OECS

ORGANISATION OF EASTERN CARIBBEAN STATES

Commonwealth Marine Economies Programme

Enabling Safe and Sustainable Marine Economies across Commonwealth Small Island Developing States

Funded by UK Government

Managing National Marine Data - Challenges to managing a national data centre

MARK HEBDEN (BRITISH OCEANOGRAPHIC DATA CENTRE, UK)

WITH MATERIAL DEVELOPED BY:

CLARE POSTLETHWAITE (MEDIN)

GRAHAM ALLEN, MATTHEW DONNELLY, LUCY DICKINSON, KATIE GOWERS, SEAN GAFFNEY & ROB THOMAS (BRITISH OCEANOGRAPHIC DATA CENTRE)

FUNDED BY THE UK FOREIGN & COMMONWEALTH OFFICE

MEDIN

marine environmental data & information network

National Oceanography Centre

NATURAL ENVIRONMENT RESEARCH COUNCIL

BODC

noc.ac.uk

Managing National Marine Data-
Challenges to managing a national data centre

Strategy:

- **Our objective is to help the Caribbean region maximise their access to local marine data**
- National Oceanographic Data Centres (NODCs) play a key role in meeting this objective
- We will share our experiences and challenges of managing the UK NODC
- But, challenges quickly require solutions beyond a single NODC
- So we will share the wider context in which NODC operate as an overall approach to wide/easy access to regional marine data
- Transfer learnings to help solve regional challenges

Managing National Marine Data-
Challenges to managing a national data centre

Agenda

1. Introduction to British Oceanographic Data Centre
   - The UK National Oceanographic Data Centre (NODC)

2. Challenges (& some solutions) of a NODC

3. Share Caribbean regional experiences & challenges
   - Identify potential lessons learnt from BODC that could transfer to Caribbean
   - Future collaborative projects under the CME Programme

1. Introduction to British Oceanographic Data Centre
- The UK National Oceanographic Data Centre

1.1 What we do:
- Core Service
  - Long term availability of data
- Services
  - Vocabulary Server
  - Near real time delivery system
  - Rich metadata/information
    - e.g Cruise inventory
- Some metrics

1.2 How we do it:
- Data ingestion Projects
- Data management Projects
- Technology/System Projects

1.3 Structure
- BODC
- Networks:
  - NERC
  - MEDIN
  - SeaDataNet

1.4 Compliance
- NERC Data Policy
- EU data policy
- Open Data

1.5 Funding

1. Introduction to British Oceanographic Data Centre

1.1 What we do

Core service

Data delivered to BODC

We check the data for completeness

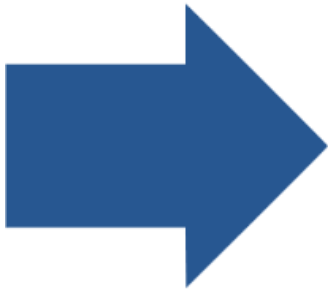
We archive a copy (in future proof format)

We record all ingestion steps and decisions

We research the context

We document details of the dataset

What happens to data?



We convert the data to a standard format

We review the quality control of the data

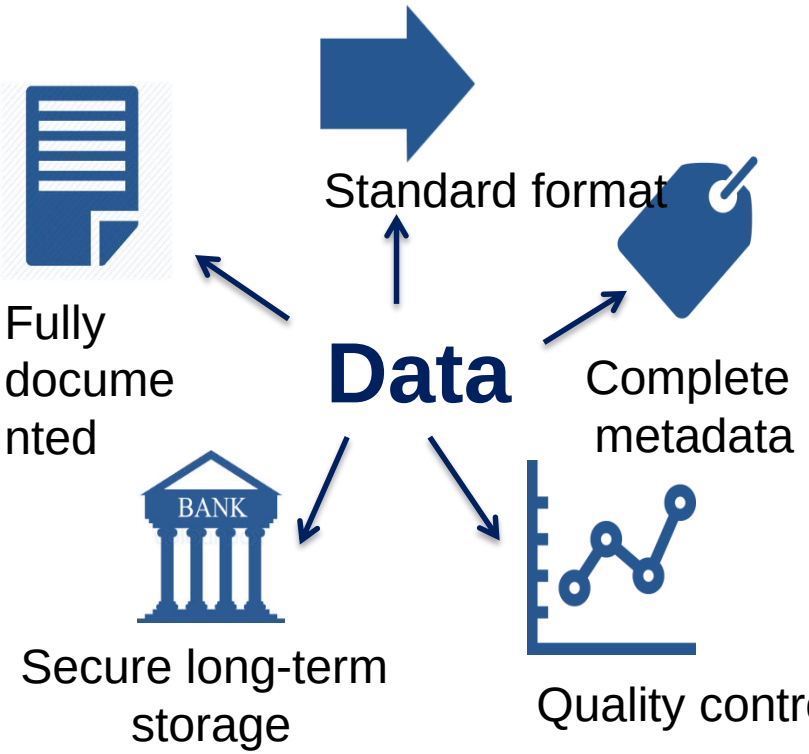
We tag with metadata from controlled vocabs

We check our processing

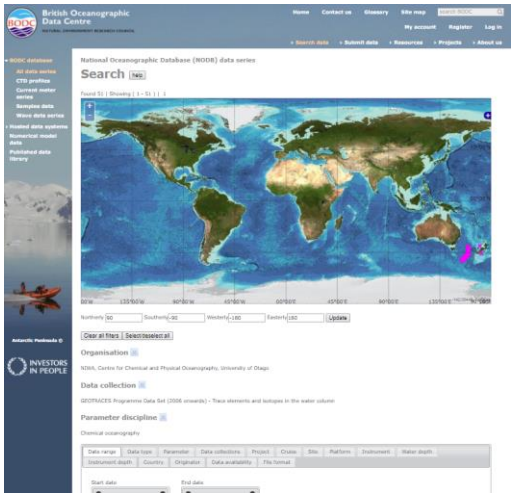
We bank the final version

We expose the dataset online

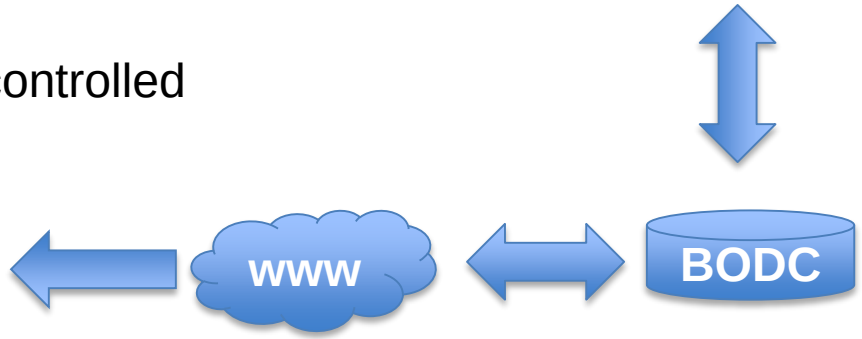
The end result



1. Data available via UI



www.bodc.ac.uk/data/bodc_database/nodb/



2. Data available via web to other systems/networks

Services - NERC Vocabulary Server - NVS

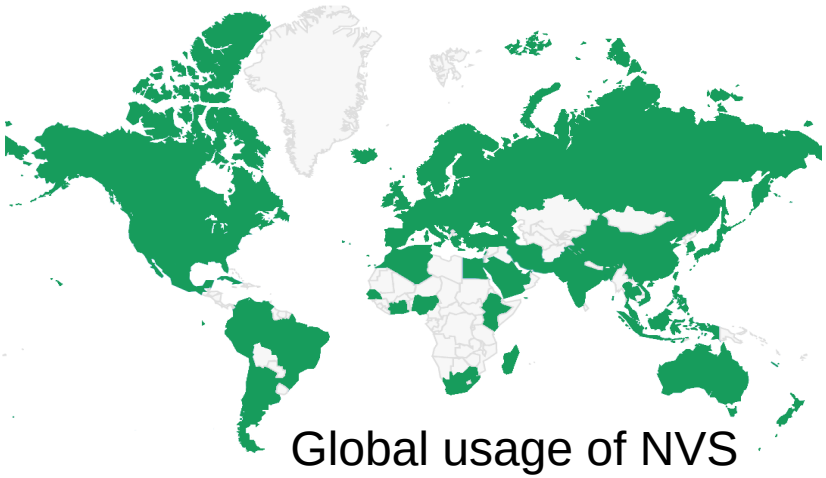
Problem: The Curse of Free Text

Solution: Controlled vocabularies

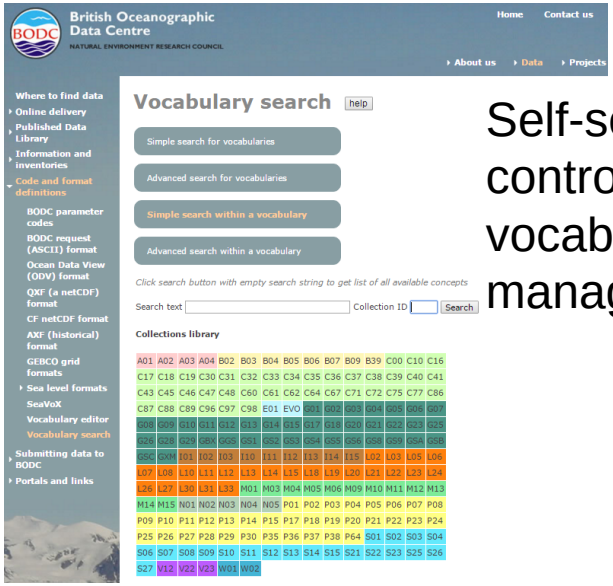
Controlled vocabulary =
‘..the use of predefined, authorised terms that have been preselected.’

- Specified when entering metadata

Services - NERC Vocabulary Server - NVS



Global usage of NVS



Self-service controlled vocabulary management

Controlled vocabulary = standard terms

'..the use of predefined, authorised terms that have been preselected.'

https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/

Submitted to the IODE Ocean Data Standards and Best Practices Project (ODSBP) as **the European regional standard for controlled vocabularies in** metadata and data formats descriptions of Marine and Oceanographic Datasets.

Services: Provide a near real time delivery system for marine autonomous systems

Services: Rich metadata & marine information

For example: Cruise inventory reports

Cruise inventory

Search for

With this search interface y
explanation of each specific

Free search
Ship name
Begin date
End date
Status
Ocean/sea area
Discipline
Data type
Host laboratory
Order by
Retrieve

Search Clear

Cruise inventory

RV Calanus Clyde 15

Cruise summary report

| New query | Results | Found 18 | Show 9 | Previous | Next |

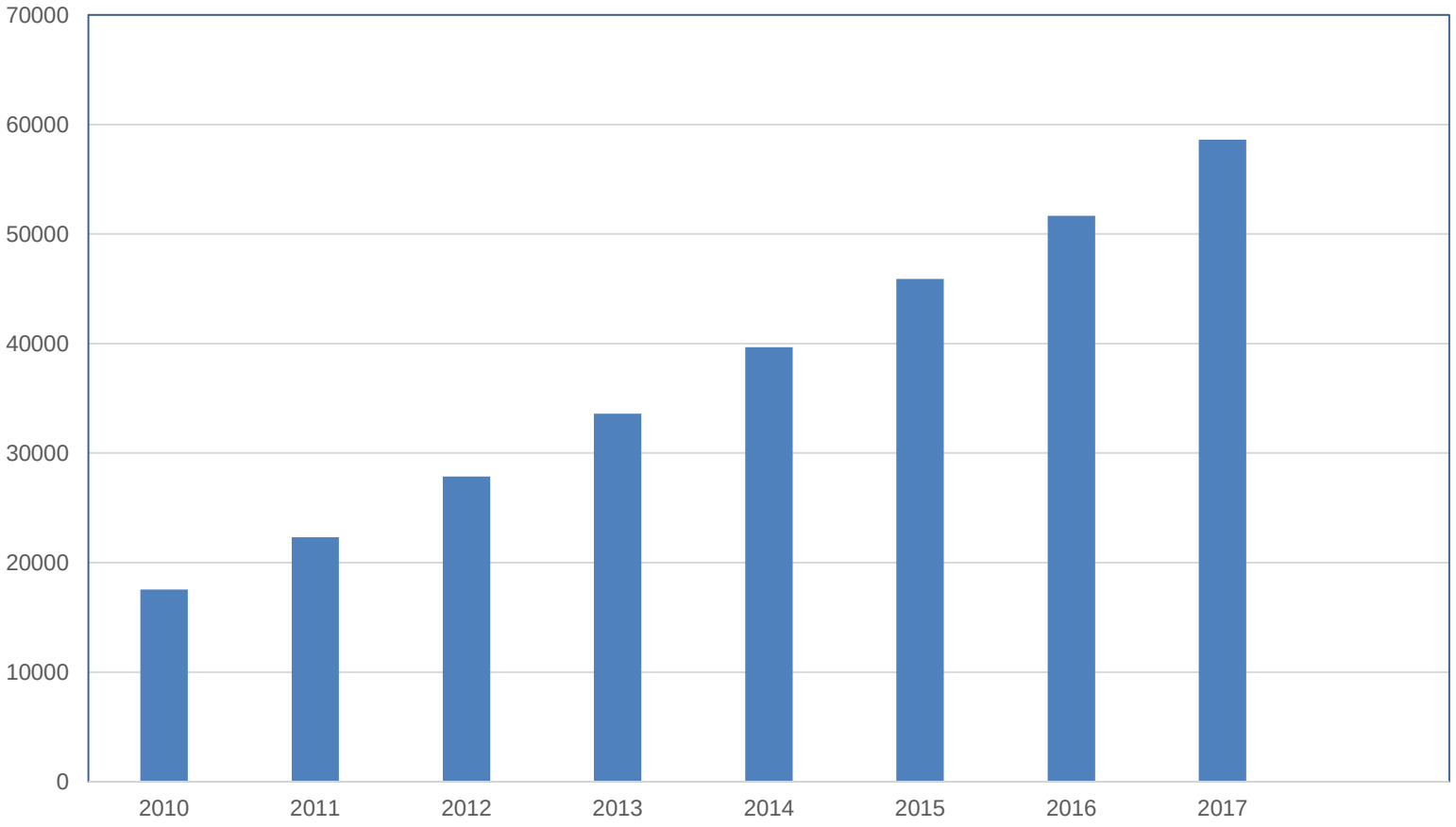
le a detailed

Cruise Info.	
Ship name (ship code)	RV Calanus (745R)
Cruise identifier	Clyde 15
Cruise period	1994-03-25 — 1994-03-27
Status	Completed
Port of departure	Dunstaffnage, United Kingdom
Port of return	Dunstaffnage, United Kingdom
Purpose	Research
Objectives	Clyde 15 Objective was to make observations of the seasonal cycle of water column structure and productivity in the Clyde Sea and North Channel of the Irish Sea with the aim of determining the dynamics and role of the physical forcing on the cycles of phytoplankton and nutrients.
Chief scientist	Rik Midgley (University of Wales, Bangor School of Ocean Sciences)
Project	PROFILE EU MAST
Coordinating body	POL, Bidston Observatory, Birkenhead, UK
Ocean/sea areas	
General	Inner Seas off the West Coast of Scotland Irish Sea and St. George's Channel
Specific	NW European Shelf, Clyde Sea, North Channel of Irish Sea
Measurements	
Physical oceanography	
Water bottle stations	Quantity: number of stations = 23
CTD stations	Quantity: number of stations = 23 Description: Continuous recording
Transparency (eg transmissometer)	Quantity: 23 Description: Transmissometer attached to CTD
Current meters	Quantity: 2 Description: 8 RCMs
Chemical oceanography	
Phosphate	Quantity: number of samples = 23 Description: From bottle samples at 5, 10, 20, 40, 60, 80, 100, 120, 150 m
Nitrate	Quantity: number of samples = 23 Description: From bottle samples at 5, 10, 20, 40, 60, 80, 100, 120, 150 m
Nitrite	Quantity: 23
Ammonia	Quantity: 23
Silicate	Quantity: number of samples = 23 Description: From bottle samples at 5, 10, 20, 40, 60, 80, 100, 120, 150 m
Moorings, landers, buoys	
Physical oceanography	
Transparency (eg transmissometer)	55° 32' 59" N 4° 55' 59" W — 2 transmissometers
Current meters	55° 32' 59" N 4° 55' 59" W — 1 surface T/S logger and 4 RCMs
Current meters	55° 43' 59" N 5° 10' 59" W — 1 surface T/S logger and 4 RCMs

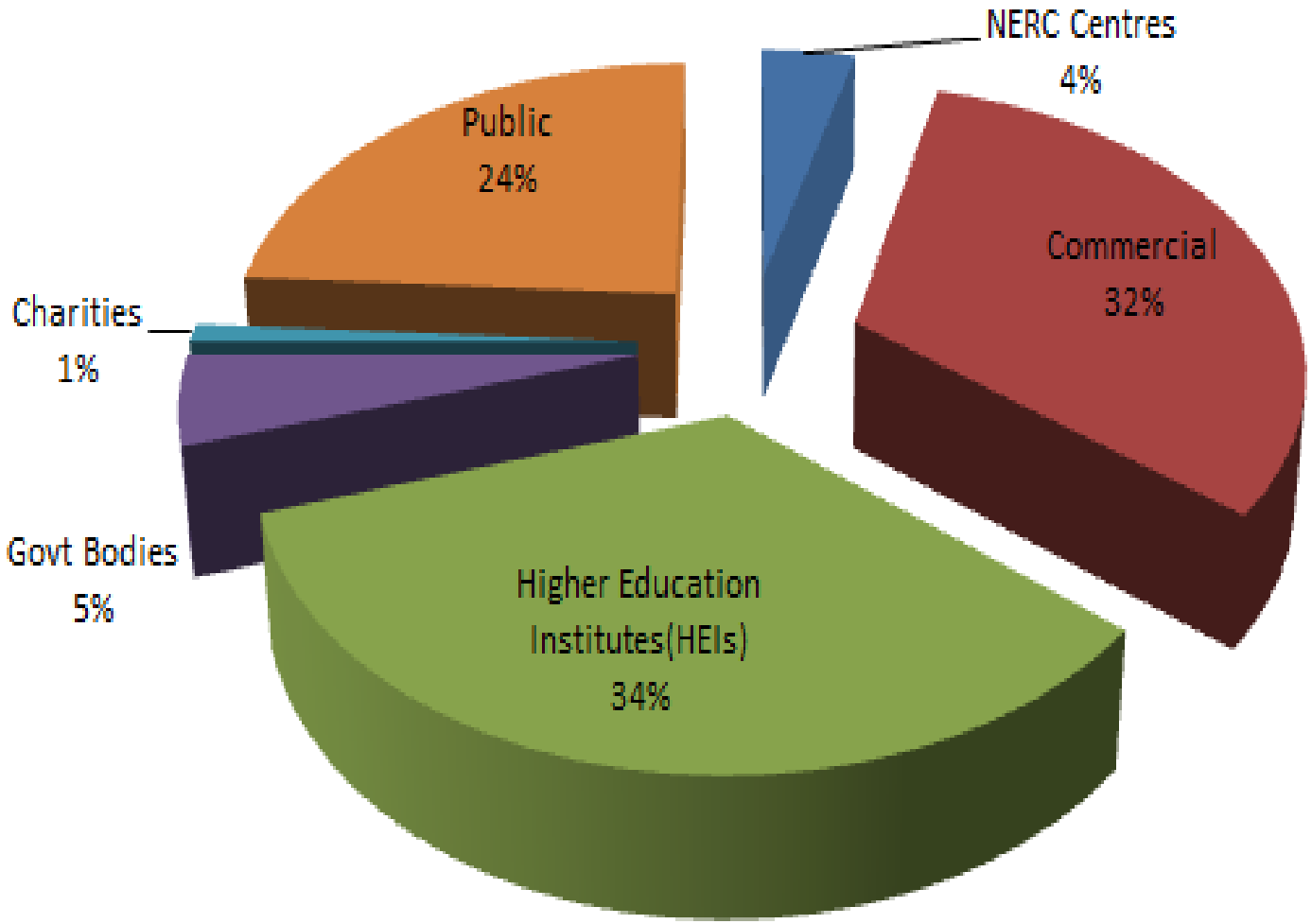
| New query | Results | Found 18 | Show 9 | Previous | Next |

Some BODC metrics…

Registered users



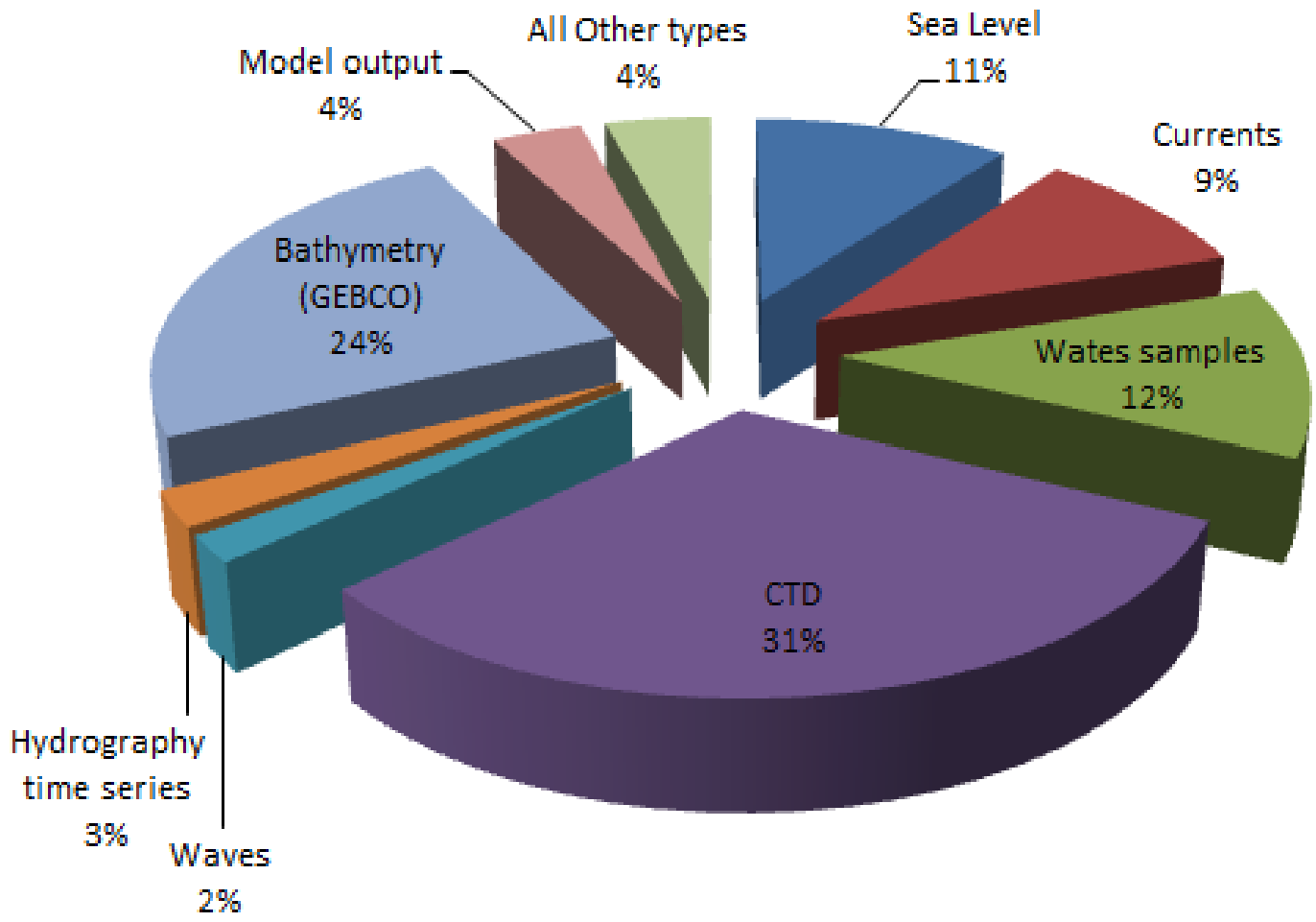
BODC user types



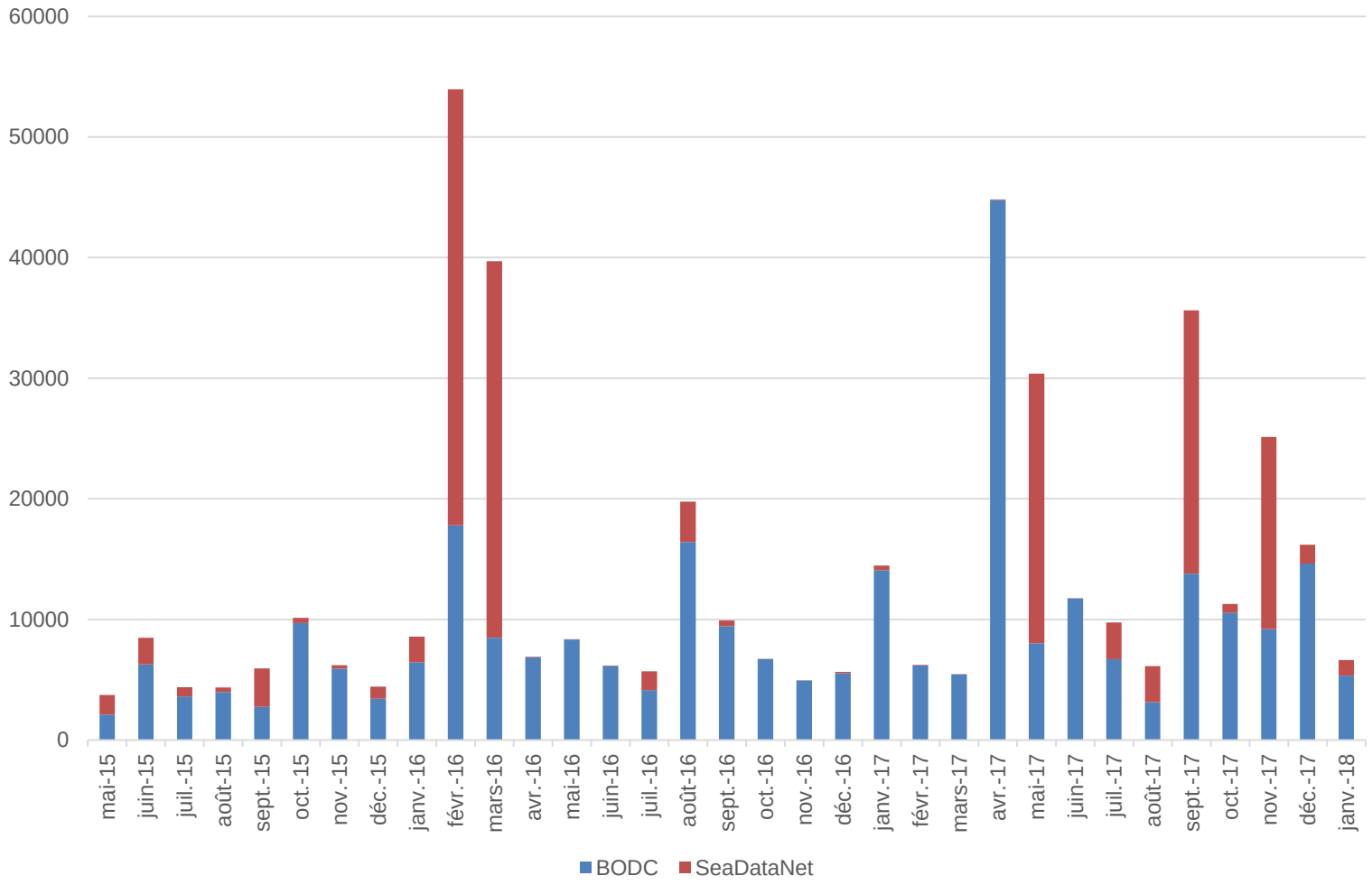
BODC in numbers

	2015	2016	Δ
Datasets downloaded	74,000	**101,187**	↑ 37%
Datasets available online	97,674	**110,063**	↑ 13%
NVS calls	350,000	**570,950**	↑ 63%
GEBCO grid downloads	34,000	**20,985**	↓ 38%
GEBCO WMS calls	2,375,147	**4,018,361**	↑ 69%
Web hits	920,000	**979,686**	↑ 6%
Seal tag dataset	20,660	**32,815**	↑ 59%

BODC data by type



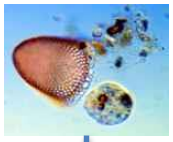
BODC downloads though BODC data portal



1.2 How we do it

Data ingestion Projects













BODC

Biological, chemical, physical and geophysical *water column* data

Embedded data management

Data in

Data management & system projects

RAGNARoCC

Radiatively active gases from the
North Atlantic Region and Climate Change

PSMSL

GLOSS

ARGO

UK Tide Gauge Network

UK Ocean Acidification

FASTNEt

BRIDGES

Fluxes Across Sloping Topography of the North East Atlantic

General Bathymetric Chart of the Oceans GEBCO

iGlass
Using interglacials to assess future sea-level scenarios

SeaDataNet

SHELF SEA BIOGEOCHEMISTRY
www.uk-ssb.org

DIMES

Diapycnal and Isopycnal Mixing Experiment
in the Southern Ocean

GEOTRACES

Biogeochemical global cycles

plus ENVRI

AtlantOS
Optimising and Enhancing the Integrated Atlantic Ocean Observing Systems

EMODnet
European Marine Observation and Data Network

Rapid Climate Change

RAPID-watch

Atlantic Meridional Overturning Circulation

OSMOSIS

Ocean Surface Mixing,
Ocean Sub-mesoscale Interaction Study

Atlantic Meridional Transect

MERMAN –
Clean Safe Seas Environmental Monitoring Programme (CSEMP)

imorc
SYSTEM OF INDUSTRY METOCEAN DATA FOR THE OFFSHORE AND RESEARCH COMMUNITIES

Sense OCEAN

ANDREX

Antarctic Deep Water Rates of Export

ODIP II

Ocean Data Interoperability Platform

National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

BODC

noc.ac.uk

1.3 Structure

Who are we ?

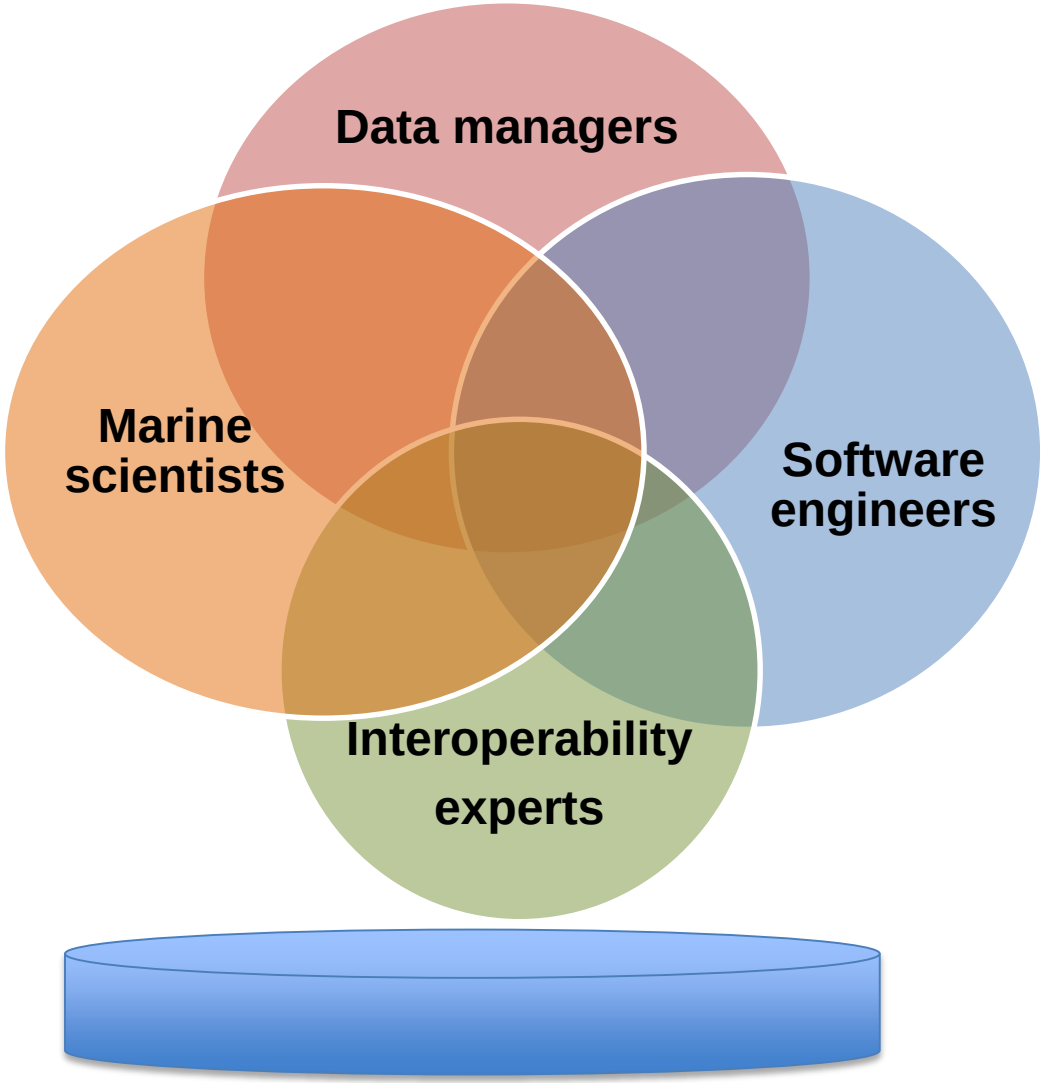
Software

Data Management

- 50 people
- Liverpool and Southampton (42:8)
- Data managers: Software engineers (35:15)
- Gender: F-M: 31:19
- 20 PhDs in Marine Science
- Tenure: days to 44 years

What is BODC Today ?

A team of 50



and a data management system

Structure: NERC Environmental Data Network

1. Marine

BODC

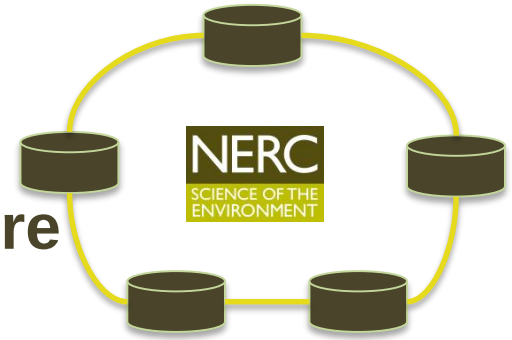
National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

2. Terrestrial & freshwater

Environmental Information Data Centre

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

5 Environmental Data Centres



3. Polar and cryosphere

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Polar Data Centre

4. Geosciences

National Geoscience Data Centre

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

5. Atmospheric & Earth Observation

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Measure once, use many times

MEDIN

marine environmental data & information network

National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

cfpo@bodc.ac.uk

oceannet.org

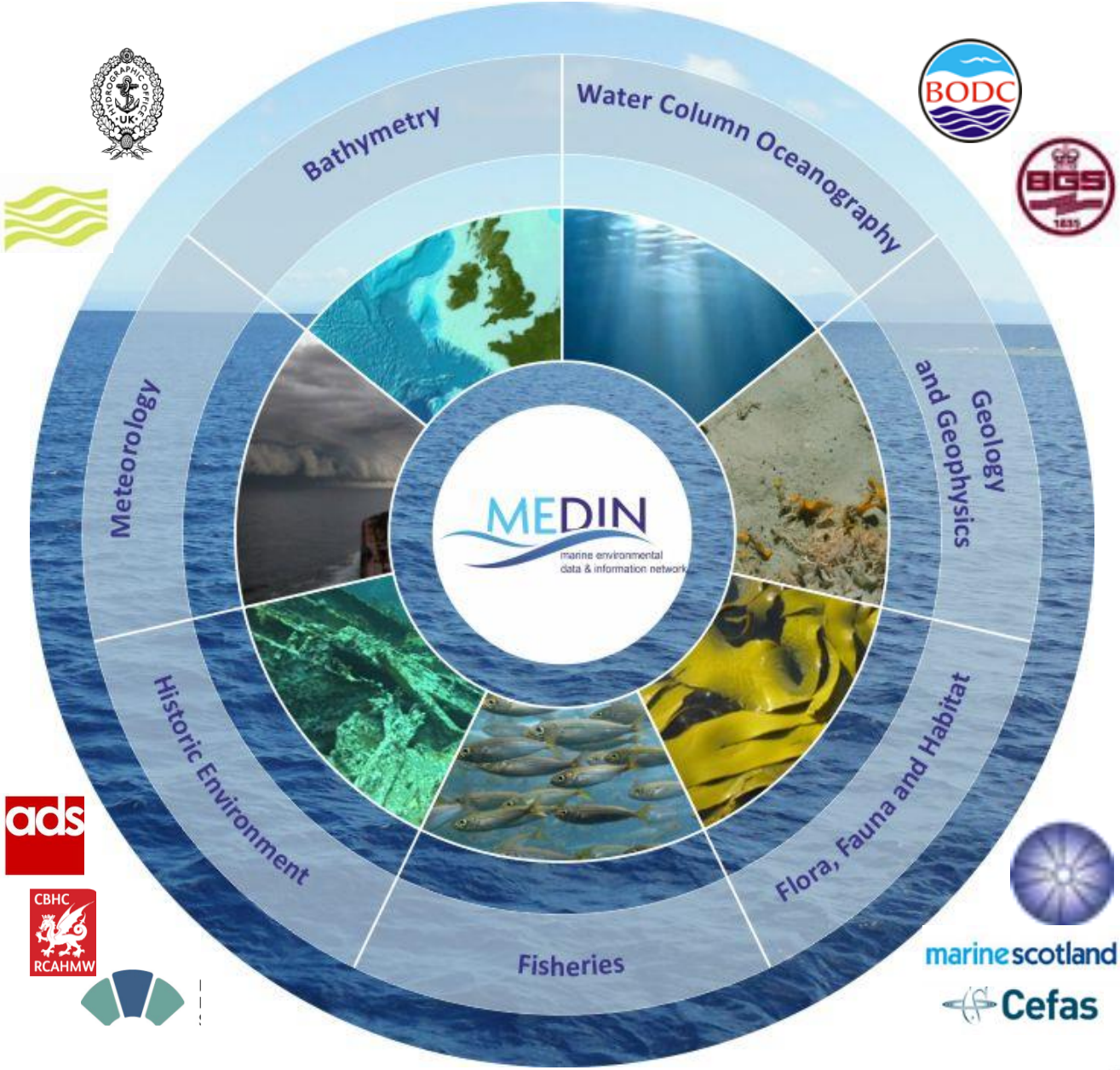
noc.ac.uk

What is MEDIN? – UK marine data network

Open partnership

Report to Marine Science Co-ordination Committee

A network of UK marine data centres



Structure: SeaDataNet – EU marine data network

BODC was very actively involved in SeaDataNet and SeaDataNet II

We not only provide data and metadata but are also involved in

- Managing Vocabularies for data processing, communication and quality assurance
- Developing European standards for metadata
- European wide metadata directories
- Designing and implementing interoperable systems

Structure: A network of networks

IODE International Oceanographic Data and Information Exchange

NERC SCIENCE OF THE ENVIRONMENT — Data Catalogue Service *sharing NERC data*

EMODnet European Marine Observation and Data Network

MEDIN marine environmental data & information network

BODC Data Portal

SeaDataNet

ICSU WORLD DATA SYSTEM

GEO GROUP ON EARTH OBSERVATIONS | GEOSS Portal
Discover, Access, Contribute
Earth Observations, Information and Services

esa

Data out

1.4 Compliance

Compliance - NERC Data Policy

Home / Research / National capability / Data centres / Data policy

Data policy

NERC has a well-established Data Policy that sets the ground rules that all those funded by NERC must follow in managing the data that they collect. The Data Policy details our commitment to support the long-term management of environmental data and also outlines the roles and responsibilities of all those involved in collecting and managing environmental data. Central to the policy is that NERC-funded scientists must make their data openly available within two years of collection and deposit it in a NERC data centre for long term preservation. The aim is that all NERC-funded data are managed and made available for the long-term for anybody to use without any restrictions.

NERC Data Policy (PDF, 28KB)

Guidance notes for the NERC Data Policy (PDF, 209KB)

Licensing & Charging policy

This policy describes why, when and how NERC will and will not apply charges for Environmental Data and Information Products. It underlines our commitment to apply any charges in a transparent and consistent way.

We have created the policy to be consistent with relevant legislation and Government guidance on charging for information, on open data provision and on licensing.

Central to the policy is that NERC will make its environmental data available free of charge apart from special cases that involve third party data. There may be a charge for information products. Where there is a charge, the charges and licensing arrangements will be clearly explained, compliant with Government legislation and guidance, and applied in a consistent manner.

NERC Licensing & Charging Policy (PDF, 37KB)

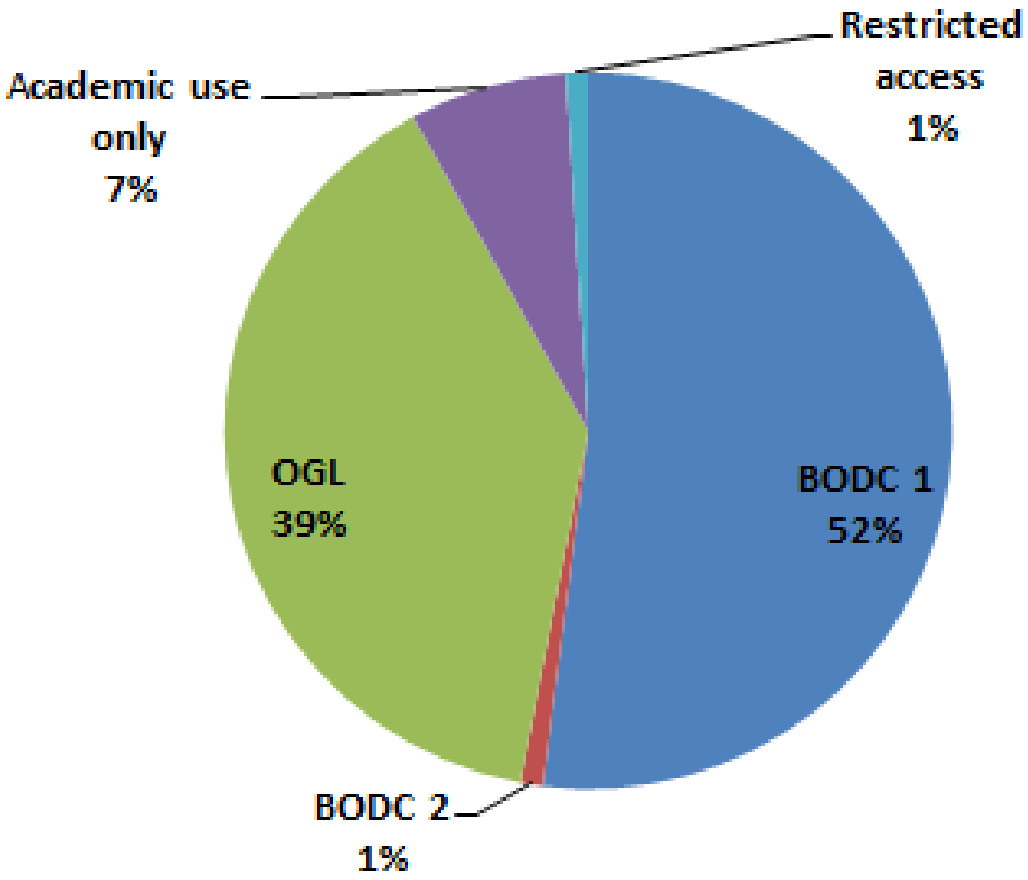
If you have any comments or questions, or require further information on the NERC Licensing & Charging Policy, please contact NERC's data management co-ordinator:

Mark Thorley
☎ 01793 411567
✉ mrt@nerc.ac.uk

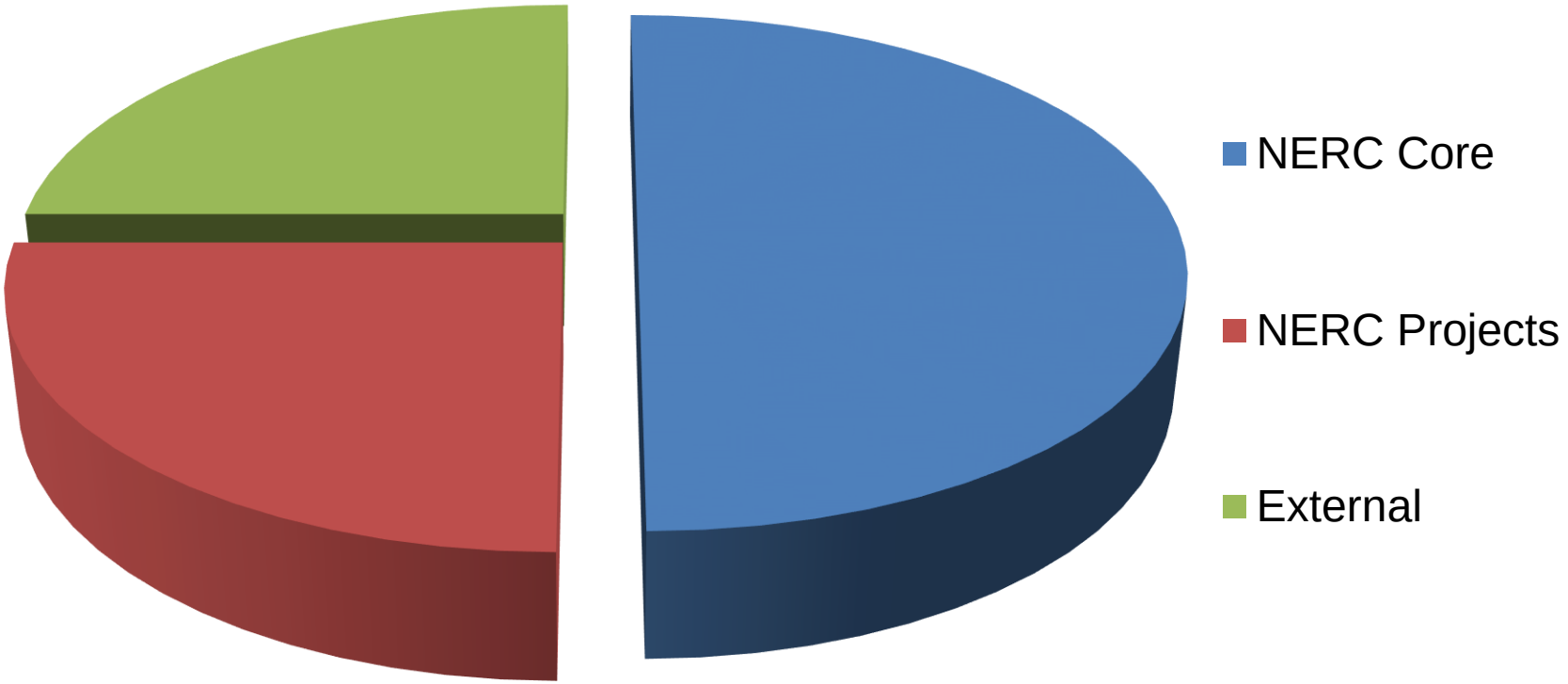
'…NERC-funded scientists must … deposit the data in a NERC Data Centre…'

'…NERC will make its env data available free of charge …'

Compliance - Open Data Policy



1.5 Funding



2. Challenges of a National Oceanographic Data Centre

2. Challenges of a National Oceanographic Data Centre

1. ‘No man is an island’

John Donne (1572-1631)

‘Human beings do not thrive when isolated from others’

A single data portal is not the solution:
The need for networks and Application Programming Interfaces (APIs) to disseminate data and metadata

Lesson learnt:

The need for networks and APIs

2. Challenges of a National Oceanographic Data Centre

2. (lack of) **Funding**

- To build the system
- To run the system
- To use the system to ingest data

Lessons learnt:

- Cost to ingest included in projects that generate the data
- The role of compliance

3. Take but not give

- Everyone wants a data portal to go and get data
- But less enthusiasm for the effort to provide data

Lessons learnt:

- Ease of ingestion – make it easy
- Working with the data collectors – add value in the project
- Funding: include ingestion costs in collection project budget

4. Data complexity and volumes

- Lots of different data types – complex ingestion system & hard User Interfaces (UI) to discover and access
- Large volumes
- Real time data streams

…Getting worse due to technology accelerations

Lessons learnt:

- Engage users when building systems
- Get feedback

5. A sea of (data) islands

- **There are so many data portals!**
- Where are they
- Which one to use
- How to search across
- How to merge data

Lessons learnt:

- Don’t try and do anything
- Don’t re-invent the wheel and build another data portal
- Re-use existing systems
- Importance of interoperability

2. Challenges of a National Oceanographic Data Centre

John Donne (1572-1631)

1. ‘No man is an island’

A single data portal is not the solution
The need for networks and APIs

2. (lack of) **Funding**
   To build
   To run
   Ingestion costs

3. **Take but not give**
   - everyone wants a data portal to go and get data
   - But less enthusiasm for the effort to provide data

5. A sea of (data) islands
- **There are so many data portals!**
- Where are they
- Which one to use
- How to search across
- How to merge data

4. **Data complexity and volumes**
- Lots of different data types – complex ingestion system & hard UI to discover/access
- Large volumes
- Both of above getting worse due to technology accelerations

3. Share Caribbean regional experiences & challenges