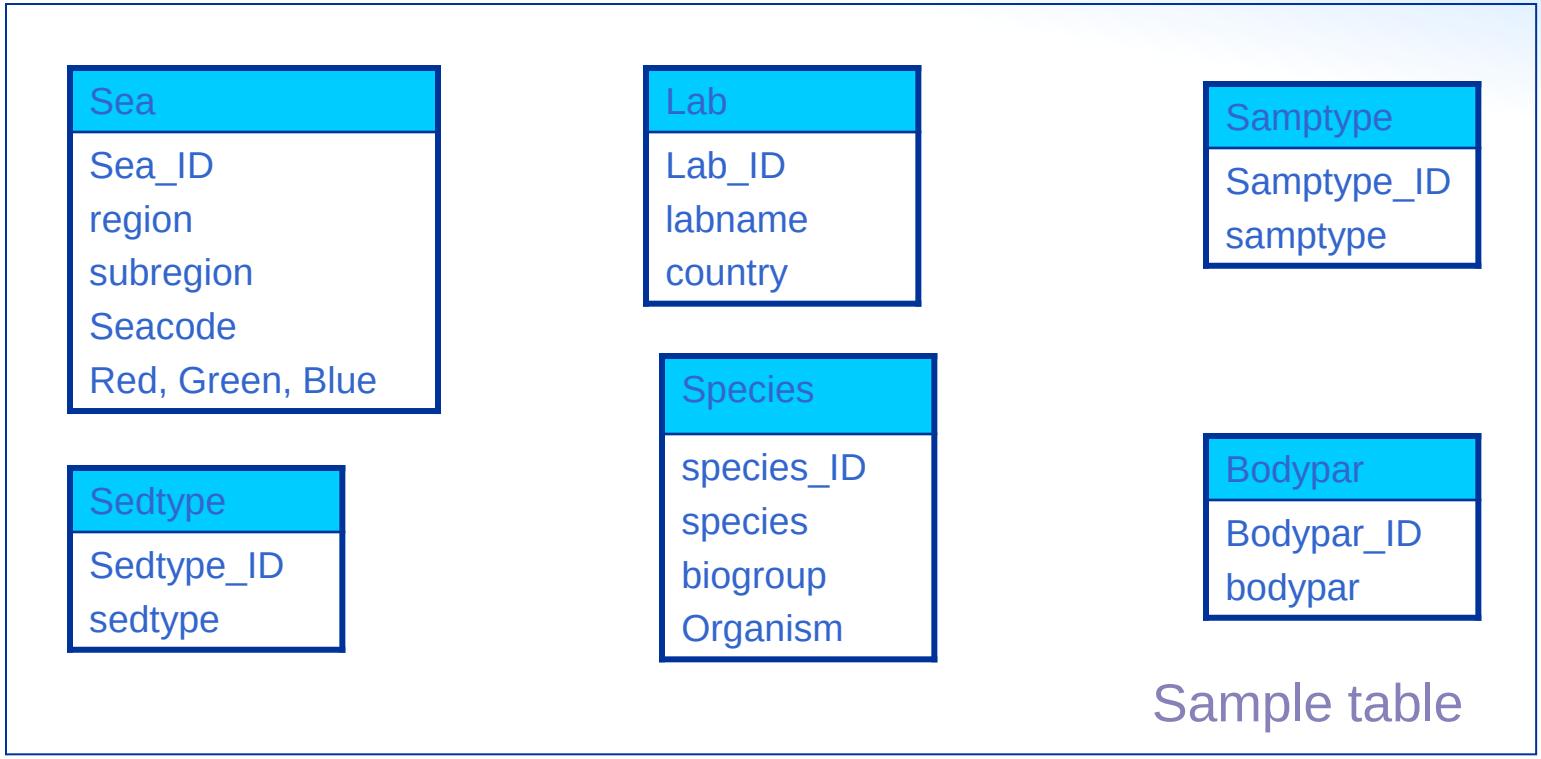
IAEA 60 Years Atoms for Peace and Development

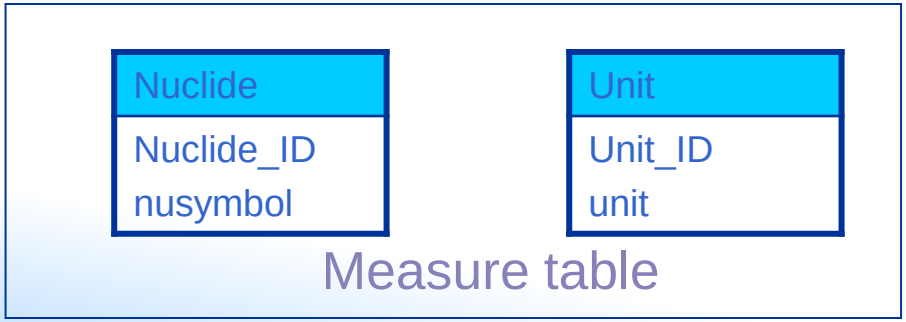
Ref
ref_id
datbase
institute
author
finit
yearpub
jname
jvol
fpage
bookt
bookp
title

Sample		
sample_id	station	totdepth
	latchar	sampdepth
samptype_id	lat	salinity
ref_id	latd	temperatur
lab_id	latm	filtered
sea_id	lats	oxygen
sedtype_id	longchar	sliceup
species_id	long	slicedown
bodypar_id	longd	plannet
	longm	begperiod
	longs	endperiod
	longitude	
	latitude	

Measure
sample_id
nuclide_id
unit_id
profile
activity
uncertaint
detection

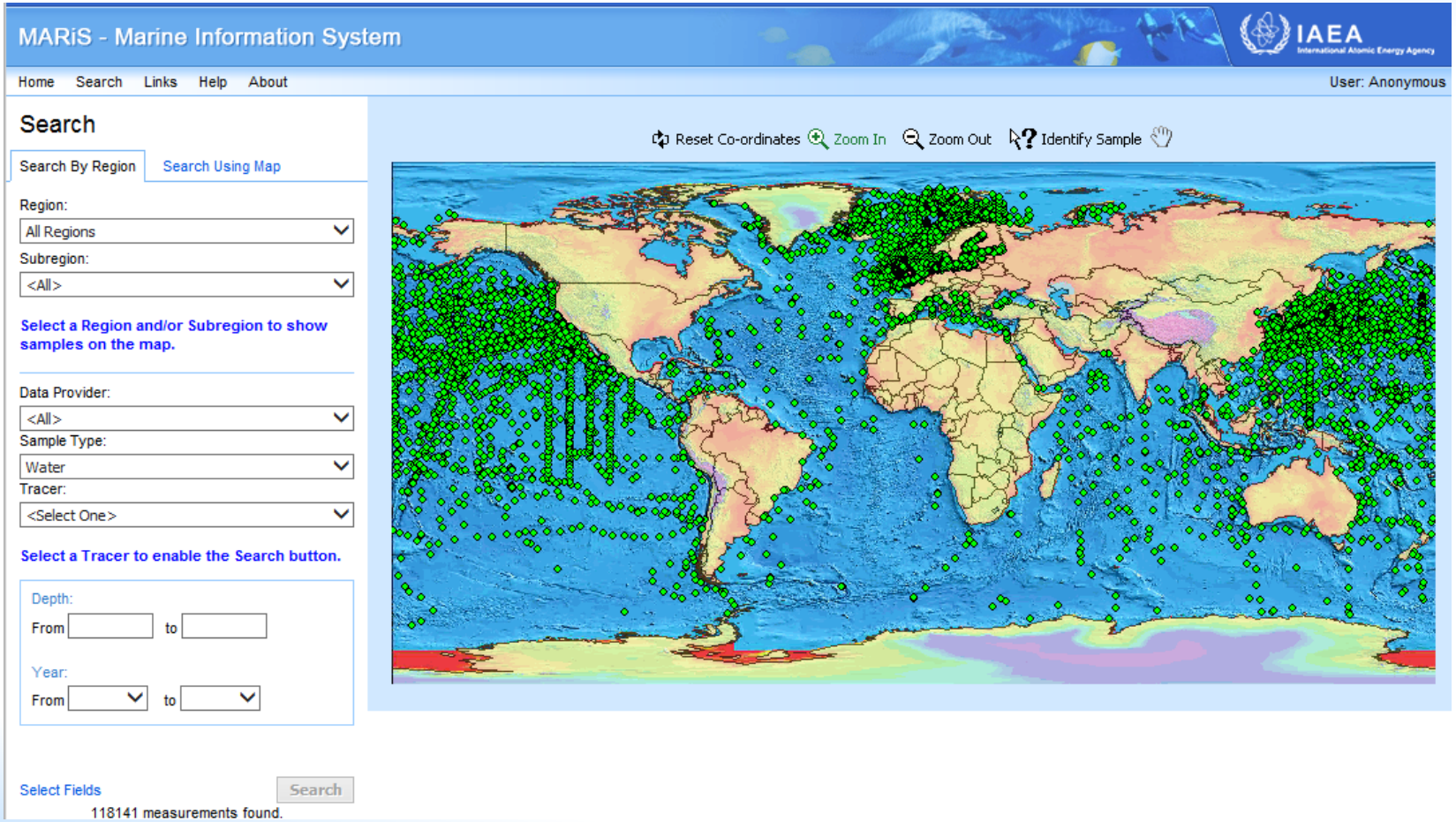
MARiS – lookup tables



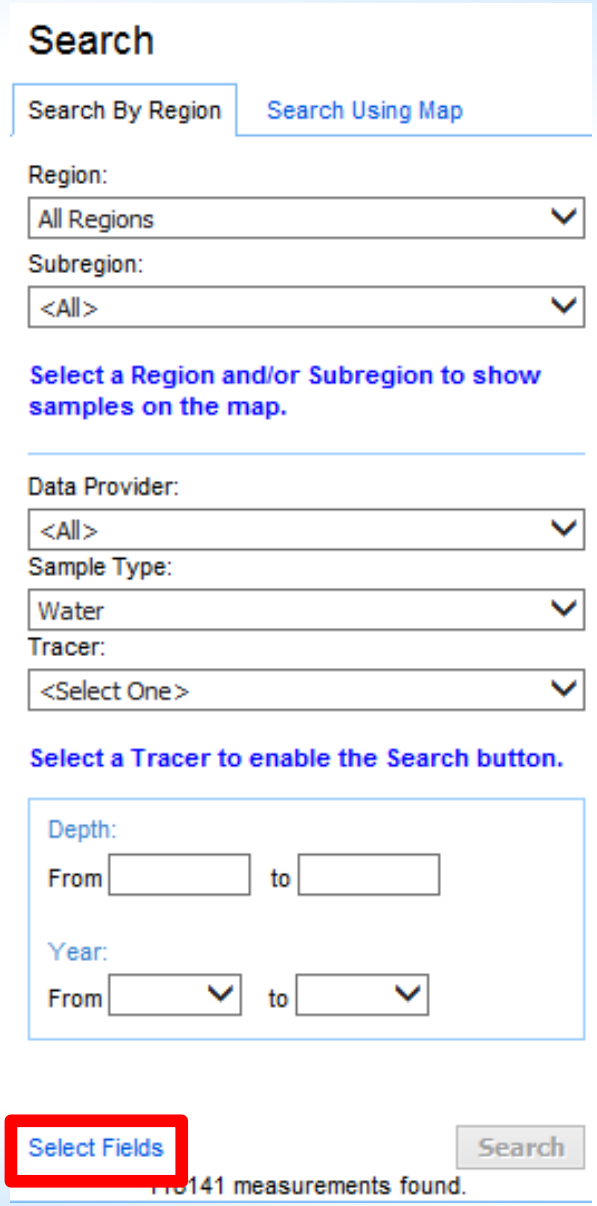


MARiS – Marine Information System

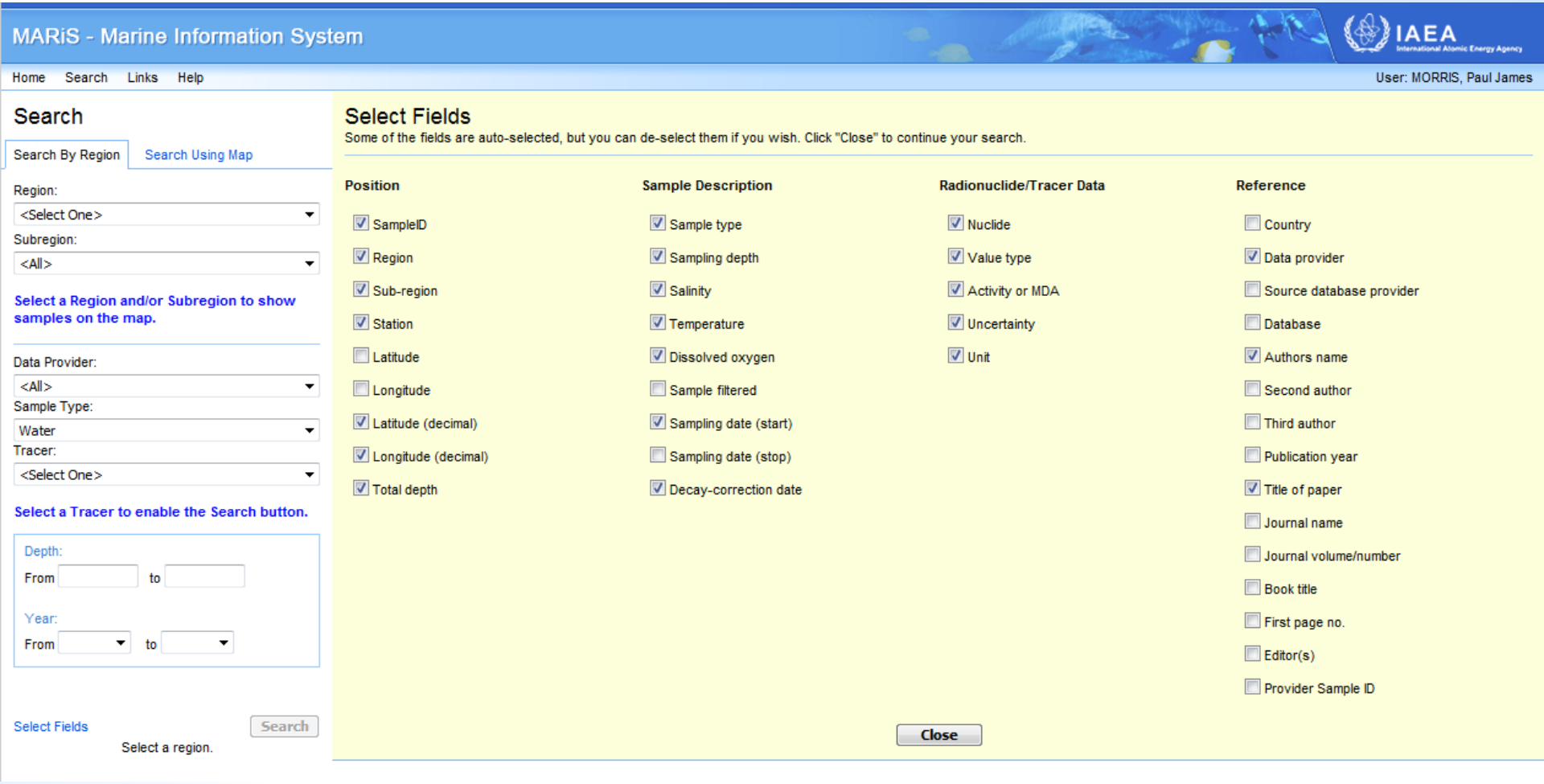
- Publically accessible database of marine radionuclides: maris.iaea.org



MARiS – Select search criteria



MARiS – Select fields



MARiS – Select fields

IAEA 60 Years Atoms for Peace and Development

Search Results - Returned 54269 Records

Export to CSV | Back to Search

SampleID	Region	Sub-region	Station	Latitude (decimal)	Longitude (decimal)	Total depth	Sample type	Sampling depth	Sampling date (start)	Decay-correction date	Nuclide	Value type	Activity or MDA	Uncertainty	Unit	Country	Data provider
1	North Atlantic Ocean	North Sea	5 - 9	52.05	4.15		WATE	0	1987-03-09		137Cs	=	14.4	1.584	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
2	North Atlantic Ocean	North Sea	5 - 26	52.67	4.18		WATE	0	1987-03-09		137Cs	=	10.9	1.308	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
3	North Atlantic Ocean	North Sea	5 - 1	52.92	4.68		WATE	0	1987-03-09		137Cs	=	15	1.65	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
4	North Atlantic Ocean	North Sea	5 - 27	52.77	4.00		WATE	0	1987-03-09		137Cs	=	12.6	1.512	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
5	North Atlantic Ocean	North Sea	5 - 2	52.87	4.63		WATE	0	1987-03-09		137Cs	=	12.3	1.476	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
6	North Atlantic Ocean	North Sea	5 - 28	52.85	3.82		WATE	0	1987-03-09		137Cs	=	13.4	1.474	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
7	North Atlantic Ocean	North Sea	5 - 3	52.77	4.57		WATE	0	1987-03-09		137Cs	=	14.8	1.628	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
8	North Atlantic Ocean	North Sea	5 - 29	52.93	4.08		WATE	0	1987-03-09		137Cs	=	14.3	1.573	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
9	North Atlantic Ocean	North Sea	5 - 4	52.65	4.50		WATE	0	1987-03-09		137Cs	=	15.5	1.705	Bq/m3	FRANCE	IPSN, Institut de Protection DPEI - Commissariat à l'En
10	North	North	5 - 5	52.88	4.43		WATE	0	1987-03-09		137Cs	=	14.7	1.617	Bq/m3	FRANCE	IPSN, Institut de Protection

MARiS - data

- Data originates from
  - Databases of collaborating institutes
  - Reports
  - Published scientific literature

- New data entered into the database
  - New data is entered into GLOMARD*
    - *GLOMARD (Global Marine Radioactivity Database) is the IAEA's in-house database
  - Data is checked and validated
  - Data eligible for public release is then uploaded into MARiS

- Intellectual property protected
  - Data is traceable to the originator or providing database
  - Clearance to include unpublished data is sought from the data provider
  - An acknowledgement reminder is displayed when data is downloaded

MARiS – status June 2016

- ~173,500 radionuclide activity measurements
  - 91,300 samples
  - 54 references (inc. 10 database contributions)
  - Seawater, sediment, biota and suspended matter
  - Main nuclides: ^{137}Cs, ^{134}Cs, ^{90}Sr, $^{239,240}Pu$, ^{3}H, ^{14}C
  - >60 natural and anthropogenic radionuclides

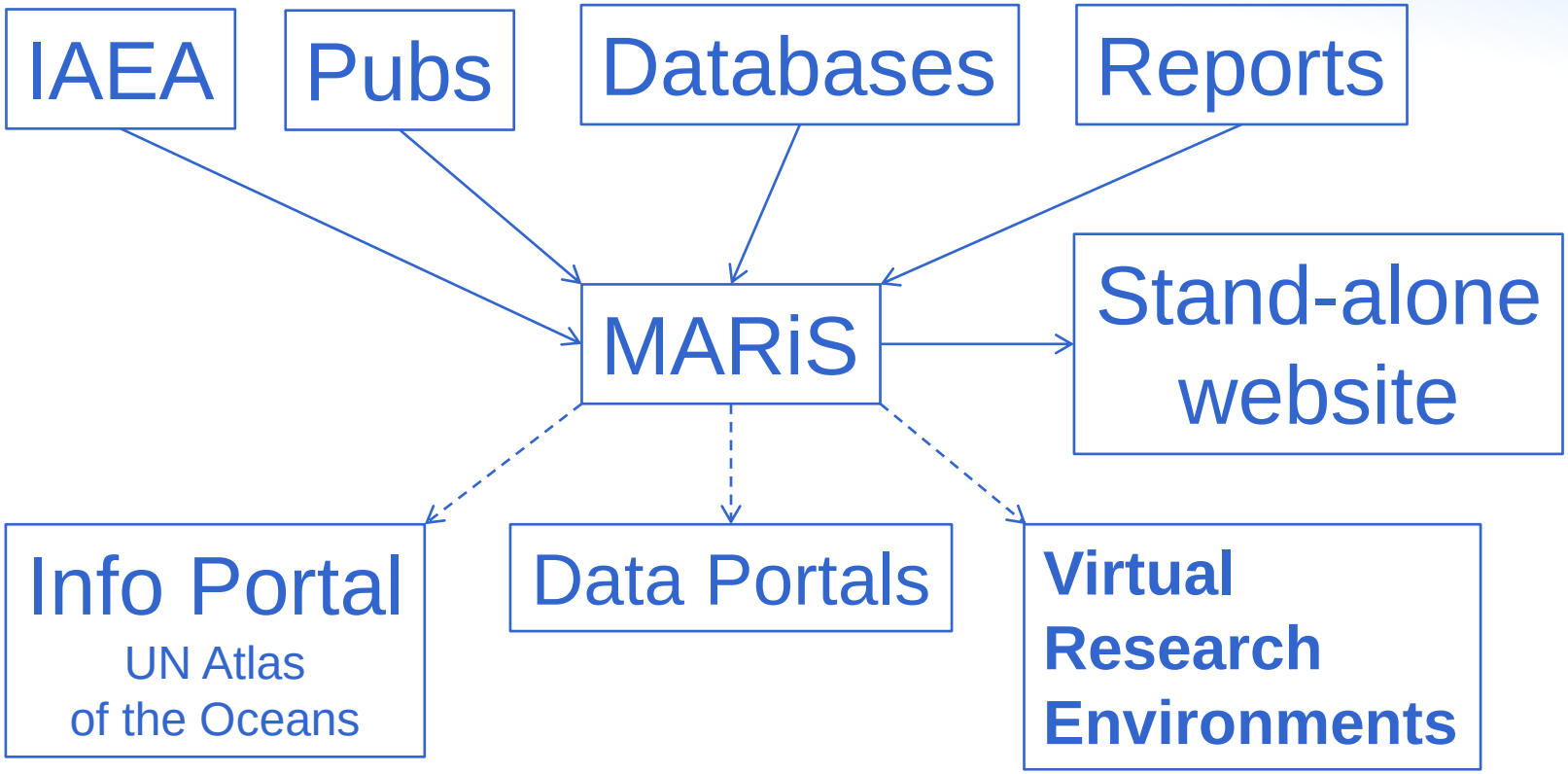
MARiS – data governance

- Data pathway or lifecycle
  - Data received from data provider/originated
  - Check to see if the data is already in MARiS
  - Import data and metadata in to MARiS
  - Document and archive the data reference
  - Serve the data to MARiS database users
    - Please cite the data provider/originator if you use data
    - A citation to MARiS is also encouraged

MARiS – data visualisation

- Tableau
  - Specialist data visualisation software under trial
  - In development and on trial for incorporation into MARiS website
  - 4 predefined visualisations for discovering what references, samples and measurements are contained in the MARiS database
  - Visit this link to see some example visualisation: http://public.tableau.com/profile/paul.mcginnity6639#!/

MARiS - data network



Links

– **MARiS** database website
– www.maris.iaea.org

– **Data visualisation** using Tableau under trial
– http://public.tableau.com/profile/paul.mcginnity6639#!/

- **User feedback** for MARiS database development
- http://goo.gl/forms/gzVsJdq1Df
- Link on MARiS homepage
- Your feedback will give us the power to develop MARiS

IAEA 60 Years
Atoms for Peace and Development

Thank you!