A Data Science Preparatory Course (?)

Wolfram Horstmann

State and University Library Göttingen – Director
Humboldt University Berlin – Lecturer

INTRODUCTION

Background

„Future graduates must move securely in the digital world"

- Digital Transformation of research and society poses challenges in teaching: we need to educate the future academic workforce able to tackle societal and scientific problems.
- Computer and data based methods has penetrated most areas of research: experimenting, analysis, interpretation, presentation and publishing. Knowing them is a must!
- Researcher of all disciplines can access all kinds of data and information, digital and analog, from everywhere, everytime, with minimal effort, share with others and securely preserve. Only skilled graduates are competitive.

Rationale

1. Data Science is a multidisciplinary effort
2. Skills need to draw from multiple sources
3. Elements and modules distributed in faculty
4. Curricular integration needs to be flexible

>>

Prep-Course w/ polyvalent structure is needed

>>

„PROPAEDEUTIKUM DIGITALE"

Disclaimer – all this is still conceptual but already in preparation (Campus Göttingen)

Proaedeutikum Digitale

- The Prep-Course as a polyvalent toolbox for elements, sessions and modules that can be re-used in multiple application scenarios
  - Admission criterion for Bachelor/Master/PhD
    - Similar to Medicine / ‚Latinum' / Math
  - Integration in existing Bachelor/Master/PhD
    - As part of curriculum or as additional modules
  - Application in High Schools
  - Application in Services – **Libraries, IT Services, Enterprises**

EXAMPLE CONTENTS

Example contents: Basic

- Secure handling of essential digital tools and methods
  - e.g. storage media, productivity/office, eMail etc.
- Good research practice
  - e.g. citation, anti-plagiarism
- Standards in information practice
  - e.g. desk research and literature management
- Data protection, cybersecurity, IPR, Ethics
- Introduction to subject-specific tools
  - depending on subject, to give a 'flavour' of diversity

Example contents: Advanced

- Good scientific practice based on subject specific examples
- Data analysis software
  - e.g. Excel, SPSS, R
- Essentials of Data Management
  - e.g. documentation, repositories, metadata, Git, Confluence
- Essentials concepts in computer science
  - e.g. databases, programming, algorithms
- Electronic publishing and Open Access models
- Societal impact of digitalisation (of research)
- Method Introductions
  - e.g. Simulation, Visualisation, Mining

Example contents: Expert

- Digital research methods in disciplines, e.g.
  - statistics with R
  - simulation and visualization with MatLab
  - interview analysis with AtlasTI
  - digital editions with TextGrid
- Professional Research Data Management
  - e.g. encoding, documentation, preservation
- Project and service planning and management
  - e.g. ITIL, Prince, ‚Agile'
- Research and Information Infrastructure
  - e.g. Virtualization/Cloud, Genbank, VO's

Example Contents: A Full Prep-Course

- Basic knowledge and skills in data management
  - Good scientific practice in terms of data management
  - Funders policies and data management plans
  - Data protection, ethics, anonymization, cybersecurity, IPR
  - Introductions: database, search/indexing technology, APIs
  - Data documentation, metadata, ontologies, linked open data
  - Electronic publishing and open access models
  - Conets of programming, algorithms and data structures
  - Typical application contexts
    - statistics with R, simulation with MatLab etc (see above)

APPLICATION SCENARIOS

Proaedeutikum Digitale

- The Prep-Course as a polyvalent toolbox for elements, sessions and modules that can be re-used in multiple application scenarios
  - Admission criterion for Bachelor/Master/PhD
    - Similar to Medicine / ‚Latinum' / Math
  - Integration in existing Bachelor/Master/PhD
    - As part of curriculum or as additional modules
  - Application in High Schools
  - Application in Services – **Libraries, IT Services, Enterprises**

Admission criterion for Bachelor/Master/PhD

- *If you want to study XYZ, you [must | are requested to | should consider to] have ABC!*
- Before studying or as bridge course, e.g. to change after bachelor level to computer science or to be admitted to grad school
- Possible Examples
  - Bioinformatics
  - Digital Humanities
  - Quantitative Economics or Social Sciences
  - ...

Integration in Existing Curricula

- Diverse scenarios and durations – **1 hour < x > 1 semester**

  - As regular seminar / module

  - Individual session in a module

  - As an additional (non-compulsory) module

  - Recommended Trainings, e.g. in graduate schools

Application in High Schools

Open for discussion

Application in Services

- Libraries, IT Services, Enterprises might re-use but require different routes
  - Training: a plethora of services
  - Expert recruitment: employing skilled staff
  - Learning-on-the-job: engage in projects
  - Online Learning: e.g. MOOCs
  - Degrees: iSchools and others
- cf. „How to maximize research data skills in Libraries" Resarch Data Alliance **– Library BoF**

CONCLUSION

Conclusion

- Future graduates must move securely in the digital world

- Data Science practice and teaching is already existing, but ‚patchy'

- A Prep-course is a good way to weave the currcular fabric – if(f) it is modular and flexibly re-usable in multiple application contexts