DaLiCo Glossary

Author:
Ulrike Spree – HAW Hamburg

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 GLOSSARY

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 Glossary

AUTHOR

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

CONTRIBUTORS

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

ABSTRACT

DaLiCo Glossary (Data Literacy in Context Glossary) is a collection of relevant key concepts in the field of data literacy (education) developed in cooperation with the partners from the ERASMUS+ Project "Data Literacy in Context" (DaLiCo) (https://dalico.info/about/). It is structured as a thesaurus following the DIN ISO 25964-1:2011).

DATE: 31.10.2022

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

The DaLiCo Glossary described in this result description is part of the Intellectual Output One – The DaLiCo data literacy map (DaLiCo map) (IO 1) The goal of IO1- 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 constitutes the framework for the activities within DaLiCo. It uncovers existing gaps and bridges them by offering context. The DaLiCo map links existing competencies/personal expertise, resources, projects, and possible scenarios as well as good practice examples.

The DaLiCo map also supports the knowledge management of the DaLiCo project team. (University of Applied Sciences Hamburg 2019, p. 80 ff.)

The DaLiCo map consists of the following interconnected deliverables,

1. a glossary of data literacy terms (DaLiCo Glossary described in this result description),
2. a suggestion for a facetted structure for the description of data literacy competencies (for details refer to IO1-Result Description DaLiCo Dimensions)
3. an online accessible reference data base that provides access to a curated collection of existing data literacy resources. The materials developed in the course of the DaLiCo in Context project are also made accessible via the collection. The resources are indexed with keywords from the glossary and assigned to one or more facets from the DaLiCo Dimensions (for details refer to IO1-Result Description DaLiCo References.)
4. The integrated tool DaLico App makes the DaLiCo Map available online.

During the work on the DaLiCo project, it became clear that there exists no uniform terminology within the data literacy communit(ies) regarding the meaning of central concepts in the field of data literacy. Instead, we experience a kind of language gap especially between agents in the field of data literacy and in the field of data science (Gläser, Spree 2022). This is also true for the DaLiCo team as members come from such different academic backgrounds as economics, pedagogy, e-learning, data science or library and information science. The collaborative development of a glossary of basic terms in data literacy education seemed an appropriate way to reach a common understanding within the DaLiCo team.

Development of the Output

The glossary was developed as a controlled vocabulary in a collaborative process. Before starting the project we explored whether we could reuse existing controlled vocabularies consulting various resources as the Bartoc register (Bartoc 2022). We identified a number of vocabularies defining concepts from the fields of data science and data literacy but none that served the intended purposes. However, existing vocabularies (NNLM National Center for Data Services (NCDS) (2022), were checked for terminology to integrate into the DaLiCo glossary as seeds. Following, the entries were added and completed by the partners in a collaborative

process. After entering a number of terms as 'seed' terms the partners were asked to provide suggestions for entries as candidates. The actual vocabulary work (deciding on preferred terms for the vocabulary, providing definitions and relating the terms as well as the editorial work) was done by the Hamburg team. The vocabulary is in English, but all entries are translated into the four DaLiCo partner languages (German, Hungarian, Dutch, Spanish).

Based on a selection of various materials (curricula, module handbooks, resource collections, existing frameworks like Dama 2017, Ridsdale 2015, Schüller 2020) with the aim of creating a suitable data model for the glossary the initially pursued approach of developing an ontology was rejected in favor of a faceted knowledge order. This decision was taken because the original idea of mapping the whole domain in the form of an ontology proved to be too complex for an ERASMUS+ project. Contrary to our assumptions there exists too little prior work (meta-ontologies, vocabularies, taxonomies) we could build on.

The glossary is structured as follows:

Based on a facet analysis considering fundamental categories we generated the following hierarchies:

Table 1: Facet structure of DaLiCo Glossary

Activity	Container used to collect academic disciplines
Academic discipline	
Agent	Container used to collect agents relevant in the field of data literacy education
Organization	
Person	
Basic Concept	Container for the actual data literacy concepts
Data Literacy Frameworks	Container for frameworks that describe data literacy competencies or skills.
D DaLiCo Dimensions	
F Future Skills	
R Ridsdale Matrix	
Document Type	
Educational Term	
Glossary_suggestion	The glossary will be continuously supplemented and expanded in a collaborative process. Suggestions for glossary candidates can be entered as glossary candidates for later inclusion in the glossary

Parallel to the development of the hierarchical structure the DaLiCo team chose a development tool. Two suitable thesaurus software tools were identified.

The two tools Xtree (Digicult-Verbund and Tematres (Tema Tres) were tested. After a testing phase a decision was made in favor of Tematres

(https://vocabularyserver.com/web/). The functionality of Xtree is more advanced but the source code is not yet completely open source which bears some risks for the sustainability of the application. In the DaLiCo context it is particularly positive that Tematres is being further developed as an open source project and that Tematres is multilingual and a Spanish documentation is also available (Tema Tres 2022). A test instance has been installed in Hamburg (https://www2.bui.haw-hamburg.de/tematres/vocab/index.php).

Since Tematres allows for the creation of metaterms, expressed by using pointed brackets, the facet structure for the presentation of data for the DaLiCo Map is also implemented directly in the DaLiCo Glossary as DaLiCo Dimensions.

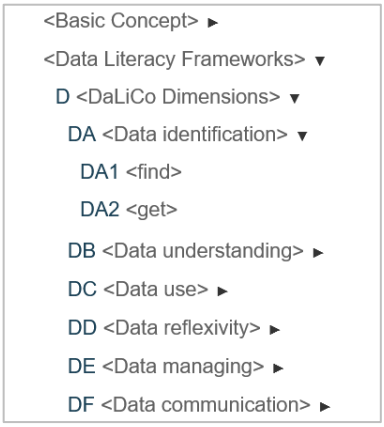

Figure 1: Embedding of DaLiCo Dimensions in DaLiCo Glossary

The structure is based on the suggestion of the data competencies used by Guido Ongena (Ongena 2022) in IO 4 for the Data Literacy measurement assessment instrument (for more details refer to IO4 Result Description).

Structure and Features

The DaLiCo Glossary (Data Literacy in Context Glossary) is a collection of relevant key concepts in the field of data literacy (education). It is structured as a thesaurus following the recommendations of the international Norm Thesauri and interoperability with other vocabularies - Part 1 (ISO 25964-1:2011). The thesaurus draws on the following keys and abbreviations to denote relationships between terms:

- <>: Indicates that this term is a “meta-term” meaning it is only used for hierarchical purposes. Deviating from the Thesaurus norm the meta-terms below are used to assign references to facets.

- BT: Broader Term – Indicates the “parent” of the term, in the hierarchical tree structure.
    - BTG: Broder Term Generic - is used when a generic is_a relation between the "parent" of the term exists. The generic relationship is the link between a class or category and its members or species.
- NT: Narrower Term – Indicates the “child” of the term, in the hierarchical tree structure.
    - NTG: Narrower Term Generic - is used when a generic is_a "child" of relation exists.
- RT: Related Term – Indicates any terms that are related in meaning or in scope to the term being viewed. USE: Use reference - Indicates that the current terms is "non-preferred" and that it should not be used for indexing purposes. UF: Used for - references to non-preferred equivalent term(s).
    - Translations of the terms into Dutch, German, Hungarian and Spanish are referenced as specialized UF Relations. UFDE - references the German translation UFES - references the Spanish translation UFHU - references the Hungarian translation UFNE - references the Dutch translation.

A complete glossary entry consists of the following elements:

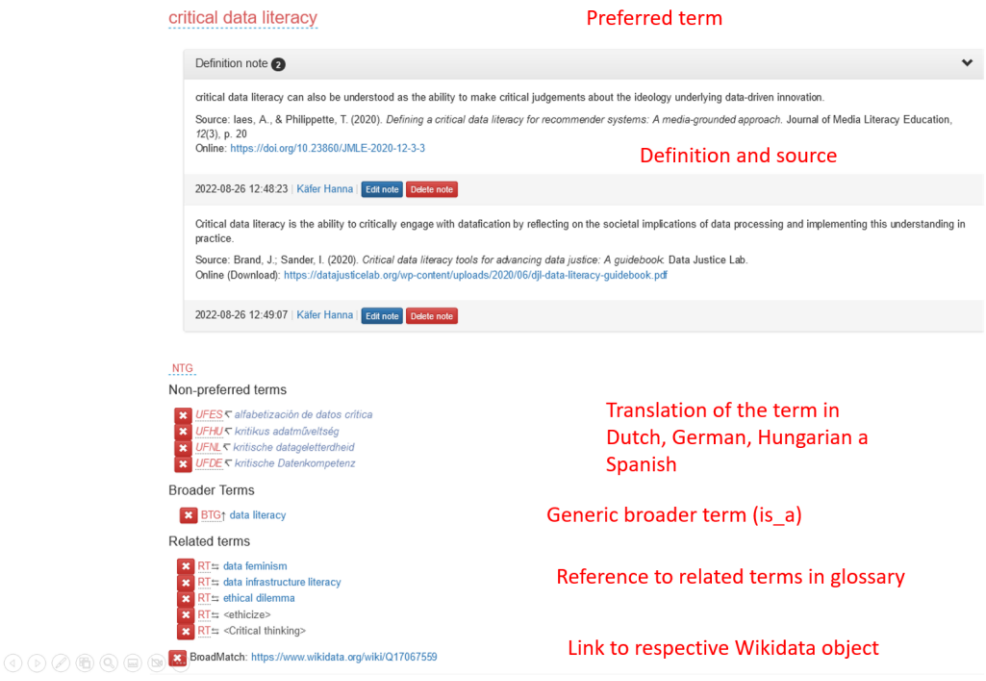

Figure 2: Complete descriptor set of DaLiCo Glossary for the entry “critical data literacy” (screenshot from DaLiCo glossary, edited by Ulrike Spree)

Target group

Primarily the DaLiCo Glossary serves the purpose of vocabulary control for indexing and classifying the resources collected in DaLiCo References (detailed description in DaLiCo References).

The glossary is also published as a stand-alone application (DaLiCo Team 2022) and may serve as a glossary for multipliers in the field of data literacy education as well as for learners to quickly look up the meaning of a term.

Usage and Impact

The DaLiCo Glossary is used to describe and index the intellectual outputs respectively the OERs like the Data Literacy Learning Space (accessible via viaMINT (https://viamint.haw-hamburg.de/course/view.php?id=271) that have been created in IO3 as well as the Train-the-trainer handbook (IO2).

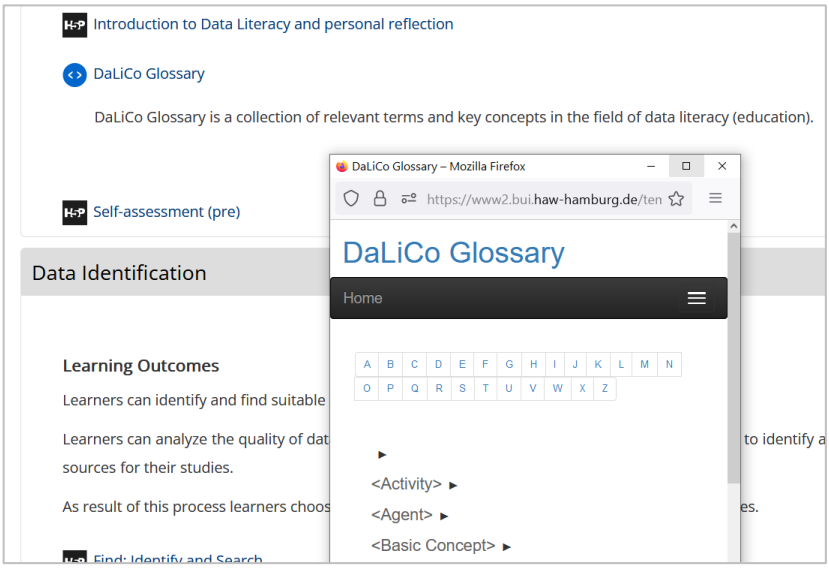

Figure 3: Integration of DaLiCo Glossary in the Data Literacy Learning Space (https://viamint.haw-hamburg.de/course/view.php?id=271)

Future developments

As the glossary is available via an open API (https://www2.bui.haw-hamburg.de/tematres/vocab/services.php) and a SPARQLE endpoint. In the future the controlled vocabulary could be used for automatically mining online resources on data literacy concepts or to enhance semantic web projects.

Dissemination

The data from the DaLiCo Glossary will also be made accessible via the DaLiCo App, directly integrated into the frontend of the web-application.

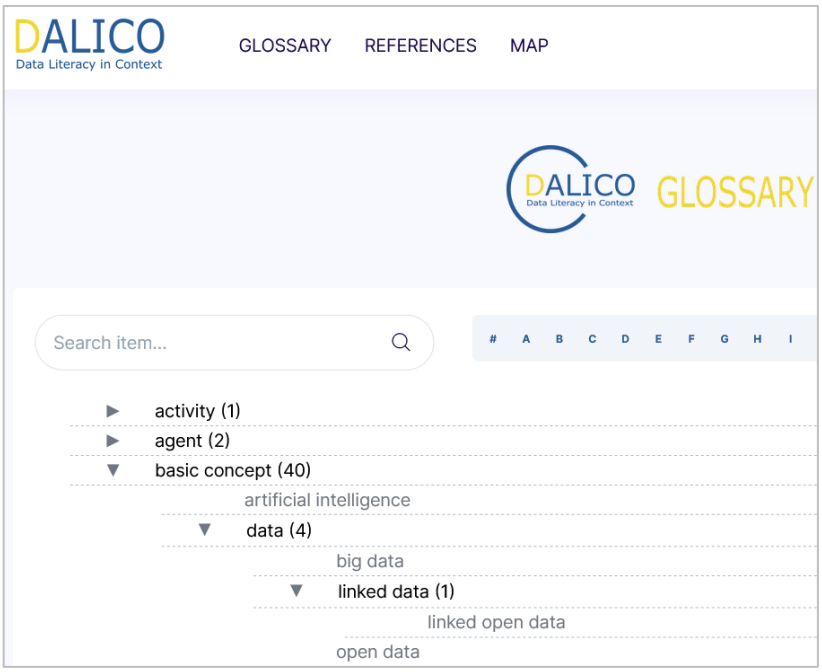

Figure 4: Prototype of DaLiCo Glossary integrated into the DaLiCo App

It will be further disseminated via various HAW-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

Bartoc (2022). *Bartoc.org. Basic Register of Thesauri, Ontologies & Classifications*. Verbundzentrale des GBV (VZG). Online: https://bartoc.org/. Retrieved: 2022-11-07

Center for Humanitarian data (2022). *Glossary*. Online: https://centre.humdata.org/glossary. Retrieved: 2022-11-07.

DaLiCo Team (2022). *DaLiCo Glossary*. Online: https://www2.bui.haw-hamburg.de/tematres/vocab/. Retrieved: 2022-11-07

Dama International (2017). *DAMA-DMBOK: data management body of knowledge*. Technics Publications, LLC.

Digicult-Verbund. https://xtree.digicult-verbund.de/

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.

ISO 25964-1:2011. Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. Beuth Verlag GmbH, 2011

Komweid (2022). Komweid (**Kom**petenzen **wei**terentwickeln im **d**igitalen Wandel (engl. developing competencies in the digital transformation) is a project of the Vice President for Studies, Teaching and Equal Opportunities, Prof. Dr. Monika Bessenrodt-Weberpals and Prof. Dr. Olga Burkova, Vice President for Digitalisation at the University of Applied Sciences Hamburg. It is funded by the Foundation for Innovation in Higher Education from 1.8.2021 to 31.07.2024 as part of the programme Strengthening Higher Education Teaching through Digitisation. Online: https://www.haw-hamburg.de/qualitaet-in-der-lehre/komweid/. Retrieved: 2022-11-22

NNLM National Center for Data Services (NCDS) (2022). Data Glossary. Online: https://www.nnlm.gov/guides/data-glossary. Retrieved: 2022-11-07

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.

Ridsdale, C. et al. (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

Tema Tres (2022). TemaTres : *Manage, Publish, Share, Reuse. The open source way to manage formal representations of knowledge*. Online: https://vocabularyserver.com/web/ Retrieved: 2022-11-07

University of Applied Sciences Hamburg (2019). *Data Literacy in Context. Application Call 2019 Round 1 KA2 - Cooperation for innovation and the exchange of good practices. KA203 - Strategic Partnerships for higher education. FormId KA203-FAEC1BD3*