



Science & Technology
Facilities Council

Data manager

“Plumber”

Sam Pepler

Roles and Responsibilities for Effective
Data Management. 28-29 November
2008



**British Atmospheric
Data Centre**

NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL



Science & Technology Facilities Council

Rutherford Appleton Laboratory



**National Centre for
Earth Observation**

NATURAL ENVIRONMENT RESEARCH COUNCIL



Overview

- The organisation of my data centres
- The people doing the work as data managers and their career paths
- Summary

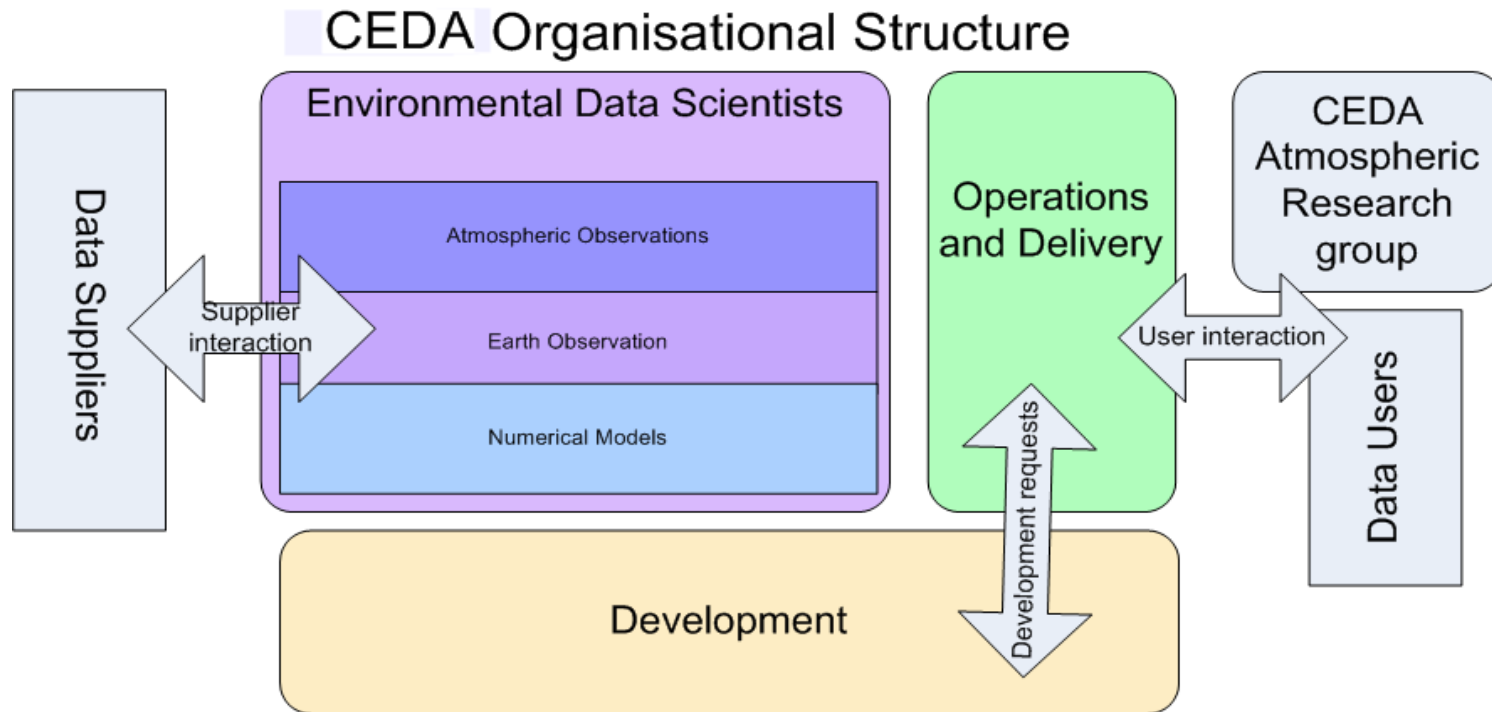


What is a data manager

- “THE SKILLS, ROLE AND CAREER STRUCTURE OF DATA SCIENTISTS AND CURATORS: AN ASSESSMENT OF CURRENT PRACTICE AND FUTURE NEEDS”

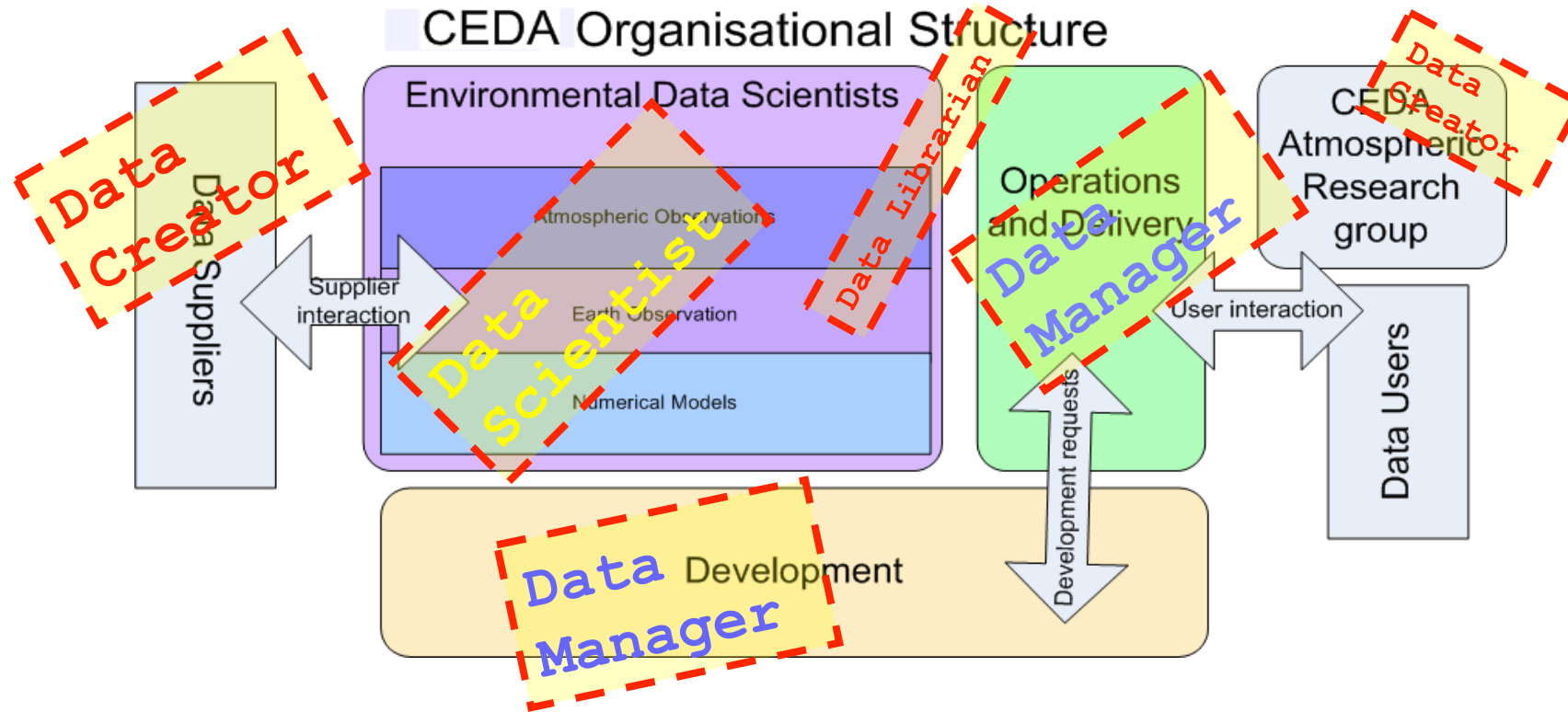
- Data managers: people who are **computer scientists, information technologists or information scientists** and who take responsibility for **computing facilities, storage, continuing access and preservation of data**. They **liaise extremely closely with data scientists**, ensuring that the right **technological** facilities are **available for the research group** to be able to carry out its work effectively. Some data managers described their role as data **‘plumber’**, piping data from one place to another, ensuring data flows operate properly and that valuable data are not lost
- I would not call this a data manager myself.

British Atmospheric Data Centre (BADDC) Structure





... Where the roles fit in



Rob Harper - Storage coordinator

- **Background:** Sys admin at NRPB > Computer Science degree while working for BADC > Left to do sys admin work in different department.
- **Skills:** Sys admin.
- **Tasks**
 - Compiling the BADC system roadmap. Including volume planning, costing the options available and gathering storage requirements.
 - Development of a media management system for BADC.
 - Assignment of storage for datasets.
 - Dataset backup management.
 - Monitoring disks arrays for failure, overfilling or bad performance.
 - Writing scripts to automate functions.
 - Working with the system manager to implement and configure new storage.
 - General query handling, particularly queries relating to storage problems and other system failures.
- **Interactions:** Data scientists, System support

Dominic Lowe - Developer

- **Background:** Physics with env science BSc > Technical publishing > GIS MSc > GIS consultancy > BADC
- **Skills:** Web service Programmer, software engineer.
- **Tasks:**
 - Data modelling
 - Programming
 - Liaison with GIS standards community
 - Develop the Climate Science Modelling Language (CSML)
- **Interactions:** Standards community, Data scientists, Development projects

Andrew Harwood - Infrastructure Manager

- **Background:** BSc Physics > MSc IT and medical physics > Astronomy programmer > BADC
- **Skills:** Sys admin, programming, software management, database admin.
- **Tasks:**
 - Maintenance and tracking of in house software.
 - Deployment of software
 - Software development
 - User admin
 - Database admin
- **Interactions:** Data scientists, Development group, Researchers

Matt Pritchard – Development Manager

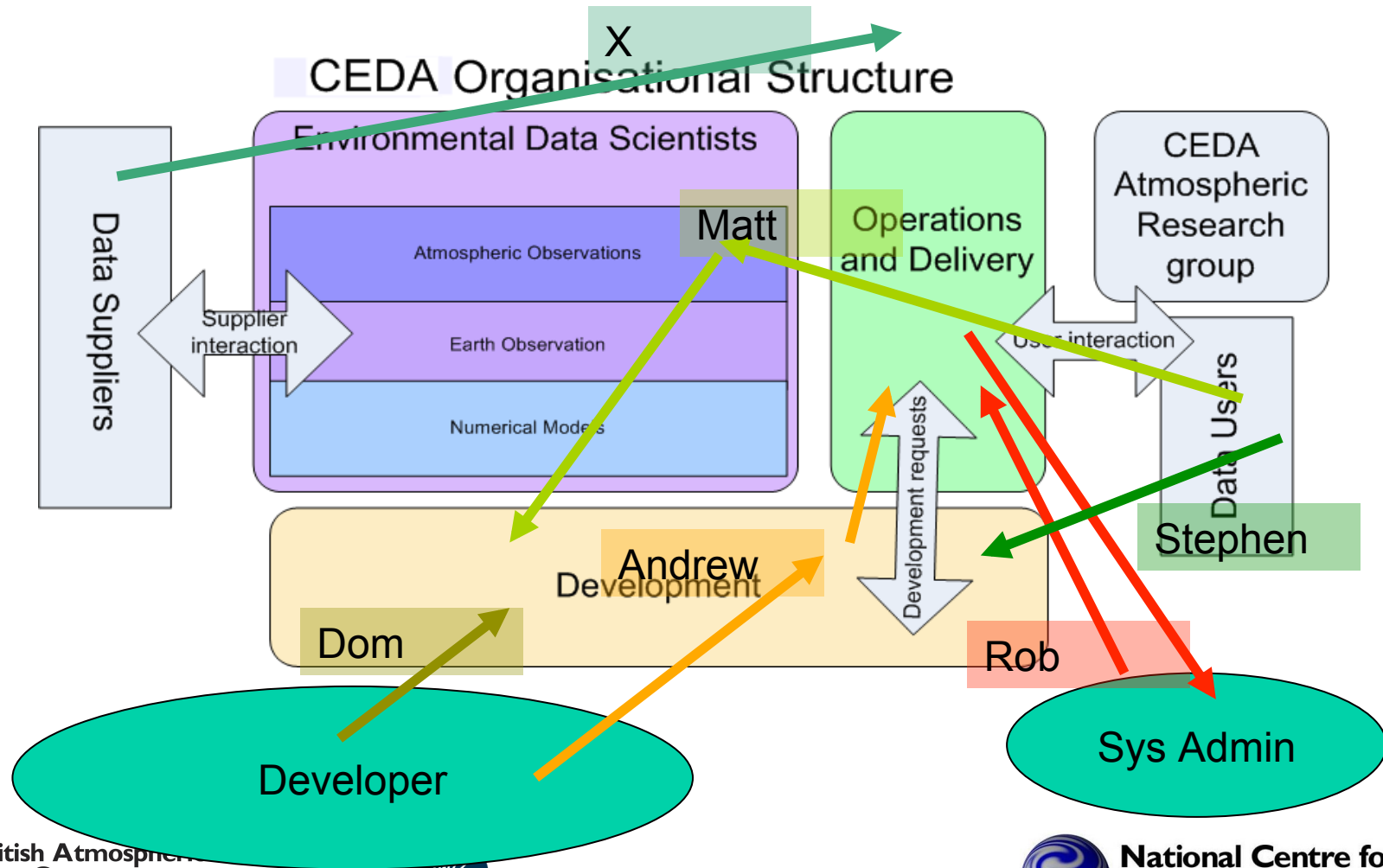
- **Background:** Geology PhD > NEODC manager > Development manager
- **Skills:** Web programming, people management, EO data
- **Tasks**
 - Organise and prioritise the development projects
 - Work on development project
 - EO data management
- **Interaction:** Developers, EO researchers, operations.

Dr X

- **Background:** Climate Modelling > HPC requirements for NERC climate
- **Skills:** Running numerical models. HPC configuration. Programming.
- **Tasks**
 - Support of Climate scientists
 - negotiation for system modification
 - Managing computer time
- **Interaction:** Researchers, data centres, HPC facilities



Development paths



Summary

- I would not call this a data manager myself. Plumber is a better description. “Data infrastructure maintainer” is a bit of a mouthful.
- We separate developers and maintainers.
- Its not just sys admin or software development, but it's a big part of what's needed.
- Where is the line between sys admin, developer and a data manager?
- Career paths are not obvious.