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-l&DM Human Dimensions

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

- Data complexity: 20.1%
- Data volumes: 19.5%
- Finding relevant existing data - knowing what’s out there: 16.5%
- Lack of data standards and exchange standards: 17.1%
- Dealing with multiple data types: 9.1%
- Data access and file transfer: 9.1%
- Data management and storage: 16.5%
- Other: 9.1%
Needing most improvement

<table>
<thead>
<tr>
<th>Issue</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reluctance to share data or models</td>
<td>87</td>
</tr>
<tr>
<td>Lack of computational and numerical analysis skills</td>
<td>69</td>
</tr>
<tr>
<td>Lack of awareness of relevant and potential data sources</td>
<td>67</td>
</tr>
<tr>
<td>Poor programming style or data analysis workflow</td>
<td>67</td>
</tr>
<tr>
<td>Lack of well-maintained libraries and codes</td>
<td>53</td>
</tr>
<tr>
<td>Inability to process large amounts of data</td>
<td>52</td>
</tr>
<tr>
<td>Overuse of spreadsheets</td>
<td>52</td>
</tr>
<tr>
<td>Lack of multi- and interdisciplinary collaborations</td>
<td>43</td>
</tr>
<tr>
<td>Poor compliance with data management requirements</td>
<td>39</td>
</tr>
<tr>
<td>Inefficient numerical analysis, no awareness of advances</td>
<td>36</td>
</tr>
<tr>
<td>Reliance on old fashioned languages or legacy systems</td>
<td>32</td>
</tr>
<tr>
<td>Other</td>
<td>16</td>
</tr>
</tbody>
</table>
Workshop

DATA DISCOVERY

Discovery is the first step (where you go)

Innervation - techniques in use

IMPROVE - both local to system + the skill chain

AGENDA - every week we have a problem or challenge

ARTIFICIAL INTELLIGENCE

MADE IN - creation of skills/competencies

ARTICULAR - knowing what you want to know

EXPLAIN THE SKILLS - essential for training

DATA MANAGEMENT: Know your data

INP_ICHOSE TO LEARN

DATA QUALITY

CAN YOU FIND YOUR OWN DATA?

DOES (+ RESIDENCE) - is
can you ask?

DATA COMPLEXITY

Competencies + skills + understanding

Priorities:
- Terminology & vocab common across all
- Basic understanding of definitions
- Real data relationships in a database
- Social research skills curriculum
- What data related skills not there
- Reliability + accuracy + pitfalls of data sources & fields

Self assessment - test - challenge - difficult to self-assess confidence

Workshop: when you bring own problem & share in interests.

Deliver - classroom/hands on

Barriers: access to data sets

Funders need to find data.
Curricula Goals

- Data handling more efficient
- Ease data sharing
- Research more reproducible
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

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

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

- Background
  - Skills gap analysis
  - Workshop outcomes

- Curricula
  - Goals
  - Structure
  - Example course

- Q&A

v.lucas@the-iea.org