DaLiCo Dimensions

Authors:
Christine Gläser – HAW HH
Ulrike Spree – HAW HH

Co-funded by the Erasmus+ Programme of the European Union

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

DaLiCo Dimensions

PROJECT INFORMATION

PROJECT TITLE:
Data Literacy in Context

PROJECT ACRONYM:
DaLiCo

PROJECT NUMBER:
2019-1-DE01-KA203-005066

ERASMUS+ PROGRAMME:
KA2 Partnerships for Innovation and the Exchange of Good Practices

WEBSITE:
www.dalico.info

PARTNERS:

HAW HAMBURG
HOGESCHOOL UTRECHT UNIVERSITY OF APPLIED SCIENCES
UNIVERSITY of DEBRECEN
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Co-funded by the Erasmus+ Programme of the European Union

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

DaLiCo Dimensions

AUTHORS

Christine Gläser (HAW Hamburg, Faculty of Design, Media, Information, Dept. Information)
Ulrike Spree (HAW Hamburg, Faculty of Design, Media, Information, Dept. Information)

CONTRIBUTORS

Kristin Ameis (HAW Hamburg, Faculty of Design, Media, Information, Dept. Information)
J. Alberto Conejero (Universitat Politècnica de València, Director of the Dep. of Applied Mathematics)
Judit Fazekas-Paragh (University of Debrecen and National Library, Head of Department at Education and Research Support Department)
Hanna Käfer (HAW Hamburg, Faculty of Design, Media, Information, Dept. Information)
Guido Ongena (Hogeschool Utrecht University of Applied Sciences)
Rüdiger Weißbach (HAW Hamburg, Faculty of Business and Social Sciences)

ABSTRACT

DaLiCo Dimensions is a suggestion for basic dimensions for the description of generic data literacy competences in the field of higher education. The dimensions address 6 core competencies (Data identification, Data understanding, Data use, Data reflexivity, Data managing, Data communication) roughly oriented towards the data life cycle.

DATE: 31.10.2022

This document was written by partners of the "DaLiCo" project and is published as open educational resource. The project was funded through the European Commission. The content of this document exclusively reflects the opinions of the authors. The EU Commission and the National Agency at the DAAD assume no liability arising from the use of the information presented here.

LICENCE

2022. This work is licensed under a
Creative Commons License Attribution 4.0 International CC-BY-SA.

Co-funded by the Erasmus+ Programme of the European Union

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Contents

INTRODUCTION 1

DEVELOPMENT OF THE OUTPUT 1

COLLECTING QUESTIONS DALICO MAP SHOULD ANSWER 1
EXPERIMENTING WITH DEVELOPING A DATA LITERACY ONTOLOGY 2
REVIEWING EXISTING DATA LITERACY FRAMEWORKS 3

STRUCTURE AND FEATURES 5

TARGET GROUP 6

USAGE AND IMPACT 7

DISSEMINATION 7

SOURCES 8

Introduction

The DaLiCo Dimensions described in this result description are part of IO1. The goal of IO1-The DaLiCo data literacy map (DaLiCo map) is to provide a survey of existing data literacy resources and activities with a focus on the four DaLiCo partners in order to make these resources discoverable and accessible. Included are physical and virtual infrastructure (like labs, digital libraries), resources (e-learning/hybrid modules or courses) and tools.

The DaLiCo map consists of the following interconnected deliverables,

1. a glossary of data literacy terms
2. a suggestion of dimensions for the description of data literacy competencies **(DaLiCo Dimensions described in this result description)**
3. a collection of literature references of useful data literacy resources assembled with the reference management software Zotero. The resources are indexed with keywords from the DaLiCo Glossary and assigned to one or more facets from the DaLiCo Dimensions
4. The integrated tool Dalico App makes the DaLiCo Map available online.

So far, there exists no unifying and accepted definition of neither the concept of data literacy nor about the competencies a data literate person should have. Different communities of practice use slightly different terminologies (Spree & Gläser 2022; Ongena 2022). Within the ERASMUS+ Project DaLiCo we saw the need to define core competencies a data literate person in the field of higher education should have and the terms (labels) used to describe them. For this purpose the DaLiCo Dimensions describe six interconnected dimensions of beeing literate with data.

Development of the Output

The development of a common understanding of data literacy within the DaLiCo Team took place in several overlapping working steps.

Collecting questions DaLiCo map should answer

Within the project proposal the scope of the DataLiteracy Map is described in general terms as “to provide a survey of existing data literacy resources and activities with a focus on the four DaLiCo partners in order to make these resources discoverable and accessible“ (Project proposal, p. 83). With the aim of narrowing down the prior understanding and the expectations of the partners of the field of data literacy the partners on the occasion of one of the regular international online meetings collected questions the map should answer and rated them according to importance. Quite a few questions showed the need to clearly define data literacy – as understood by the partners - and at the same time to make the references and overlaps with other literacies and fields of competence comprehensible. (What is the common understanding of Data Literacy shared by the partners? Which stakeholders use which definition of Data Literacy? Where can I get help on a specific data literacy task? What are the main data literacy tasks? What generic and specific competences do data literate university graduates need? Where can I find re-usable material focusing on special competencies? What distinguishes Data Science from Data Literacy?)

Experimenting with developing a data literacy ontology

In order to arrive at a common understanding the initial idea was to create an ontology to conceptualize the domain of data literacy bringing together application areas, academic domains, data literacy competences as well as competence levels and teaching levels. We had the expectation that we could reuse and supplement existing ontologies. First experiments (by students of a Hamburg masterclass on Data Literacy and Open Science in the ontology-tool Protégé showed that until now there exist no reusable formal ontologies for data literacy. The Data Science Ontology (2022) is still under construction and focusses on a specialized data science related choice of concepts. The master students experimented with modelling the domain based on the structure suggested by Ridsdale (2015). While modelling the Ridsdale matrix proofed well suited to get a rough overview of the competence field, but difficult when it comes to delineating competences from each other without overlaps. For example, the distinction between Data Ethics and Identifying Problems Using Data was not unequivocal to us. In total, the Ridsdale categories did not proof suitable for filtering purposes because they are compound categories.

Figure 1. Early version of DaLiCo ontology experiments in Protégé.

Reviewing existing data literacy frameworks

In addition to Ridsdale (2015) further frameworks were scanned for transferability (Gläser & Spree 2022, Ongena 2022). The actual phrasing of the dimensions is based on a study by Guido Ongena (DaLiCo partner from Utrecht) on employee's data literacy. The DaLiCo dimensions are based on a small number of frequently received frameworks on data literacy (DAMA-DMBOK 2017, Ridsdale, et al. 2015, Sternkopf and Mueller 2018, Schüller 2020, Pangrazio and Selwyn 2019). Following comparative analysis and evaluation we propose 15 skills which jointly constitute data literacy. To capture prevalent skills – understood as the ability to do something well (Attwell 1990) – related to dealing with data we use verbs. We grouped the 15 individual skills into 6 more general concepts in order to develop higher-order constructs.

Figure 2. Visualisation of DaLiCo Dimensions created with Skos-Play

The original version presented by Ongena consisted of five dimensions focusing on competencies directly linked to generating value from data (Ongena 2022). For the purpose of the DaLiCo project data managing was added as a sixth dimension. The subordinated skills are based on the DAMA-DMBOK (2017). The description of the dimensions is partly taken over verbatim (albeit abbreviated from Ongena 2022).

Data identification

A first step in turning data into value is to discover and to obtain relevant data. One should be able to ***find*** this data, what it describes, in which systems it is collected, and what type it is. In addition to its discovery, it is imperative to know how to access the data and to import it into applications that are suitable to explore the data. To support the exchange of data between devices filetypes like XML or JSON need to be understood. These formats are specifically used when the data is made accessible through API's which makes it possible for an organization to ***get*** this external data.

Data understanding

Assessing the data quality and cleaning and preparing, sometimes 'dirty', raw data is an important task in data analytics. Data quality problems are present in single data collections due to misspellings during data entry, missing information or other invalid data. The task becomes more complex and pressing when several sources are combined. It becomes necessary to ***clean*** the data. The neglect of proper data cleaning can result in inaccurate analytics and unreliable decisions. The importance of data cleaning has increased by the renaissance of big data analytics. Cleaning the data is intertwined with ***analyzing*** the data as exploring the data has to be done first in order to determine actions for cleaning the data (e.g., which records or cases should be deleted or altered). This is often referred to as exploratory data analysis that is used to increase insight into the data and to spot anomalies.

Data use

According to Schüller data literacy is the cluster of all efficient behaviours and attitudes for the effective execution of all process steps for creating value or making decisions from data (Schüller 2020). Value is created by ***defining*** research (or business) questions to the data. The outcomes of the data analyses need to be evaluated against business or project goals. In other words, they need to be ***verified***. Data initiatives are relevant to the IT departments as well as to the remaining functional organizations. Cultivating data literacy, or ***embed*** data within the organizational culture, is the prerequisite of an organization, be it a public university or a private company, to becoming data literate. This also includes the idea that the use of data should be supported by the management and is seen as an enabler or opportunity and not as a threat.

Data communication

Using data to communicate a message to a particular audience (Bhargava & D'Ignazio 2015) is an integral part of data literacy definitions. This means that a data literate person should be able to ***present*** data analysis outcomes in a comprehensive way to relevant stakeholders. Turning data insights into a narrative and communicating this narrative to an audience is often linked to processes like storytelling. Storytelling presumes a controlled delivery or presentation of information and often goes hand in hand with visualizations in order to convey and support the message. The successful communication of (especially quantitative) data presupposes having skills to create and comprehend all kinds of visualizations like mapped data, graphs, pie charts, thus the ability to ***visualize*** the data, becomes an important skill that allows to navigate data sets and supports data understanding.

Data reflexivity

The frameworks we looked at in one way or another address attitudes and values as well as an awareness for the consequences the utilization of data driven technologies will have on a particular business or organization or the society in general (Schüller 2022, p. 19 ff.; Pangrazio & Selwyn 2019, Ridsdale 2015). To characterize the associated abilities, attitudes and values terms like critical thinking and ethics are used. For the purpose of the DaLiCo dimensions we delineated the dimension data reflexivity. Data reflexivity involves examining one's own

judgments, practices, and belief systems when utilizing data. The goal of being reflexive is to identify any personal beliefs that may have affected the gathering, processing and interpretation of data. With regard to data this pertains to the question what the implications are of the utilization of data. Data reflexivity is demonstrated by ***assessing*** data in terms of interpreting data, outputs of data analyses and results confidently and critically on the basis of internalized explicit evaluation criteria. It also demands for ***ethizising***, thus the conscientious and responsible consideration of the impacts of data use based on appropriate ethical decision models and defined guidelines for responsible data handling.

Data managing

Data managing comprises the activities related to a responsible and sustainable management of data along the data-life-cycle. We distinguish four skills. Data and information need to be gathered and ***collected*** systematically and methodologically. To allow for easy transfer and (re)using data ***creating and using metadata*** secures the suitable packaging and labelling of data. This is done by assigning appropriate descriptors to data or datsets and applying and assessing appropritate data organization requirements like sorting and classifying to make the data useful. ***Storing*** is the ability to archive data in electromagnetic or other forms for use by a computer or device. Data storing competencies include the knowledge on retention schedules, technical storing requirements, accessiblity and sharing needs. Data ***curation*** is the supervision, organization and integration of data collected from various sources. It includes all the processes needed for systematic and controlled data creation, maintenance, and management. In an academic context, data curation in the sense of research data management refers to the process of extraction of important information from scientific texts, such as research articles by experts, to be converted into an electronic format.

Structure and Features

For the implementation, the dimensions are integrated into the DaLiCo glossary on the Tematres platform. The meaning of the six dimensions as well as the 15 skills is explained by scope notes. To add further value, the dimensions are mapped to equivalent or similar frames (related frame (RTRF) in the Ridsdale matrix (2015) and in the Future Skills (2020) framework. Where suitable, glossary terms from the Basic concepts are indicated as related terms.

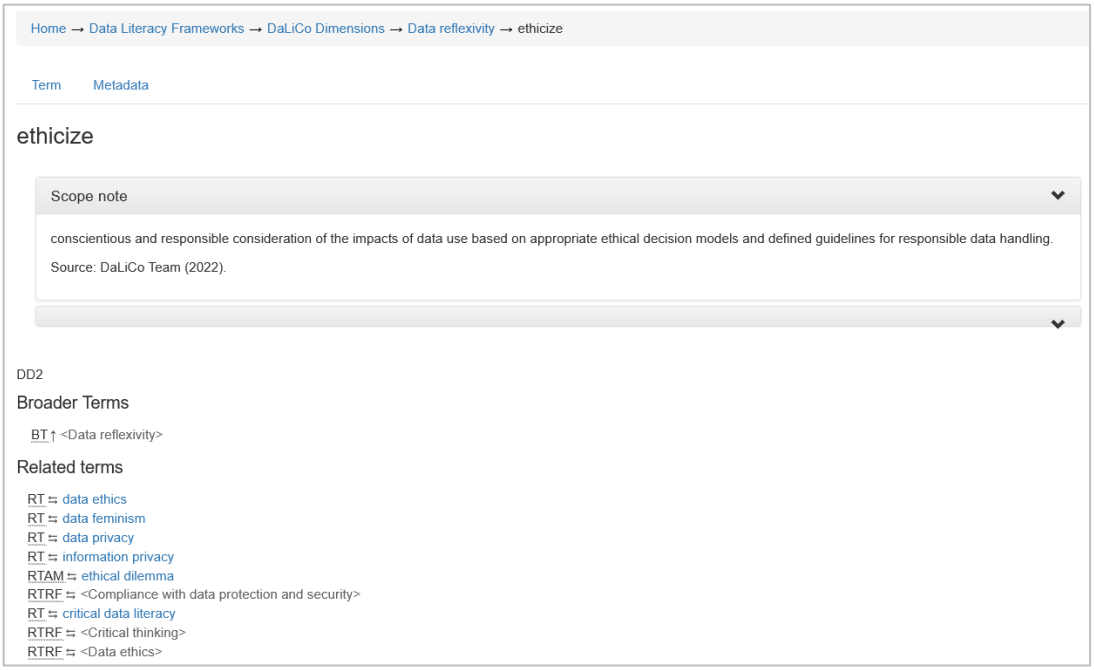

Figure 3. Full entry of the skill ethicize in the DaLiCo Glossary. Online: https://www2.bui.haw-hamburg.de/tematres/vocab/index.php?tema=1975&/ethicize

Target group

Primarily the DaLiCo Dimensions serve the internal purpose of vocabulary control for classifying the resources collected in DaLiCo References (detailed description in DaLiCo References).

The DaLiCo Dimensions are integrated into the DaLiCo Glossary and are published as a stand-alone application and may be re-used by multipliers in the field of data literacy education as well as for learners.

The DaLiCo Dimensions are used as a basic structure for the modules of the Data Literacy Learning Space (accessible via viaMINT (https://viamint.haw-hamburg.de/course/view.php?id=271) that have been created in IO3. They also served useful for phrasing the learning outcomes.

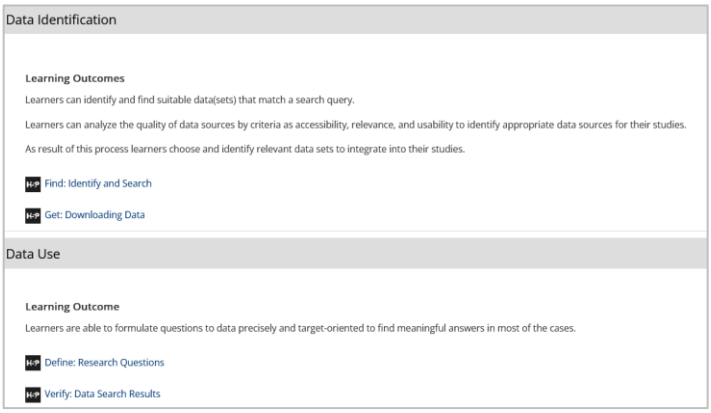

Figure 4. Structure of modules in Data Literacy Learning Space in Viamint (https://viamint.haw-hamburg.de/course/view.php?id=271)

Usage and Impact

The DaLiCo Dimensions are the backbone of the controlled vocabulary and can be used for automatically mining online resources on data literacy concepts.

One of the learnings of the DaLiCo project is that it is futile to try to establish a uniform structuring and terminology for the interdisciplinary field of data literacy. However, a mapping between existing frameworks seems a promising approach. Within the DaLiCo Glossary we already explored the mapping of categories between two influential frameworks (Ridsdale 2015, Schüller 2020). In this respect the DaLiCo Dimensions could be used for further mappings to other structuring approaches from different domains and in different languages like the (DaLi Kompetenzrahmen 2022 (in German) or the or the Framework for information literacy in Higher Education (American Library Association 2015)

Dissemination

The DaLiCo Dimensions will also be made accessible via the DaLiCo App. Here they are used as a means for searching and filtering the resources.

They will be further disseminated via various internal (Komweid 2022) and external communication channels like academic journals (IWP, BuB) and networks (like the Data Literacy Education Netzwerk https://www.stifterverband.org/data-literacy-education#netzwerk).

Sources

Attwell, Paul (1990). What is skill? *Work and Occupations* , 17(4), 422-447. Online: https://journals.sagepub.com/doi/abs/10.1177/0730888490017004003

Bhargava, R., & D'Ignazio, C. (2015). Designing Tools and Activities for Data Literacy Learners. *Wed Science: Data Literacy Workshop.* Oxford, UK.

DaLi Kompetenzrahmen (2022). Online, Retrieved: 2022-12-01: https://www.th-koeln.de/informations-und-kommunikationswissenschaften/data-literacy-basismodul_88414.php#bottomcontext

American Library Association. (2015, February 9). *Framework for Information Literacy for Higher Education* [Text]. Association of College & Research Libraries (ACRL). https://www.ala.org/acrl/standards/ilframework

Gläser, C., & Spree, U. (2022). Finding Access Points for Data Literacy: The Example of the ERASMUS+ Project DaLiCo (Data Literacy in Context). In *European Conference on Information Literacy* (pp. 109-121). Springer, Cham.

Henderson, D., Earley, S., & Data Administration Management Association (Eds.). (2017). *DAMA-DMBOK: Data management body of knowledge* (Second edition). Technics Publications. https://www.dama.org/cpages/body-of-knowledge

Data Science Ontology (2022). IBM Research AI; Stanford University Statistics (ed.). Online: https://www.datascienceontology.org/about. Retrieved: 2022-11-28

Francart, Thomas; SPARNA (2022). Skos-Play. Version 0.9.1. CC-By-SA. Online: https://skos-play.sparna.fr/play/upload. Retrieved: 2022-11-28

Ongena, Guido (2022). Data literacy for improving governmental performance: A competence-based approach and multidimensional operationalization, In: Digital Business, 2022, 100050, https://doi.org/10.1016/j.digbus.2022.100050.

Pangrazio, L., & Selwyn, N. (2019). Personal data literacies : A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society, 21*(2), 419-437.

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., . . . Matwin, S. W. (2015). *Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report.* Halifax: Dalhouse University

Schüller, K. (2020). *Future Skills: A Framework for Data Literacy – Competence Framework and Research Report*. Working Paper Nr 53. Hochschulforum Digitalisierung, Berlin

Sternkopf, H., & Mueller, R. (2018). Doing Good with Data:Development of a Maturity Model for Data Literacy in Non-governmental Organizations. *Proceedings of the 51st Hawaii International Conference on System Sciences*, (pp. 5045–5054). Waikoloa Village, Hawaii