Towards an Empirically Founded Data Literacy Competency Model

Andreas Grillenberger & Ralf Romeike
Computing Education Research Group, Friedrich-Alexander-Universität Erlangen-Nürnberg

---

Data as (Re-)Emerging Topic of Computer Science and other Sciences

**Computer Science**
- data has been a fundamental area of computer science from its beginnings
- often, mainly the role of data in databases and data structures were considered
- big data, data management and data science gave the topic “data” new impetus
- in recent years, both science and practice put increasing emphasis on data

**Data Management**
- focuses on the static aspects of storing and handling data
- model of key concepts of DM:
  - acquisition
  - cleansing
  - modeling
  - optimization
  - analysis
  - visualization
  - evaluation
  - sharing
  - archiving
  - measure

**Data Science**
- emphasizes the dynamic aspects of handling and processing data
- is an uprising topic of (general and higher) CS education
- gives insight into the use of data
- fosters an understanding of the potential that data analysis involves

**Data Literacy**
- particularly involves aspects of CS and mathematics
- is related to several concepts of databases, data management and data science
- needs to be distinguished from information literacy
- is necessary for everyone and in every discipline when handling and processing data

---

Current Research on Data Literacy and Related Fields

**Data Literacy Competencies**
- Ridsdale et al. (2015) built a competency model by analyzing best-practice approaches
- these approaches cover an interdisciplinary perspective
- as data literacy includes various CS concepts, investigating this field from a CS perspective gives additional insights and foundation

---

Developing a Competency Model: Approach

- a data literacy competency model can be developed by theoretically deriving competency areas from existing work
- empirical data on the contents of data science and the key concepts of data management can serve as a basis
- dividing the model into process and content areas is suitable for considering both perspectives on this topic

---

Results & Exemplary Competencies

```
<table>
<thead>
<tr>
<th>Content Areas</th>
<th>Process Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>data and information (C1)</td>
<td>gather, model and cleanse (P1)</td>
</tr>
<tr>
<td>data storage and access (C2)</td>
<td>implement and optimize (P2)</td>
</tr>
<tr>
<td>data analysis and machine learning (C3)</td>
<td>analyze, visualize and interpret (P3)</td>
</tr>
<tr>
<td>data ethics and privacy (C4)</td>
<td>share, archive and erase (P4)</td>
</tr>
</tbody>
</table>
```

C1/P1: verify if captured data appropriately represents the original information
C2/P1: structure gathered data in a way suitable for storing them
C2/P2: compress data to increase storage efficiency
C3/P3: visualize the results of data analyses
C4/P2: implement specific data analysis methods using appropriate tools
C4/P4: discuss data storage for further uses from an ethical perspective
C4/P3: evaluate potential ethical issues raised by data analyses

The process and content areas can be combined in different ways and hence cover various competencies

---

Future Work

- evaluate the competency model and its competencies
- divide the model into several levels representing different target groups and respective competency goals from “generally everyone” over “every scientist” to “data professionals”
- implement interventions that strive for fostering data literacy and evaluate them