Data Skills Curricula Framework

Vicky Lucas

Training Development Manager, IEA

Content

- Background
  - Beginnings
  - Skills gap analysis
  - Workshop outcomes
- Curricula
  - Goals
  - Structure
  - Example course
- Q&A

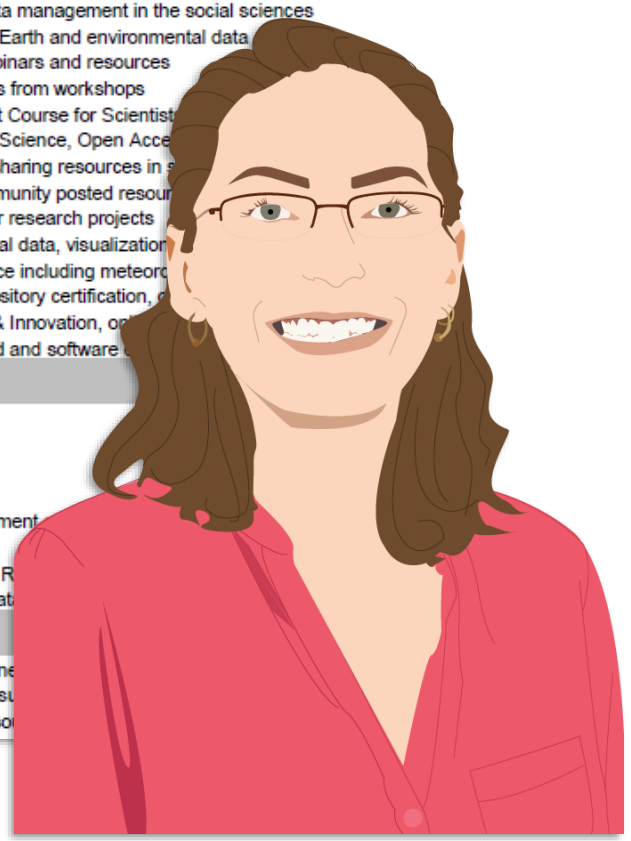
Background

- Belmont Forum
  - Partnership
  - Interdisciplinary and transdisciplinary science
  - Global environmental change
- e-I&DM
  - Data sharing, e-infrastructures
  - Action Theme 4
    - Capacity building and human dimensions
    - Digital skills for data intensive global change research
    - A Place to Stand – guiding document

e-I&DM Human Dimensions

8. RECOMMENDED TRAINING ACTIVITES

ONLINE	
ANDS	23 Research Data Things, resource for managing data
CMMI Institute	Enterprise data management training and certification
CESSDA	Training on research data management in the social sciences
DataONE	Supporting discovery of Earth and environmental data
DPC	Digital preservation, webinars and resources
ESA	Resources and materials from workshops
ESIP	Data Management Short Course for Scientist
FOSTER	Online courses in Open Science, Open Acce
IASSIST	Community repository, sharing resources in
GEOCAB	Earth Observation, community posted resour
MANTRA	Managing digital data for research projects
MELODIES	Linked and environmental data, visualizatio
MetEd	Resources for geoscience including meteorc
Research Data Alliance	Data management, repository certification,
RRI	Responsible Research & Innovation, o
UK Data Service	Datasets, topics, method and software
MOOCs	
Illinois Urbana-Champaign	Data Mining
Illinois Urbana-Champaign	Data Visualization
Johns Hopkins	Data Science
Johns Hopkins	Statistical Inference
North Carolina Chapel Hill	Research Data Management
Stanford University	Machine Learning
PRACE	Managing Big Data with R
University of Southampton	Introduction to Linked Dat
SUMMER SCHOOLS AND WORKSHOPS	
Alpbach (Austria)	Space science and engine
CODATA-RDA	1st run 2016, Unix, R, visu
CEDA (UK)	Scientific computing, reso



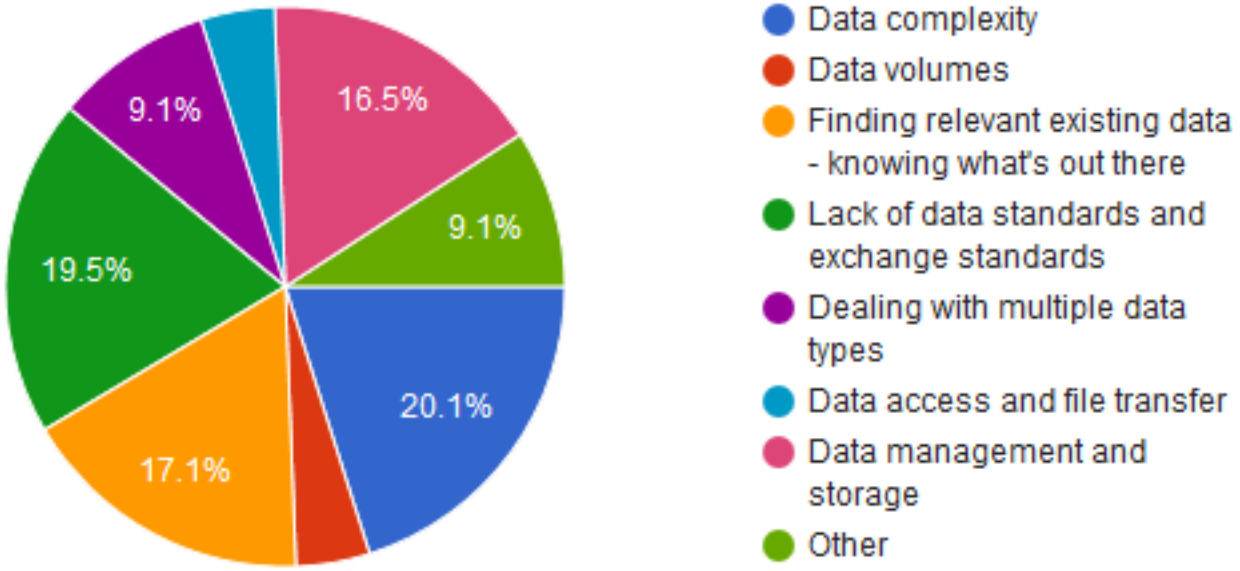
Digital skills for data intensive global change research (2016-17)

- Skills gap survey
- List: recommended training
- Workshop at EGU 2017
- Curricula
  - Endorsed by Belmont Forum 2017
- Human Dimensions Champion
  - Supported by Rowena, Bob, Tina

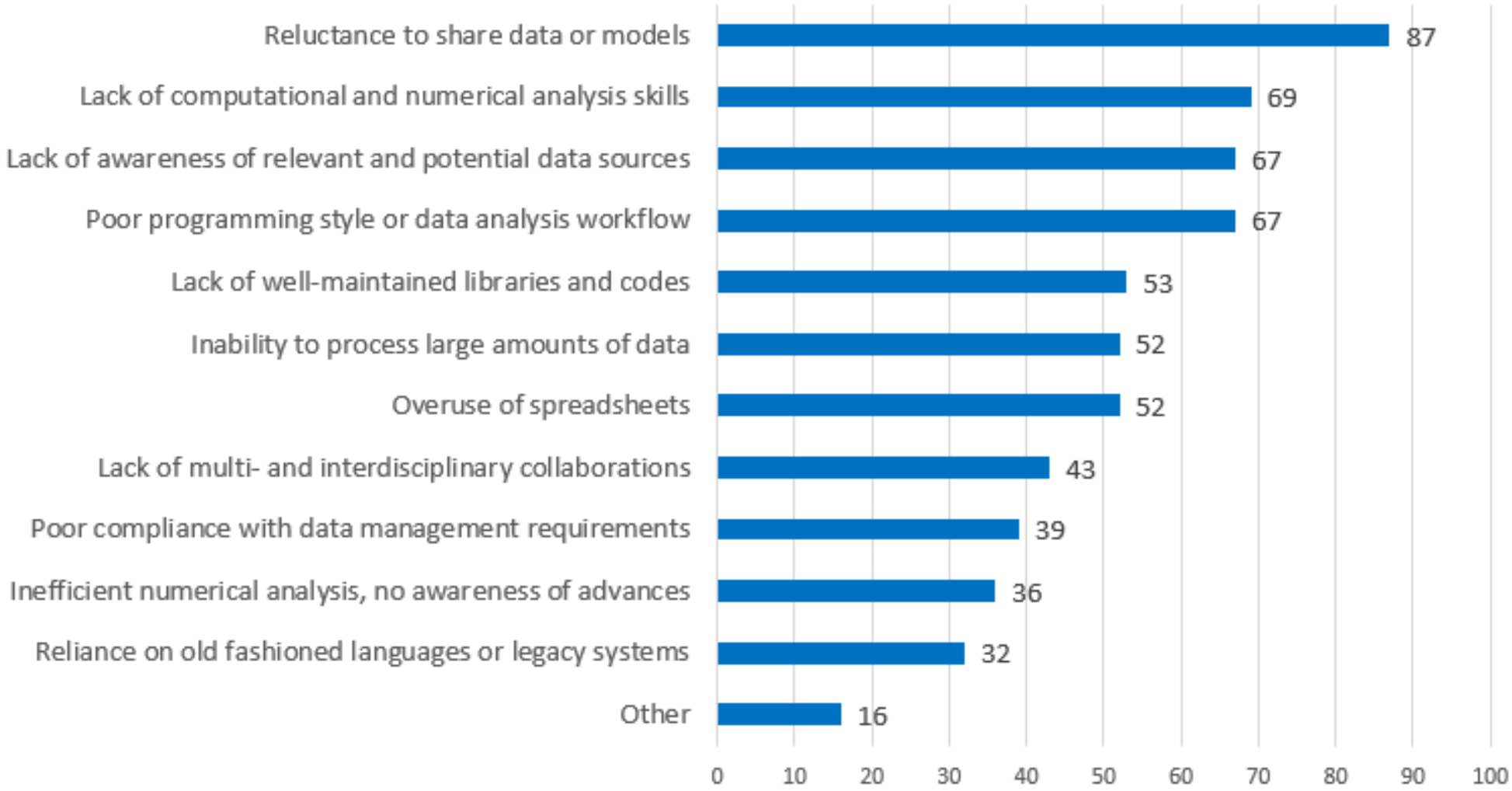
List of recommended training & Vicky Lucas, Human Dimensions Champion

Skills gap survey

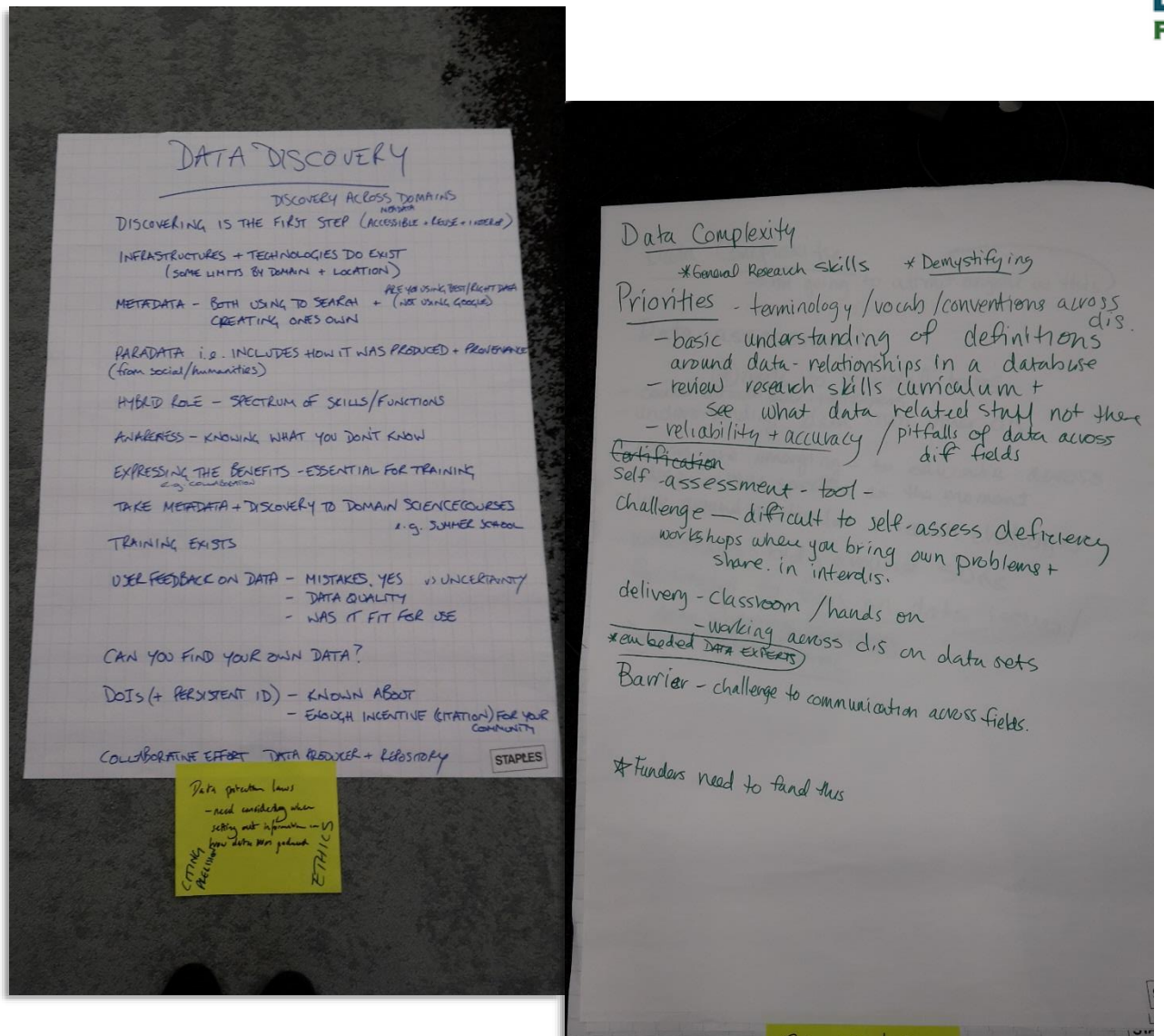
Largest data use challenge



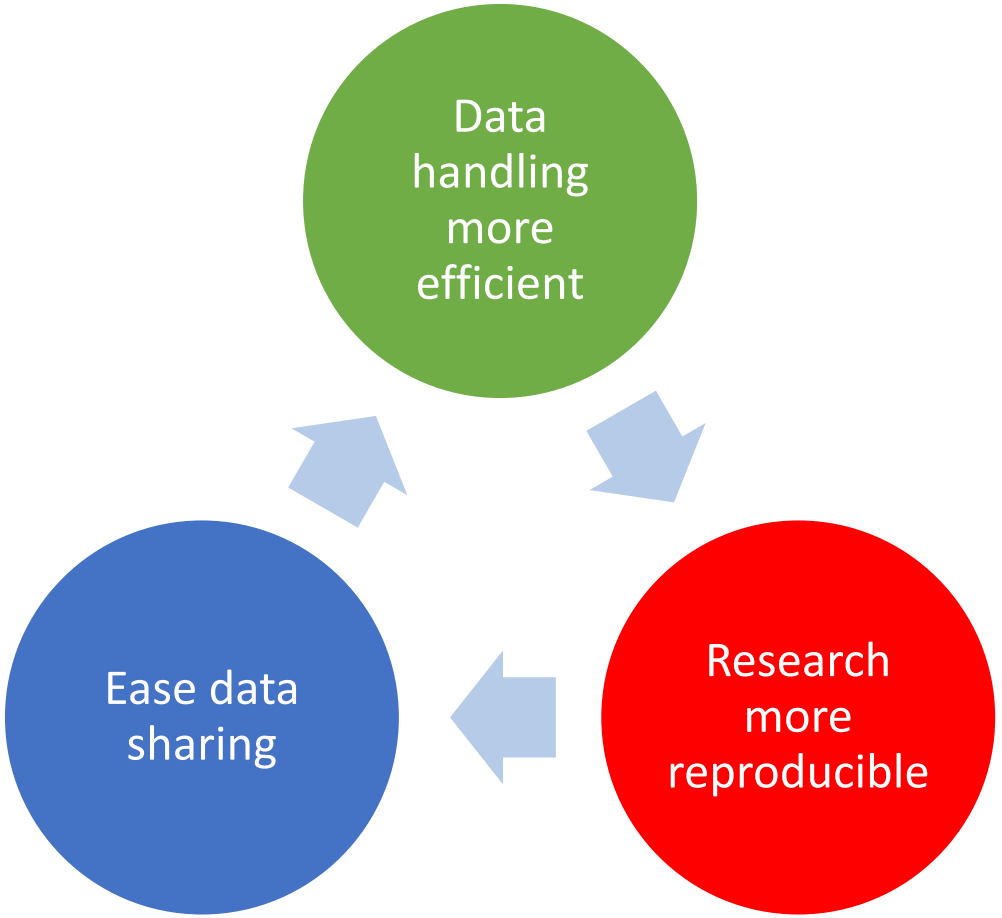
Needing most improvement



Workshop


DATA DISCOVERY
DISCOVERY ACROSS DOMAINS
DISCOVERING IS THE FIRST STEP (ACCESSIBLE + REUSE + INTERED)
INFRASTRUCTURES + TECHNOLOGIES DO EXIST (SOME LIMITS BY DOMAIN + LOCATION)
METADATA - BOTH USING TO SEARCH + CREATING ONES OWN
PARADATA i.e. INCLUDES HOW IT WAS PRODUCED + PROVENANCE (from social/humanities)
HYBRID ROLE - SPECTRUM OF SKILLS/FUNCTIONS
AWARENESS - KNOWING WHAT YOU DON'T KNOW
EXPRESSING THE BENEFITS - ESSENTIAL FOR TRAINING
TAKE METADATA + DISCOVERY TO DOMAIN SCIENCECOURSES
TRAINING EXISTS
USER FEEDBACK ON DATA - MISTAKES, YES vs UNCERTAINTY
- DATA QUALITY
- WAS IT FIT FOR USE
CAN YOU FIND YOUR OWN DATA?
DOIs (+ PERSISTENT ID) - KNOWN ABOUT
- ENOUGH INCENTIVE (CITATION) FOR YOUR COMMUNITY
COLLABORATIVE EFFORT DATA REDUCER + REPOSITORY
STAPLES
CITING
ETHICS
Data Complexity
*General Research skills
*Demystifying
Priorities - terminology / vocab / conventions across dis.
- basic understanding of definitions around data - relationships in a database
- review research skills curriculum + see what data related stuff not there
- reliability + accuracy / pitfalls of data across dif fields
Certification
Self-assessment - tool -
Challenge — difficult to self-assess deficiency
workshops where you bring own problems + share in interdis.
delivery - classroom / hands on
- working across dis on data sets
*embedded DATA EXPERTS
Barrier - challenge to communication across fields.
*Funders need to fund this

Curricula Goals



Curricula overview

Core

- programming, environmental data, visualisation, management, interdisciplinary data

Optional

- software development, object orientated, data science, data organisation

Principal Investigators

- DMPs and repositories, team skills and development

Core Modules

i. Programming for data intensive research
ii. Environmental data: expectations and limitations
iii. Visualising environmental data
iv. Data management
v. Interdisciplinary data exchange

Jon Blower – CTO & data visualisation expert at the IEA

Optional Modules

CODATA-RDA Research Data Science Summer School 2017
Course co-director Hugh Shanahan & delegate Shaily Gandhi

- Software development
- Object-orientated programming
- Data science topics
  - Databases
  - Machine Learning
- Data organisation
  - Workflow
  - Code sharing facilities

Principal Investigators

DMPs and data repositories

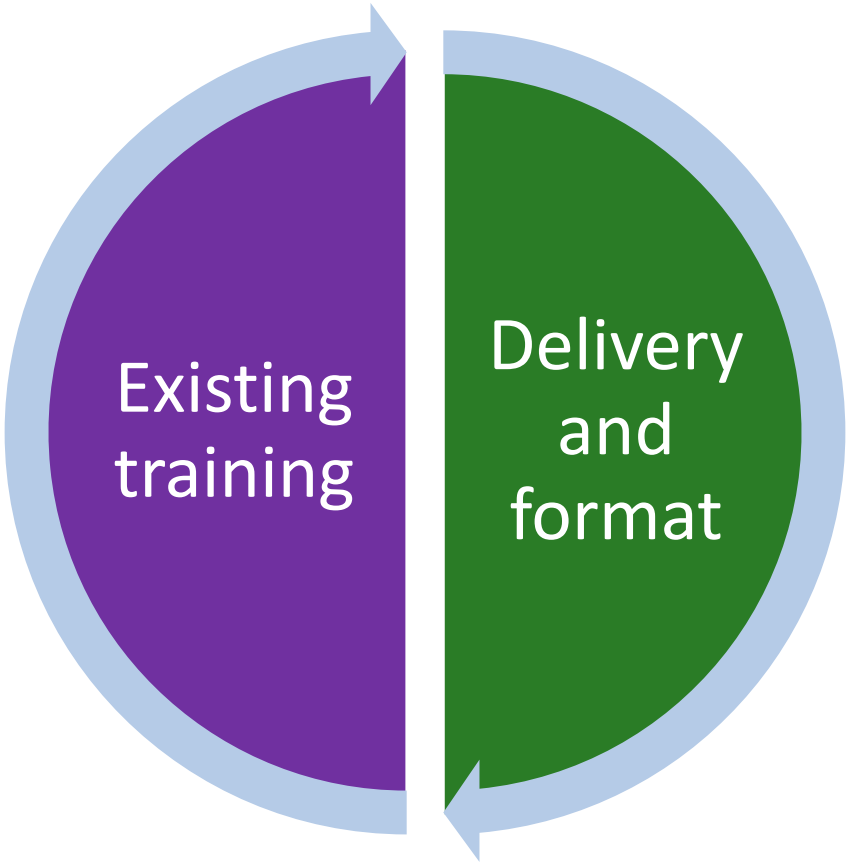
- Prioritising data management
  - Funder requirements
  - DMPs as living documents
  - Resourcing

Skills for data intensive research

- Team roles
  - Software engineering is a specialism
  - How to recognise contributions of ‘non-publishers’

Curricula context

- Delivery
  - Conference training slots
  - Messaging that skills are valued
  - Emphasis on benefits
  - Real-life datasets
  - Online is convenient
  - Face-to-face for interdisciplinary
- Existing training
  - Much on data management
    - Application to be encouraged
  - Less on PI, interdisciplinary data and environmental data



Example course

- Open online course, launch June 2018
- 20 hours of content
- Science PhD students and early career researchers
- Face-to-face promotional workshops June/July
- NERC funded

Data Management

1. CONTEXT
2. PRACTICALITIES
3. NERC SPECIFICS

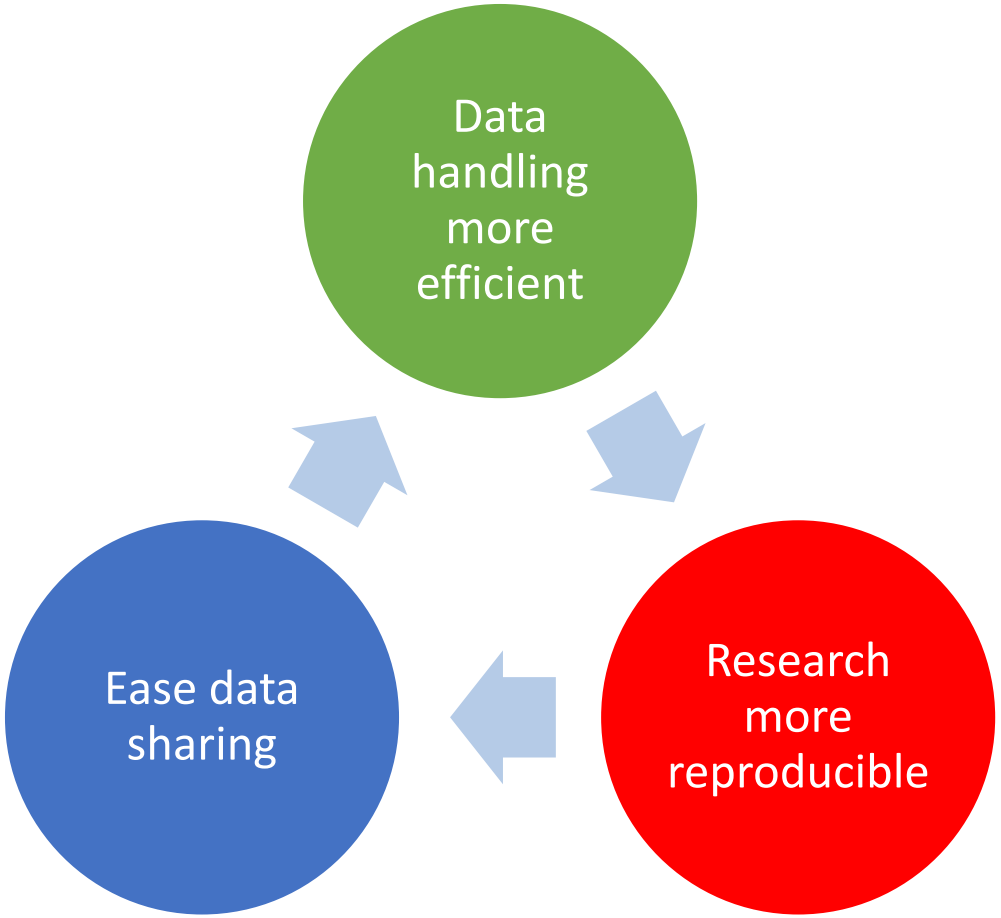
Data Application

4. VARIABILITY
5. VISUALISATION

End-Users

6. POLICY
7. BUSINESS
8. MEDIA & PUBLIC

Any questions?



- Background
  - Skills gap analysis
  - Workshop outcomes
- Curricula
  - Goals
  - Structure
  - Example course
- Q&A

v.lucas@the-iea.org