



Training the new data scientist

EUDAT User Forum – Barcelona, 2012

Adam Carter

A.Carter@epcc.ed.ac.uk



8th March, 2012



Introduction

- EUDAT has some big plans for a collaborative data infrastructure
 - but we also want to make sure that people are well trained in how to make the most of it
- Task 3.3: Education & Training
 - Part of the Dissemination work package
 - 36 person-months of effort over the three year project
 - Partners:
 - EPCC (lead), CINES, INGV, UCL, BSC, CSC



Approach

- Year One:
 - Training Needs Analysis
- Years Two & Three:
 - Training Programs and Activities
 - Development of Training Material and Courseware Repository
- We may also co-ordinate some training events (possibly made available by videoconference) during year one



Training Needs Analysis

- Approach:
 - Start with core communities
 - Broaden scope where appropriate
- Core Communities: We will approach you (we possibly already have!)
- Core Communities and All Others:
 - Please feel free to approach us too
 - This user forum provides an ideal opportunity



Training Needs Analysis (Approach)

- We're not just going to send out questionnaires
 - Generally a low and often unrepresentative response
- We're going to identify contacts and “typical users” and discuss needs with them
 - An informal interview
 - Discussion areas listed in advance to ensure all areas are covered
 - We welcome input from anyone who can contribute
- Users might not know what they need to know!
- More personal engagement allows us to excite and interest users in what EUDAT can provide



Challenges

- Many different communities
 - Possibly with existing training provision
- Data is a broad subject! Many different aspects to cover
 - Make initial study of needs open and broad, but make users aware of those services likely to be rolled out first by EUDAT



Planned Activities

- October 2012:
 - Training Needs Analysis Workshop
 - timed to coincide with the next user-forum



Planned Activities

- September 2012:
 - Deadline for submitting Needs Analysis deliverable
- October 2012:
 - Training Needs Analysis Workshop
 - timed to coincide with the next user-forum



Planned Activities

- June 2012:
 - Training Needs Feedback Meeting
 - subject to community interest
 - possibly by videoconference
- September 2012:
 - Deadline for submitting Needs Analysis deliverable
- October 2012:
 - Training Needs Analysis Workshop
 - timed to coincide with the next user-forum



Planned Activities

- June 2012:
 - Training Needs Feedback Meeting
 - Tell you what we think you've told us
 - Share needs amongst communities
- October 2012:
 - Training Needs Analysis Workshop
 - Share our findings with wider community
 - Discuss how the findings of the analysis can be turned into plans for years 2 and 3



Ideas Always Welcome!

- Data staging & movement
- Data storage
- Data-intensive computing
- Streaming data
- Workflows/Dataflows
- Datamining
- Data labelling / re-use
 - formats, metadata, etc.
- Data semantics
- RDBMSs
- Data hardware
 - disks, SSDs, controllers, networks
- Distributed data
 - incl., e.g., distributed joins
- AAI principles and Shibboleth
 - SAML, certificates, etc.
- Persistent Object Identifiers and APIs
 - Handles, DOIs, URNs, etc.
- Researcher IDs and ORCID
- Policy Rules and iRODS
- Repository Systems (Fedora, D-Space, etc.)
- Metadata Principles and Components
 - Principles, Standards, OAI-PMH, Aggregation and Indexing, Semantic Mapping, Portals, etc.
- Workflow Orchestration and Execution
 - Principles, Standards, Frameworks, etc.



Ideas Always Welcome!

- Specially arranged sessions or during existing subject-area meetings?
- Teaching the users, or teaching the teachers?
- Should we include curriculum development?
 - Which curricula?
- Online materials?
- Web? YouTube? iTunesU?
- Training infrastructure:
 - Machines to log on to and try things out?