British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Why quality control and quality assurance is important for the legacy of GEOTRACES through its database?

Adam Leadbetter (alead@bodc.ac.uk), British Oceanographic Data Centre

British Oceanographic Data Centre

NATURAL ENVIRONMENT RESEARCH COUNCIL

Outline

- Data matter!

- Why compatible data?

- The Geotraces database

- A data intensive future…

1. Data Matter!

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Data matter!

“A scholar’s positive contribution is measured by the sum of the original data that he contributes. Hypotheses come and go but data remain.”

Santiago Ramón y Cajal
(Nobel Prize winner,1906) in
Advice to a Young Investigator (1897)

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Data matter!

"You are not finished until you have done the research, published the results, *and* published the data, receiving formal credit for everything.

***Preserve or Perish*"**

Mark Parsons
US National Snow and Ice Data Center
Data Management for the International Polar Year (2006)

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

2. Why compatible data?

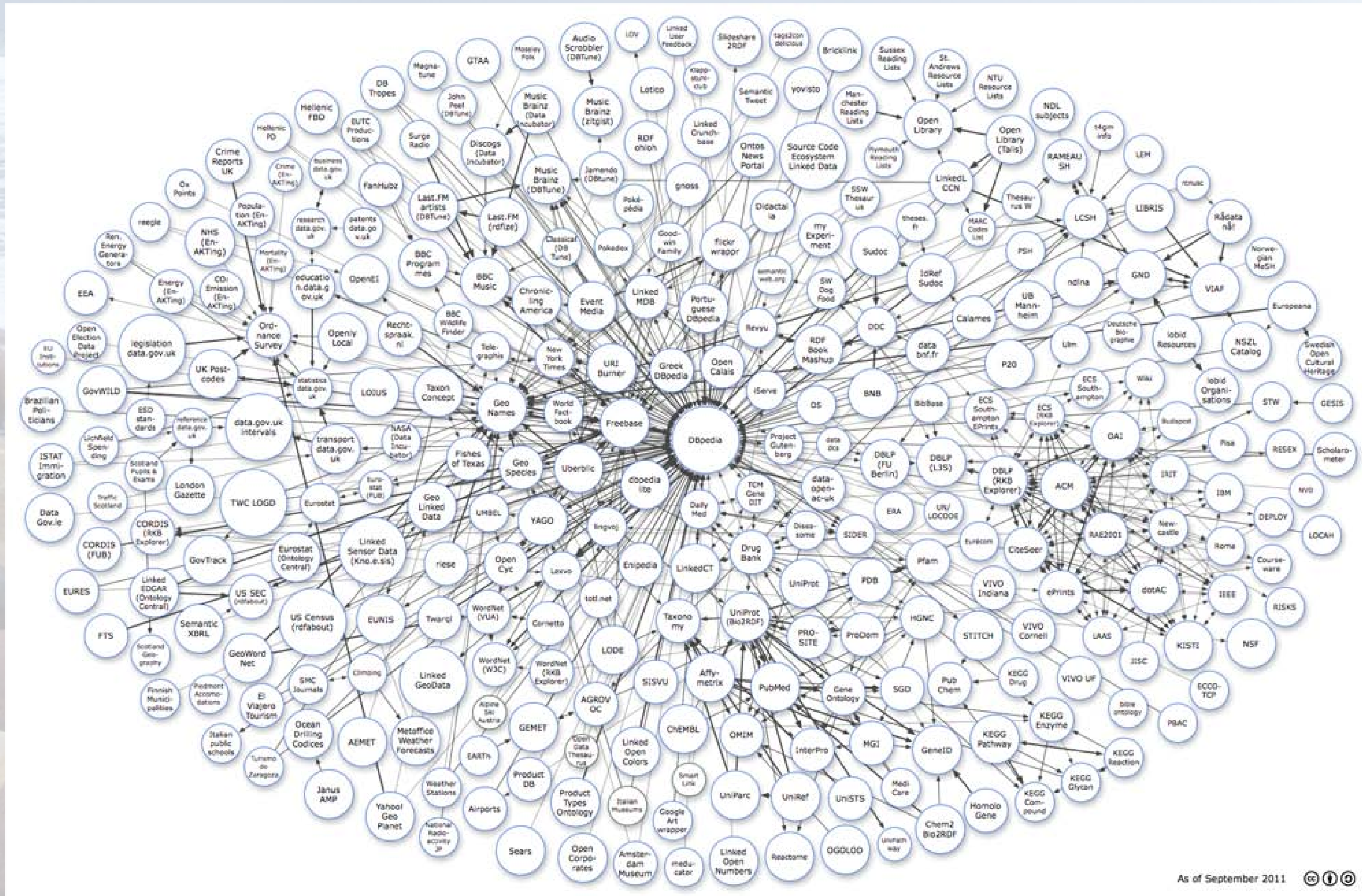
British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Why compatible data?

"If HTML and the [World Wide] Web made all the online documents look like one huge book, [compatibility] will make all the data in the world look like one huge database."

Sir Tim Berners-Lee
W3C
Weaving the Web (1999)

Why compatible data?



British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Why compatible data?

- The Linked Data cloud is built on compatible data

- Similarly, Geotraces db builds on compatible data

- How?

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Why compatible data?

- Intercalibration for QC / QA

- Only on the legacy database

- A distinction must be made where IC has not happened

- May be older “compliant data” which does not meet standards

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Why compatible data?

- Standards

  - Metadata
    - Bottle – type & make
    - Filter – type & make
    - Analytical method

  - Parameter codes

- Allows data merging & long-term data archiving

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Why compatible data?

- Merging

  - Allows easy management of “crossover stations”

  - Marked as “fixed stations” in the db

  - Enables comparison of data between cruises

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Why compatible data?

- Mantra

“To make the data accessible and usable in 5, 10, 30… years time without the need to contact the data originator.”

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

3. The GeoTraces database

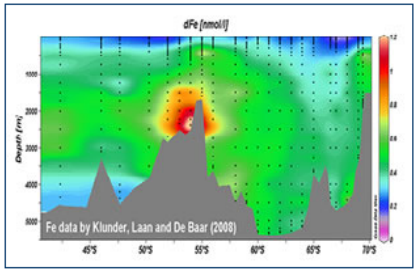
British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

GEOTRACES International Data Assembly Centre

GEOTRACES International Data Assembly Centre

GEOTRACES (www.geotraces.org) is an international programme which aims to improve our understanding of biogeochemical cycles and large-scale distribution of trace elements and their isotopes (TEIs) in the marine environment. The global field programme will run for at least a decade and will involve cruises in all ocean basins run by a variety of nations.

Planning has involved scientists from around 30 countries. GEOTRACES is expected to become the largest programme to focus on the chemistry of the oceans and will improve our understanding of past, present and future distributions of TEIs and their relationships to important global processes.



Deep ocean section of dissolved (<0.2 micron filtered) iron (Fe) at the zero meridian in the Antarctic Ocean. Data collected during expedition ANT 24-3 (2008) aboard icebreaker POLARSTERN in context of the International Polar Year GEOTRACES program. The very low dissolved Fe in surface waters and throughout the water column at 67-68° South is consistent with the overall limitation of Antarctic ecosystems due to lack of essential trace element Fe for biota. © Maarten Klunder. Enlarge image

GEOTRACES mission is:

To identify processes and quantify fluxes that control the distribution of key trace elements and isotopes in the ocean, and to establish the sensitivity of these distributions to changing environmental conditions.

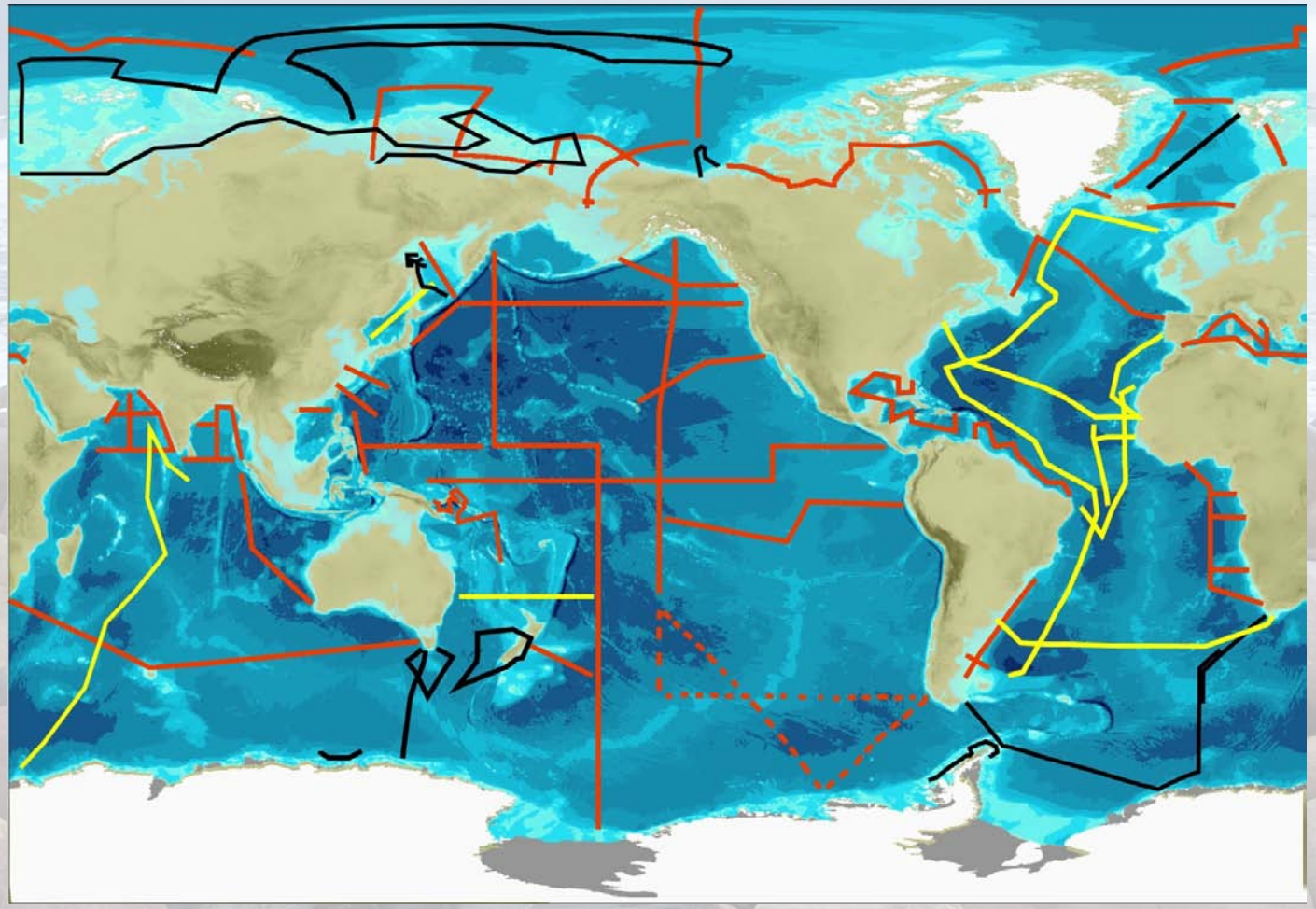
Our aim as the GEOTRACES International Data Assembly Centre (GDAC) is to provide the data management to promote data sharing and collaboration between research groups and to ensure data are made widely accessible for long-term use.

To find out more follow these links

- **Introduction** — A non-technical insight into the main goals and themes of the GEOTRACES programme.
- **Benefits** — A brief description of the long term benefits of the programme.
- **Role** — The role of the International Data Management Office.
- **Cruises** — A list of GEOTRACES cruises, past and future. Maps of all approved GEOTRACES sections, metadata forms and documentation.
- **Data** — Access to GEOTRACES data; including the Data Policy, submission guides, metadata forms and a data inventory.
- **Partners** — The partners in the GEOTRACES programme.

International Council for Science
SCOR
Scientific Committee on Oceanic Research

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL



http://www.bodc.ac.uk/geotraces/

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

The GeoTraces database

- Key parameters
  - Trace elements
  - Stable isotopes
  - Radioactive isotopes
  - Radiogenic isotopes
  - Others to allow future work to be done

- Supporting parameters
  - Salinity, Temperature, O_2, nutrients

http://www.bodc.ac.uk/geotraces/

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

The GeoTraces database

- 2014: Intermediate data product

  - It will *only* include
    - Submitted data (get your data in by 2013)
    - Intercalibrated data
    - Data passed by the IC committee

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

The GeoTraces database

- 2014: Intermediate data product

  - It will *only* include
    - Submitted data (get your data in by 2013)
    - Intercalibrated data
    - Data passed by the IC committee

DataCite
International Data Citation

doi>

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

4. A data intensive future

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

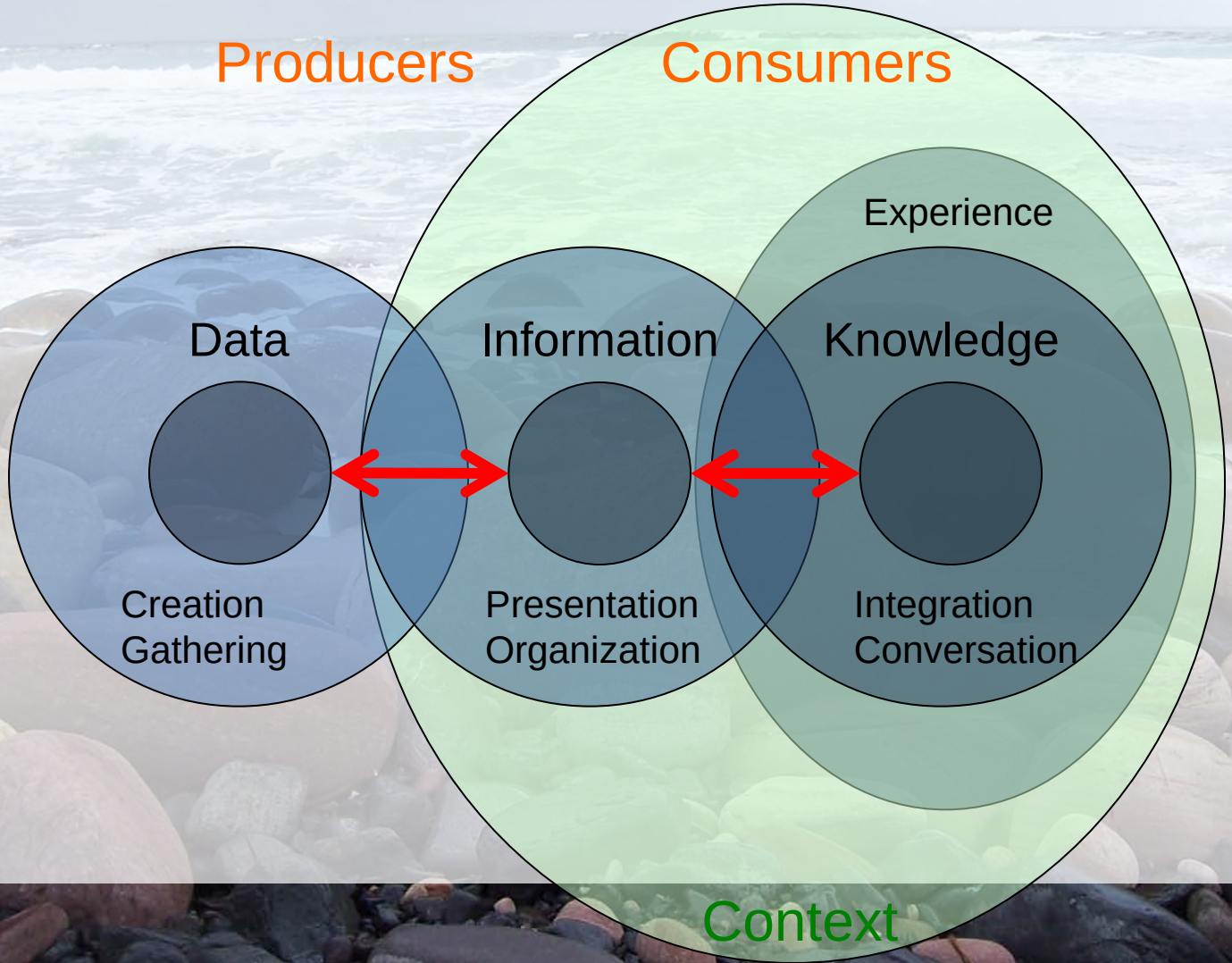
A data intensive future

“We know more than we can tell.”

Michael Polanyi
Fellow of the Royal Society
The Tacit Dimension (1967)

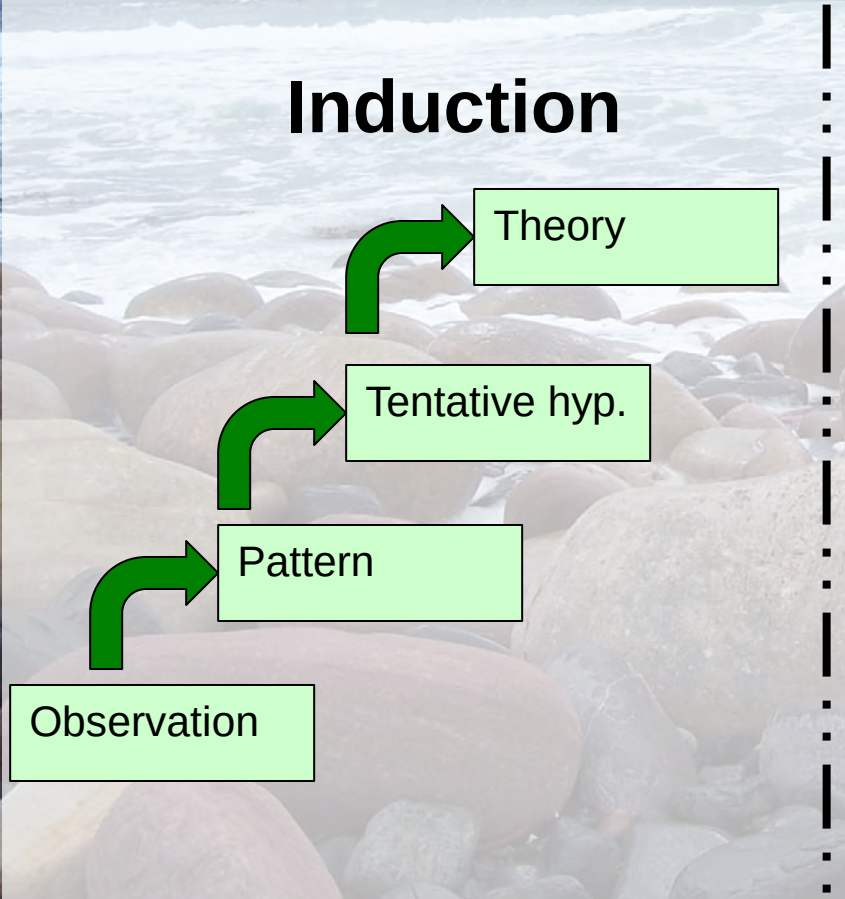
British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

A data intensive future



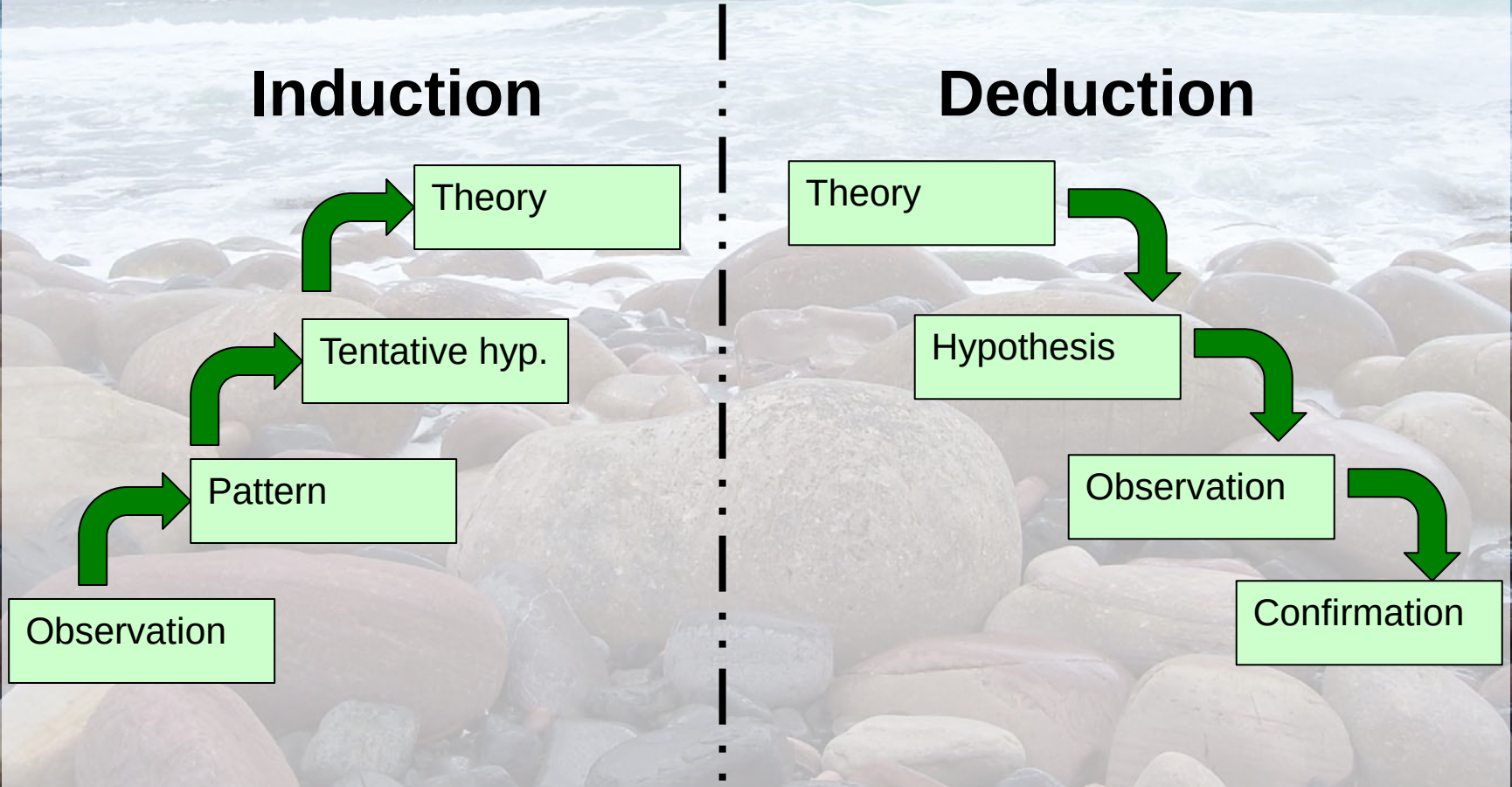
British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

A data intensive future



British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

A data intensive future



British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

A data intensive future

Abduction

Is a method of *logical inference* introduced by C. S. Peirce which comes prior to induction and deduction for which the colloquial name is to have a "hunch”

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

A data intensive future

Abduction

Is a method of *logical inference* introduced by C. S. Peirce which comes prior to induction and deduction for which the colloquial name is to have a "hunch"

- **Starts when an inquirer considers of a set of seemingly unrelated facts**
- **armed with an intuition that they are somehow connected and …**
- **But data intensive!!**
- **And this can be a job for visualization!!!**

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Conclusions

- Data matter – and increasingly so!

- The GeoTraces data assembly centre aids in making data compatible

- The GeoTraces database will be a big legacy

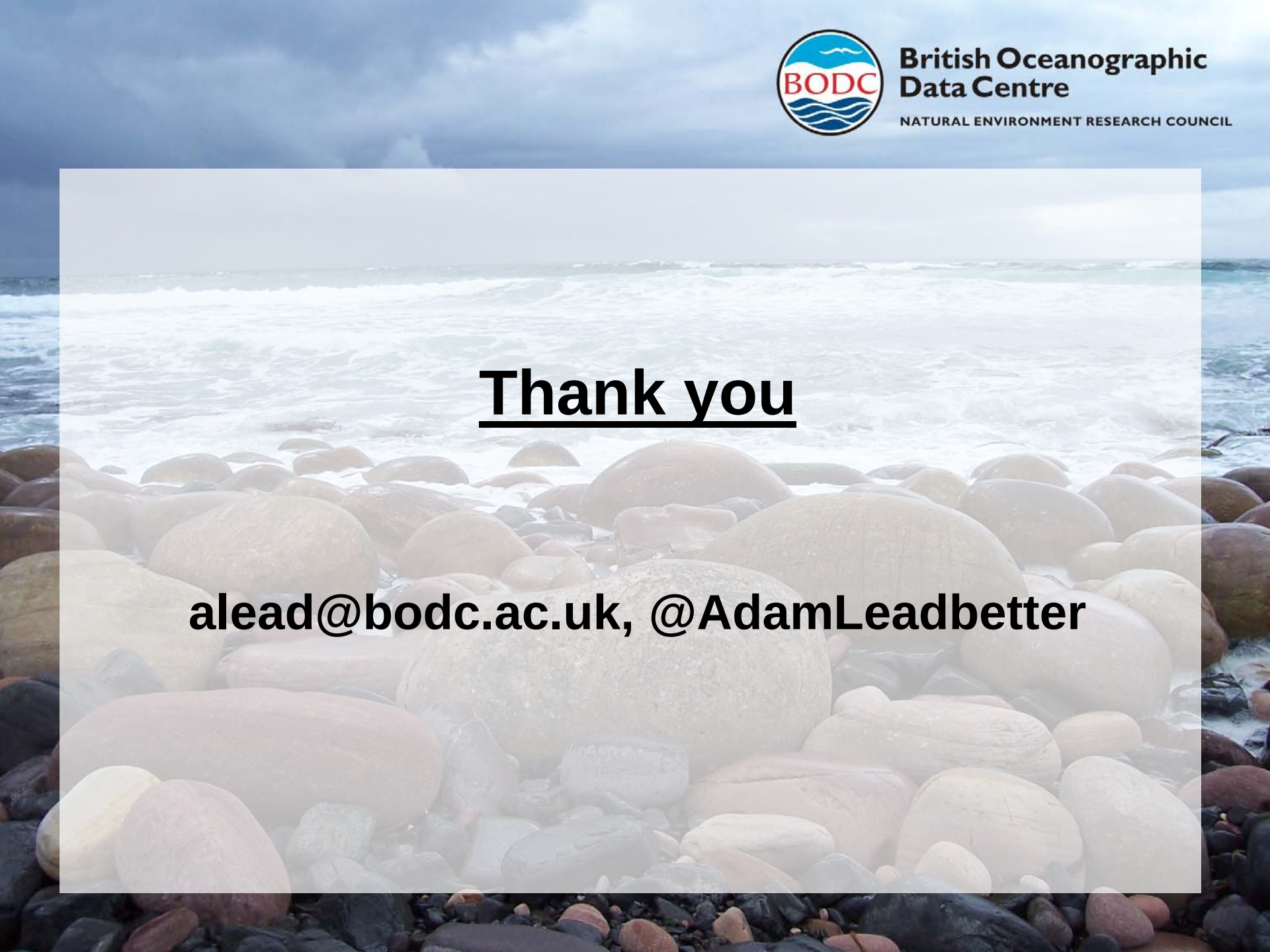
- Who knows how it may end up being used?

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Conclusions

- Low quality data have higher costs
- High quality data require communication
- Need a planned QA & QC strategy
- Investment in training
- Best practices
- Use appropriate tooling
- Extensive metadata to prevent “data entropy”

Robinson, Meyer & Lenhardt (2012). *Eos* 93(19), 189



Thank you

alead@bodc.ac.uk, @AdamLeadbetter