A Free and Open Dataset from a Prototypical Data-driven Study Assistant in Higher Education

Johannes Schrumpf[1,*][a], Felix Weber[2,*][b], Katharina Schurz[2,*][c], Niklas Dettmer[1][d] and Tobias Thelen[1][e]

[1]*Institute of Cognitive Science, Osnabrück University, Wachsbleiche 27, Osnabrück, Germany*
[2] *virtUOS, Osnabrück University, Heger-Tor-Wall 12, Osnabrück, Germany*

Keywords: Artificial Intelligence, Dataset, Digital Study Assistant, Higher Education, Educational Recommendation Engines.

Abstract: Digital study assistants (DSAs) are an as of yet sparsely explored method to build bridges between classical, on-campus higher education and novel digital education opportunities. The DSA we present in this paper (SIDDATA) aims at supporting students to identify, reflect upon and follow their personal educational goals. Over the course of 11 months, students interacted with a prototype version 2.0 of the software, generating data about what features were interacted with, users' study-related data, and which features were deemed as useful. In this data paper, we present a preprocessed version of the DSA database for research in the domain of digital higher education. We present the data model design of the DSA and its relation to its' features. We further expand on the data extraction method used to generate the present dataset from the DSA's database. We discuss potential research paths that can be explored based on the dataset as well as its limitations

1 INTRODUCTION

Digital higher education remains an innovative and fast changing field: An ever-increasing assortment of digital education opportunities blends, but also partially clashes, with classical education formats such as lectures and seminars (Castro, 2019). At the same time, emerging technologies such as Artificial Intelligence (AI) find their way onto campuses not only as a subject but also as a tool to support and enhance student education.

Project SIDDATA (Studienindividualisierung durch Digitale, Datengestützte Assistenten, *eng: study individualization through digital, data-driven, assistants*) aims to aggregate these education opportunities and streamline them into an easy-to-use DSA for long term usage, oriented towards individual educational goals. As a collaboration project between the University of Osnabrück, the University of Bremen, and the Leibniz University of Hannover, it is part of the "Innovationspotentiale Digitaler Hochschulbildung" (*eng: Innovation Potentials in Digital Higher Education*) funding line, funded by the BMBF (Bundesministerium für Bildung und Forschung, eng: *Federal Ministry of Education and Research*).

The development of the SIDDATA digital study assistant software (referenced as SIDDATA) has occurred iteratively with an agile development approach allowing user feedback to be incorporated into the latter two of the three annual releases. Research about goals in higher education and requirements for DSAs can be found in the related project publications (Reinken & Greiff, 2021, Schrumpf et al., 2021, Thelen et al., 2019; Vogelsang et al., 2019; Weber & Le Foll, 2020; Weber et al.,

[a] https://orcid.org/0000-0002-0068-273X
[b] https://orcid.org/0000-0002-7012-3378
[c] https://orcid.org/0000-0003-3804-1134
[d] https://orcid.org/0000-0001-9944-1671
[e] https://orcid.org/0000-0002-3337-6093
* Authors contributed equally

Schrumpf, J., Weber, F., Schurz, K., Dettmer, N. and Thelen, T.
A Free and Open Dataset from a Prototypical Data-driven Study Assistant in Higher Education.
DOI: 10.5220/0011038800003182
In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022) - Volume 2*, pages 155-162
ISBN: 978-989-758-562-3; ISSN: 2184-5026

2019; Weber, 2019; Weber et al., 2021a, 2021b, 2021c; Weber & Thelen, 2021).

By collecting and processing educational resource data on the one hand and user-system interaction, on the other hand, the data collected by the DSA, version P2, throughout its eleven months of operation offers a unique opportunity to probe and analyze potential opportunities and challenges for digital higher education.

Thus, we here present data generated through the interaction between users and SIDDATA. This dataset (Weber et al., 2022) serves as a basis for free exploration, offering insights into student interests, users' study-related data, the usefulness of features, and their usage frequency. Subsets of this dataset already have served as a basis for publications in the domain of educational science (Schurz et al., 2021; Weber et al., 2021c).

In the following, we give a brief overview of the SIDDATA study assistant software architecture. Further, we introduce SIDDATA's data model and briefly highlight the data extraction process. We then list descriptive statistics characterizing the dataset. We conclude with a discussion of potential research avenues based on the present dataset as well as its limitations.

2 BACKGROUND

The present data were collected from user interaction with version 2.0 of the SIDDATA study assistant software. Students from three German universities used the software, and the dataset contains data from all three universities. Data aggregation occurred over the course of 11 months, from December 2020 to November 2021.

Because the dataset follows the student assistant software's database structure, for a better understanding, we present the software architecture underlying the DSA. We highlight the hierarchical structure of database objects on a conceptual level and outline user-database object interactions facilitated by the software.

2.1 Software Architecture

The SIDDATA study assistant software is implemented as a web service by utilizing the django web framework. Figure 1 illustrates the software architecture and its components. The study assistant software consists of two interlocking parts that communicate via a RESTful-API: the SIDDATA plugin ("frontend") and the SIDDATA backend.

The frontend serves as a graphical user interface integrated into an existing learning management system (LMS) to incorporate DSA functionalities without the need to integrate them into the core of the LMS directly. SIDDATA was developed and tested with the "Stud.IP" (see https://www.studip.de/) LMS.

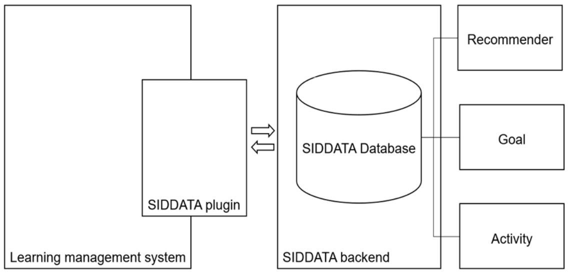

Figure 1: The components of SIDDATA's web-based software architecture.

On the one hand, the frontend provides data such as course or user information to the backend. On the other hand, it serves as an interface between users and SIDDATA backend to request and receive functionalities implemented in the backend logic.

The SIDDATA backend implements and serves SIDDATA functionalities. It receives and processes data from the LMS via the SIDDATA plugin. It also handles user interaction by evaluating user inputs and by creating appropriate answers, as implemented in individual backend functions. A central component of the SIDDATA backend is the SIDDATA database. This PostgreSQL database stores, manages, and retrieves information on individual SIDDATA functions, user data, educational resource data, and the state of interaction between users and the study assistant software.

2.2 Hierarchical Data Structure

SIDDATA's data model follows a strict hierarchy: The uppermost hierarchy level is formed by so-called **Recommender** modules. Attached to these modules are **Goal** type objects. **Activity** type objects are attached to goals and form the building blocks for user-system interactions. Recommenders can have multiple Goals, and Goals can have multiple Activities attached. Figure 2 illustrates the hierarchical structure of the SIDDATA data model.

SIDDATA's functionalities are realized in **Recommender** modules. Each recommender aims at supporting student self-regulated learning through offering a specific service. Recommenders vary in terms of complexity: While some implement, for example, a simple guide for applying and pursuing a

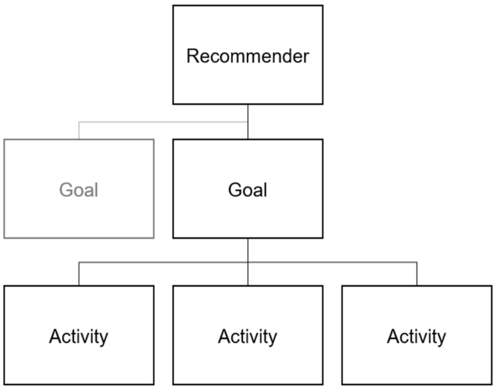

Figure 2: Hierarchical structure of the SIDDATA data model.

semester abroad, others offer more complex functionalities such as a natural language based semantic search engine for educational resources (Schrumpf et al., 2021).

Attached to each recommender are **Goals**. These objects represent abstract, long-term, or persistent sub-functionalities of Recommenders. One example of a Goal is a guide for navigating the administrative landscape for a semester abroad. Figure 3 shows an example of a Goal as displayed for the user.

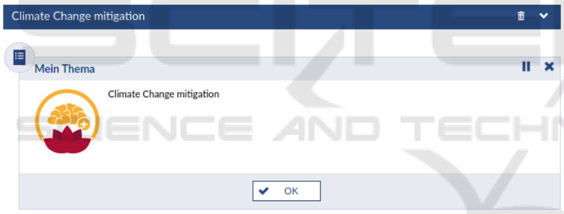

Figure 3: Example of a Goal type object from the "Academic Interests" recommender. The student has entered "climate change mitigation" as one of their interests. The Goal object is displayed as the dark blue strip (Climate Change mitigation) above an attached Activity object (Mein Thema[German: *My Topic*]). Goal type objects can hold multiple Activity type objects.

Goal type objects can be discarded by clicking the trash can symbol on the top right of the frame. This may be useful if users decide not to pursue a goal further and want to remove it from the interface.

Attached to Goals are **Activities**; these represent atomic actions the user can perform within the study assistant software. An example of an Activity is visiting the foreign office of one's home university to sign a learning agreement for a semester abroad. Activity objects are commonly displayed as a dedicated interactive graphic with a bounding box (see Figure 4).

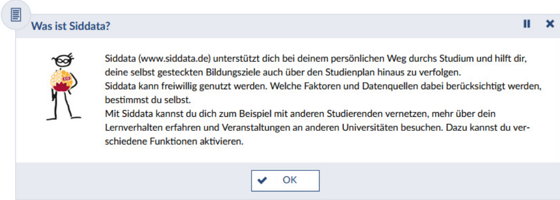

Figure 4: Example of an Activity type object as displayed in a learning management system.

Activities can be discarded and snoozed by users. A snoozed Activity is removed from the user interface and listed under the "paused" tab. Such Activities can be resumed at a later point. Discarded Activities are logged under the "discarded" tab. Finally, Activities that were completed successfully are listed under the "completed" tab.

2.3 Recommender Modules

Recommenders provide specific functionalities of the study assistant software. They serve as the highest attachment point for Goal objects. From a technical perspective, Recommenders have their own interaction logic and handle the processing of attached Goals and Activity objects. From the user perspective, recommenders are sub-functions that can be enabled and disabled and provide specific services, such as finding personally relevant learning resources (Fachliche Interessen, eng: Academic Interests) or connecting to other learners (Get-together, eng: Academic Contacts).

Figure 5: Example of the SIDDATA navigation menu displaying activated Recommenders above and links to paused, discarded, and completed Recommenders below.

Recommenders offer students the service to engage with and reflect upon individual educational goals associated with the aspect of studying associated with a recommender's coverage: The "Open Educational Resource" recommender, for example, introduces learners to and highlights key properties of, OERs and how they can assist with learning subjects relevant to the user. Recommenders thus are to be interpreted as independent interactive features of the study assistant system. Once enabled, recommender modules appear in the navigation menu and are accessible via the graphical user interface (see Figure 5).

Our dataset holds information about which users have activated individual Recommender modules, as well as which Goals and Activities are attached to Recommenders. Table 1 shows the Recommenders for which data is available in our dataset.

Table 1: Recommenders and their function.

Recommender title (German title)	**Description**
Academic Contacts (Get-together)	A tool for students to connect, based on their personal matching preferences
Academic Interests (Fachliche Interessen)	Semantic search engine for Educational Resources based on natural language processing (see Schrumpf et al., 2021)
Data Ethics (Souverän mit meinen Daten)	Providing information about data ethical aspects of the digital age
Evaluation (Evaluation)	Survey for users to give detailed feedback
Funding (Studienfinanzierung)	Providing information on how to find funding for one's studies
Learning Organization (Lernorganisation)	Providing information about methods and practices for self-regulated learning
Open Educational Resources (Freie Bildungsmaterialien)	Providing information about Open Educational Resources and a list of preselected repositories
Personality Module (Persönlichkeitsmodul)	Personality test for measuring task-switching ability and short/long-term memory
Scientific Career (Wissenschaftliche Karriere)	Providing general information about a career in the scientific field
Study Abroad (Auslandssemester)	Providing general information about study abroad

SIDDATA's design philosophy is based on a constructivist learning ideal where intrinsically motivated learners take responsibility for their personal and intellectual development. At the base of this philosophy stand personal, educational goals, which is why the substructure of recommenders is called "goals" (see goal-related references for more information). For some of the goal entities in the data set it is the case that they correspond to a specific personal goal (such as a specific personal interest in a topic or a semester abroad), while some goal entities in the data set were created by the developers to structure the program flow and the appearance to the user.

For instance, if students enter a text describing a topic of personal interest in the Academic Interests recommender, a goal appears with the interest as title, and course recommendations appear as activities within this goal. The goal here stands for an interest a student wants to follow while studying. In contrast, a goal within the Semester Abroad recommender is, for example, to find a partner university with a fitting study program. Here, a goal stands for a milestone in the process of studying a semester abroad.

Activities are the smallest units in the data structure and can be understood as atomic user interactions with the system. Examples for Activities are recommendations generated and displayed to the user or questions which ask the users for input. The term *Activity* was chosen because the system is intended to elicit actions of students towards their individual goals. All activity type objects have a title and a description displayed directly in the UI of the SIDDATA frontend. Additionally, Activities can be paused or discarded. If Activities are successfully interacted with, they are moved to the finished tab and automatically set to inactive. Activities are sorted into different types, each one implementing a unique mode of interaction:

To-do type Activities act as an item on a checklist, displaying information or reminding the user of a task to be fulfilled to progress in reaching a certain goal.
Question type Activities require the user to make a choice from a pre-selected list of answers or query the user for an answer in natural language.
Resource type Activities present online resources in the form of a weblink. This data type is intended to represent a broad range of educational resources, such as Open Educational Resources (OER), Massive Open Online Courses (MOOCs), or goal-relevant online resources, such as for instance a web-page of an institution offering scholarships that the digital assistant may identify with being of interest for the personal goal to study abroad.

Course type Activities are a recommendation for enrollment into a course from the LMS of the local and partner universities.

Additionally, users can rate any Activity on a scale of 0 to 4. This feedback function provides hints for the quality of the recommended items.

2.4 Data Extraction & Processing

In the SIDDATA database, long and cryptic IDs are used to identify activities, students, recommenders, courses, degrees, and goals for performance and security reasons. To increase usability for researchers working with the dataset and to further anonymize the data to make it even less person-relatable, these IDs were incrementally replaced by plain integer values.

In the original database, some personal or personal-relatable data is present, which had to be removed from the dataset. This includes names, telephone numbers, and email addresses exchanged between students in the Academic Contacts recommender. In some cases, other personal information such as health-related information was present. Information of these kinds was replaced by placeholders, such as for instance <emailadress>. This procedure substitutes personal data in a way that allows researchers to understand and analyze the data while effectively protecting person-relatable information.

3 DATASET PROPERTIES

The core objective of data processing from the original data, stored in a PostgreSQL relational database to one single CSV file, was optimal usability for reuse by the scientific community. While the database model was designed to avoid redundancies, the preprocessed data set repeats information and at the same time allows filtering, aggregation, and custom analyses with all kinds of data processing tools capable of reading CSV files, such as LibreOffice Calc (or a commercial alternative, such as Excel) or data analysis tools, such as R or PSPP (or a commercial alternative, such as SPSS). The size of the resulting comma-separated-value file (.csv) is 20.1 MB. The file format allows the comma or semicolon as a string separator. In this case, the semicolon was chosen because included texts do contain commata, which leads to formatting issues when working with the data. For scientists working with the data, it is important to choose a semicolon and not the comma as a string separator when importing or reading the csv file.

3.1 Dataset Structure and Statistics

To make the dataset more readable, this section introduces the data structure.

Inheriting from the hierarchical data structure of the SIDDATA database, each activity belongs to exactly one Goal, which belongs to exactly one Recommender. Additionally, we have associated each activity with a corresponding student interacting with this activity. Hence, we aggregate all activity, goal, and student information in one row of the dataset. This enhances readability with the tradeoff of increased repetition.

Students possess additional attributes such as field of study, semester of enrollment, and home university. Students can also study in multiple study programs and, therefore, different semesters of enrollment for each of their study programs. We have aggregated student information within squared brackets for dataset columns with potentially multiple entries meaning that each indexed item within the squared brackets corresponds to the data at the same index in another column with multiple entries. For example, if a student studies in two fields of study, [mathematics, philosophy], the corresponding semester of enrollment is stored as [3,5], where 3 corresponds to “mathematics” and 5 corresponds to “philosophy”, in the semester of enrollment column.

Filtering functions allow to “zoom” into the data to inspect, for instance, all activities of a user, all users of a university, all students enrolled into a specific study program, etc. Table 2 gives a detailed explanation of variables and a range of possible values for each column of the dataset.

Table 2: Summary of columns present in the dataset with a more in-depth explanation of column meaning.

Column	Explanation
activity_id	unique identifier for activities, sequential numbering
activity_title	title of the activity
activity_description	description text
activity_status	status of interactions (possible values: “new”, “active”, ”snoozed”, ”done”, “discarded”, ”active”)
activity_type	type of activity (possible values: “todo”, ”question”, ”resource”, ”course”)
feedback_scale_size	Number of choices the user has to rate an Activity. Ranges from 2 to 5, 0 stands for no feedback option.

Table 2: Summary of columns present in the dataset with a more in-depth explanation of column meaning (cont.).

Column	Explanation
Given_feedback	Feedback value for an Activity if it was rated by the user
timestamp	time of creation
question_id	unique identifier for questions, sequential numbering (optional, only for questions)
question	question text (optional, only for questions)
answers	set of possible answers for multiple choice questions (optional)
given_answers	answers given by the student (optional)
course_id	unique identifier for courses, sequential numbering (optional, only for courses)
resource_id	unique identifier for resources, sequential numbering (optional, only for recommended resources)
resource_title	title of resource (optional, only for resources)
resource_url	url of a resource (optional, only for resources)
course_title	title of a recommended course (optional, only for courses)
goal_id	unique identifier for goal, sequential numberings
goal_title	short title of a goal
goal_description	detailed description of a goal (optional)
goal_order	order in which the goal is displayed in the recommender
recommender	name of the recommender
student_id	unique identifier for students, sequential numbering
student_university	home university of the student
student_institute	institute of the student (optional)
subject*	study subject of the student (optional)
subject_id	unique identifier for subjects, sequential numbering
degree_id	unique identifier for degrees, sequential numbering
degree*	target degree for a subject of the student
semester*	semester for each combination or subject and degree for a specific student

*Subject, degree and semester together state an enrollment of a student into a study program. One student can be enrolled in more than one study program. In this dataset, subject, degree, and semester contain ordered lists. The order clarifies which subject, degree, and semester together constitute one enrollment. The number of known enrollments per student range from 0 to 5.

Table 3 shows the numbers for each entity type represented in the data.

Table 3: The dataset holds information about 735 students from three universities and the following other entity types and quantities.

Entity Type	n
Universities	3
Study Subjects	165
Degrees	19
Students	735
Recommender	9
Goals	6081
Activities	41520
Courses	1365
Questions	899
Semesters	15

4 DISCUSSION

With the SIDDATA prototype 2 dataset, we present a source of data for investigations in the domain of study assistant engineering, digital education, educational psychology, recommendation engine evaluation and the strategic considerations for learning in the digital age.

The present dataset already served as a basis to perform an effectiveness analysis of the SIDDATA assistant software (see Schurz et al., 2021). However, with this first publication, the depths of the SIDDATA prototype 2 dataset are far from being used to exhaustion.

4.1 Potential Research Avenues

We have identified the following paths as of yet unexplored research opportunities for future scientific endeavors:

So far, we have not performed analysis into student-system interaction based on the field of study. Aggregating statistical information about such usage could produce novel insights into the student behavior in terms of technology openness, the effectiveness of software features, or how well the software design fits the needs of students from different domains of expertise. General statistical information about the number of enrolled students per field of study is available at each partner university's webpage (see Appendix) and, in conjunction with the present dataset, reveals new information.

Further, we have not analyzed the feedback from students on single Activity objects, aside from a general trend which indicates which Activities from which Recommender was generally perceived to be useful or less useful, interactions between the semester of enrolment, the field of study, and likelihood to rate certain Activities with certain values may be discoverable.

Additionally, an analysis of when which recommenders were used may also give insights into the needs and interests of students over the course of two semesters.

Regarding individual recommenders, there are multiple open questions that, as of yet, have not been examined in detail:

The Academic Interests recommender (Fachliche Interessen) logs interests that were entered by students. These interests have not been analyzed for their content or correlation with the field of study, semester of enrollment, or other student data. Here, an analysis of feedback for given recommendations could yield insights into how fitting recommendations were to requests.

Another recommender whose investigation could hold valuable insight is the Academic Contacts (Get-together) recommender: Here, students are able to get matched to other students based on their common interests. Analyzing the common interests students have entered in order to find matches could give a perspective on what kind of extracurricular workshops or events could be offered by student bodies or by other university institutions to effectively facilitate student interest networks.

The present dataset may also support studies that evaluate digital study assistant systems in general or that seek to assess the state of digital higher education in general.

4.2 Limitations

The dataset presented here offers a variety of data for exploring the effectiveness of a digital study assistant system. However, some limitations have to be noted, which may impact future findings from the present data: Before using the study assistant software, users were asked whether they were willing to share their data for research purposes. The data in the dataset presented here, therefore, is only composed of interactions logged from users who affirmed the usage of their data. Therefore, a confounding effect between willingness to share data and the willingness to explore and interact with the student assistant software caused by general user technology acceptance cannot be ruled out.

Another limitation of the dataset is the quality of data: Even though general data quality can be considered high, some data such as field of study or the title of recommended courses are drawn from the learning management system of participating universities. This means that the data quality of these items may vary, as learning management system settings for one university (Bremen) allowed students to edit their field of study, while the other two universities (Hannover, Osnabrück) only gave officials the writing permissions for these data fields. For courses, some course titles may only give information in context and for a specific field of study. This holds, for example, for acronyms that carry little meaning outside a field. Further, semesters of enrollment and users enrolled in a specific field of study are not evenly distributed. This means that investigations relying on semesters of enrollment and field of study may not yield representative results and may need to be pre-selected for a sufficient number of samples before they can be used in a study.

4.3 Outlook

Since November 2021, the third and final version of the software prototype has been released and is running productively at the three partner universities. This new version includes more and refined features such as the inclusion recommendations for MOOCs as motivated by Vogelsang et al. (2019), an interactive tool for hierarchical goal setting as per Weber et al. (2021b). The feature set has been extended and improved, but the general database structure has been maintained and refined. In the first half of 2022, a new dataset with a compatible structure will be published, which allows the reuse of data analyses scripts and methods developed for the dataset published along with this paper on new data.

Because the features the SIDDATA study assistant software is offering are realized in the SIDDATA backend, data from new LMS systems are easy to integrate into the existing software. This allows to expand SIDDATA's reach beyond the three partner universities and, potentially, to extend the digital portfolio of many universities with low cost and low maintenance requirements.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany for Siddata (project number 16DHB2124).

We thank all students who donated their data.

REFERENCES

Castro, R. (2019), "Blended learning in higher education: Trends and capabilities", *Education and Information Technologies*, Education and Information Technologies, Vol. 24 No. 4, pp. 2523–2546.

Reinken, C., & Greiff, P. (2021). AIS Electronic Library (AISeL) A Multi-Perspective Research for the Investigation of Software Requirements for a Digital Study Assistant

Schrumpf, J., Weber, F., & Thelen, T. (2021). A Neural Natural Language Processing System for Educational Resource Knowledge Domain Classification. In: Kienle, A., Harrer, A., Haake, J. M. & Lingnau, A. (Hrsg.), DELFI 2021. Bonn: Gesellschaft für Informatik e.V. (S. 283-288).

Schurz, K., Schrumpf, J., Weber F., Lübcke M., Seyfeli, F., and Wannemacher, K., (2021). Towards a User-Focused Development of A Digital Study Assistant Through a Mixed Methods Design, in Sampson, D.G., Ifenthaler, D., Pedro, I., Mascia, M.L. (Eds.), 18th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2021). IADIS Press

Thelen, T., König, C., Wannemacher, K., Wollersheim, H.-W., Köhler, T., Igel, C., … Riedel, J. (2019). Digitale Werkzeuge für Studienindividualisierung und personalisierte Kompetenzentwicklung. In J. Hafer, M. Martina, & M. Schumann (Eds.), *Medien in der Wissenschaft, Band 75, Teilhabe in der digitalen Bildungswelt* (pp. 258–262). Münster, New York: Waxmann Verlag.

Vogelsang, K., Greiff, P., Tenspolde, C. and Hoppe, U. (2019), "Agile by technique – The role of technology enhanced learning in higher education ", *Beiträge Zur Hochschulforschung, 41. Jahrgang*, No. 3, pp. 28–46.

Weber, F., (2019). Goal Trees as Structuring Element in a Digital Data-Driven Study Assistant, in Sampson, D.G., Ifenthaler, D., Pedro, I., Mascia, M.L. (Eds.), 16th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2019). IADIS Press, Cagliari, Italy, pp. 413–416

Weber, F., Kernos, J., Grenz, M., and Lee, J., (2021a). Towards a Web-Based Hierarchical Goal Setting Intervention for Higher Education, in Sampson, D.G., Ifenthaler, D., Pedro, I., Mascia, M.L. (Eds.), 18th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2021). IADIS Press.

Weber, F., & Le Foll, E. (2020). A Tagset for University Students' Educational Goals. In *17th International Conference on Cognition and Exploratory Learning in the Digital Age, CELDA 2020*. https://doi.org/10.33965/celda2020_202014l004

Weber, F., Osada, S., & Thelen, T. (2019). Towards a Comprehensive Taxonomy of Study Goals of University Students. *Poster Session Presented at the EuroCogSci 2019*, 2125.

Weber, F., Schrumpf, J., and Thelen, T., (2021b). Development of a Digital Goal Setting Companion for Higher Education, Andrea Kienle et al. (Hrsg.): Die 19. Fachtagung Bildungstechnologien (DELFI), Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021.

Weber, F., Schurz, K., Schrumpf, J., Seyfeli, F., Wannemacher, K., & Thelen, T. (2021c). Digitale Studienassistenzsysteme Von der Idee zur Umsetzung im Projekt SIDDATA. In H.-W. Wollersheim, M. Karapanos, & N. Pengel (Eds.), Medien in der Wissenschaft, Band 78, Bildung in der digitalen Transformation (pp. 239–244). Münster, New York: Waxmann Verlag.

Weber, F., and Thelen, T., (2021), Characterizing Personal Educational Goals -Inter-rater Agreement on A Tagset Reveals Domain-Specific Limitations of the External Perspective, Ifenthaler, D., Sampson, D. G., Isaias, P. (Eds.). (Forthcoming 2021). Balancing the Tension between Digital Technologies and Learning Sciences. New York, NY: Springer.

Weber, F., Schrumpf, J., Dettmer, N., Schurz, K., Thelen, .T, (2022), "Siddata Study Assistant Dataset from Prototpye 2.0", https://doi.org/10.26249/FK2/YYIDKE, osnaData, V1

APPENDIX

Statistics on student enrollment can be found for each involved university by following the provided links:

Leibniz University of Hannover: https://www.uni-hannover.de/de/universitaet/profil/zahlen/studierendenstatistik/

University of Bremen: https://www.finanzcontrolling.uni-bremen.de/daten/index.htm

University of Osnabrück: https://www.uni-osnabrueck.de/universitaet/zahlendatenfakten/studierendenstatistiken/