STAG: Smart Tools and Applications in Graphics (2025)
M. Comino Trinidad, C. Mancinelli, D. Cabiddu, and D. Giorgi (Editors)

GUIDÆTA – A Versatile Interactions Dataset with extensive Context Information and Metadata

S. Lengauer[1], S.A. von Götz[2], M.T. Hoesch[2], F. Steinwidder[1], M. Tytarenko[1], M.A. Bedek[2] and T. Schreck[1]

[1]Graz University of Technology, Institute of Visual Computing, Austria
[2]University of Graz, Department of Psychology, Austria

Abstract

Interaction data is widely used in multiple domains such as cognitive science, visualization, human computer interaction, and cybersecurity, among others. Applications range from cognitive analyses over user/behavior modeling, adaptation, recommendations, to (user/bot) identification/verification. That is, research on these applications – in particular those relying on learned models – require copious amounts of structured data for both training and evaluation. Different application domains thereby impose different requirements. I.e., for some purposes it is vital that the data is based on a guided interaction process, meaning that monitored subjects pursued a given task, while other purposes require additional context information, such as widget interactions or metadata. Unfortunately, the amount of publicly available datasets is small and their respective applicability for specific purposes limited. We present GUIDEd Interaction DATA (GUIDÆTA) – a new dataset, collected from a large-scale guided user study with more than 250 users, each working on three pre-defined information retrieval tasks using a custom-built consumer information system. Besides being larger than most comparable datasets – with 716 completed tasks, 2.39 million mouse and keyboard events (2.35 million and 40 thousand, respectively) and a total observation period of almost 50 hours – its interactions exhibit encompassing context information in the form of widget information, triggered (system) events and associated displayed content. Combined with extensive metadata such as sociodemographic user data and answers to explicit feedback questionnaires (regarding perceived usability, experienced cognitive load, pre-knowledge on the information system's topic), GUIDÆTA constitutes a versatile dataset, applicable for various research domains and purposes. Alongside the data itself, we publish the software tools we use for handling and analyzing the dataset.

CCS Concepts

• ***Human-centered computing*** → ***Interactive systems and tools;*** • ***Applied computing*** → *Health care information systems;*

1. Introduction

Interaction data in the form of mouse and keyboard events or widget interaction (the target of such an event) contain vital information about users interacting with Graphical User Interfaces (GUIs). Its core advantage is that it can be collected unobtrusively and inexpensively without interfering with users' natural behavior. As such, it constitutes the basis for different objectives of different research domains. At the same time, it plays a vital role in many web-based systems where collected interaction data is used for fingerprinting, profiling, user classification/-authentication, inferring interest and other purposes.

In the field of cognitive science, behavioral indicators such as entropy, velocity, acceleration, or curvature are computed to infer users' mental activity, visual attention, cognition, emotion (e.g., stress), or interest [MKSY23, LA20]. Closely related are the objectives of the Human Computer Interaction (HCI) domain, where similar metrics based on interaction speed and timing are leveraged to improve GUIs [MSK*90, KDMH24]. Within the subfield of *interaction ergonomics*, researchers attempt to answer questions such as "how quickly can a user position the mouse and click on a button or a drop-down menu?", relating to Hick's [Hic52] and Fitts's [Fit54] laws of experimental psychology. Behavioral patterns are also used to determine self-efficacy, risk-perception, willingness to learn, or perceived usefulness/ease of use [KN18]. On a similar note, in visualization (user-adaptive visualization in particular), interaction data constitutes the core input for user modeling, which – in turn – is used to adapt GUIs according to respective users' needs [YCON25]. Such adaptions can manifest in a change of complexity, representation, or displayed content, among others [GW09, SLM*25].

In the domain of cybersecurity, interactions are used to uniquely identify users based on their behavior. This subfield known as *behavioral biometrics* is studied for the purpose of user authentication [KFH*21, ADF19, SCG12] and fending against bots [SV23, AMFVR22, WZC19]. In essence, respective approaches build upon behavioral characteristics, such as mouse dynamics (the mouse be-

DOI: 10.2312/stag.20251335

havior on GUIs), keystroke dynamics, swipe dynamics, widget interaction among others.

The common thread of these research domains is the need for large (real-world) datasets, containing all the necessary properties. With our GUIDÆTA dataset we aim to address this need. Collected through a purely unsupervised online study, it exhibits a broad spectrum of user demographics in terms of age and education level. That is, its versatility enables its use for different research objectives and applications. In summary, the core strength of the GUIDÆTA dataset is fourfold:

1. It contains exclusively *guided* interaction data, meaning that study participants completed the same set of Information Retrieval (IR) tasks for which we also report whether they have been completed successfully and correctly.
2. As opposed to most related datasets, it also includes the context of interaction events. Besides the widget (the target of an interaction), we define a set of 18 uniquely identifiable actions, which are supported by the system used for collecting the dataset. A particular event can thus have the attached information that it, e.g., triggered the pop-up of a thumbnail with specific content.
3. We provide various additional metadata. On a participant level, this includes sociodemographic information and answers to different explicit feedback questionnaires, such as the system's perceived usability and pre-knowledge on the system's information domain. On a per-completed-task level, we have participants' feedback on the experienced Cognitive Load (CL) and the required working time.
4. All of the collected data is made publicly available through the Creative Commons Attribution (CC BY) license, respecting the Findable, Accessible, Interoperable, and Reusable (FAIR) Guiding Principles for scientific data management and stewardship. In line with this mantra, we also provide all the accompanying tools for loading, filtering and analyzing the data.

In the following, we present a delineation to other related datasets (Sec. 2), before the collection process and the dataset's structure is described in depth (Sec. 3). Sec. 4 explains our purposely-developed data handling and analysis tools, while Sec. 5 gives an outlook on potential limitations, applications and follow-ups.

2. Related Work

Interaction data has been studied since the early days of GUIs, pursuing different objectives. While the analysis of the keystroke dynamics constitutes a research focus since the 1970s, mouse dynamics experienced increased attention since the beginning of the 21st century due to the immense growth of the Internet [KDMH24]. Regardless of the research domain (cognitive sciences, HCI, adaptive visualization, or cybersecurity), this interest cultivates a need for (large-scale) datasets, which allow researchers to objectively evaluate hypotheses and validate models. Depending on the intended use, the collection process for compiling said datasets varies. I.e., interactions can be collected in a purely unsupervised manner over the Internet, allowing for a huge outreach while keeping costs low, or they can be collected in laboratory setups, allowing to tap into additional information channels such as eye tracking or verbal interviews. Orthogonal to that, participants can be asked to pursue specific tasks (guided) or just engage in their day-to-day activities (unguided) [ADF19]. While the prior is crucial to study between-subjects' activity, cognition, or stress [MKSY23], the latter is completely sufficient for user authentication/identification [KFH*21, ADF19, SCG12]. Kahn et al. [KDMH24] go even one step further and define 4 tiers of guidance: (1) fixed static sequence of actions, (2) app restricted continuous data collection, (3) app agnostic semi-controlled data collection, and (4) completely free data collection.

As the focus of this paper is not on specific application domains, but datasets, we discuss the most relevant publicly available datasets in the following. Other datasets were compiled to evaluate approaches for user modeling/behavior prediction [GA10, HWB12, CLZM17, MLZM14] or for user authentication [HAG15, ZPW16, FEM*12] but are, unfortunately, not publicly available.

Balabit Published in 2016 by Fülöp et al. [FKKWP16], the Balabit dataset contains mouse pointer positioning and timing information of 10 users who connected over a remove server. The authors reason that participants can be uniquely identified, purely based on these mouse dynamics, preventing unauthorized usage of their accounts. Consequently, the collected data is unguided in nature, as users were asked to perform their regular daily duties.

Chao Shen Similarly, Shen et al. [SCG12] asked their participants to pursue their daily work while they tracked mouse movements in the background. That is, they were able to collect data pertaining to 28 individuals over a period of over two months. Each of those completed at least 30 separate sessions of about thirty minutes.

Bogazici In 2021 Kılıç et al. [KYA21] published a dataset with unguided mouse interactions of 24 users with several days of active usage. Besides positions and timestamps, they also collected window name (i.e., widget interaction) and mouse action details.

DFL Antal and Denes-Fazakas [ADF19] compiled a dataset of 21 different users, which they use for user verification. They reason that about 60 minutes of interaction data is necessary to model a user, while at least 10 mouse actions are required for identity prediction. Data collection was enabled through a background service, which users were asked to install before pursuing their daily work routine.

SapiMouse The SapiMouse Dataset, which was collected at Sapientia University in 2020 by Antal et al. [AFB21, ABF21] pursues a different objective. As opposed to long observations over the course of days or months, they have a short fixed-length observation period of 4 minutes per user. However, within this time frame participants were asked to perform very concrete actions in the form of a mini game. The interactions, collected from 120 participants, are thus highly guided, allowing for in-depth between-subject comparisons. Based on that, they present a deep learning approach for user authentication, which learns descriptive features directly from the raw data [AFB21]. As a follow-up, they also present an autoencoder-based approach for generating human-like trajectories [ABF21].

The Attentive Cursor (TAC) The largest dataset to date in terms of participants, was published in 2020 by Leiva and Arapakis [LA20]. They collected mouse dynamics of 2,909 subjects performing a transactional search task, together with attention labels and demographic attributes. Alongside timestamp, position, and event type, they also report widget information in the form of the DOM element related to an event.

ReMouse Sadeghpour and Vlajic [SV23] compiled a dataset for the very specific purpose of fending against session-replay bots. Given this particular objective, the unique characteristic of their dataset is that it contains repeated sessions generated by the same human user. Over the course of two days, they collected interaction data from 100 subjects participating in a 'Catch Me If You Can!' online game, who were recruited over Amazon's MTurk platform.

AdSERP Most recently, Latifzadeh et al. [LGL25] presented the AdSERP dataset, which combines mouse movement data with eye tracking. They aim to study user attention and purchasing behavior on search engine result pages. To this end, they tracked the interactions of 47 participants in a supersized setting with interaction sessions of up to one minute.

Table 1 shows a detailed breakdown of the specifics of the above-mentioned datasets, including the total number of users, the total number of recorded interactions the total observation period, whether they incorporate widget information, and the level of guidance during the collection process. We can observe, that solely **TAC** boasts more participants, but on the downside this dataset comprises only very short interactions – having thus a substantially shorter observation period and substantially fewer overall interaction events. Regarding the sheer observation duration, there are three datasets (**Balabit**, **Bogazici**, **DFL**) standing out with durations of multiple weeks or even months. However, these interactions stem from collection efforts which were conducted unguided and unobtrusively, asking participants to conduct their everyday work while a tracking service collected interactions in the background. With respect to these alternatives, our GUIDÆTA dataset fills a nïche as it exhibits a high count of participants, considering that it was collected from a user study with a high-level of guidance through purposely posed IR tasks. A mean observation period of 11:46 minutes ($SD = $9:57 minutes) per user, allowed us to incentivize a comparably large number of participants to partake in our study while still having a reasonable amount of data per person.

3. The GUIDÆTA Dataset

In the following, we provide details on the data collection process (Sec. 3.1), used to compile the GUIDÆTA dataset. Sec. 3.2 provides a statistical overview of the data in terms of different users and tasks, before we conclude the section with a description of the data structure we devised to store and distribute the data (Sec. 3.3).

3.1. Data Collection

The interaction data was collected in the course of a cognitive science study from May 12 till June 23, 2025.

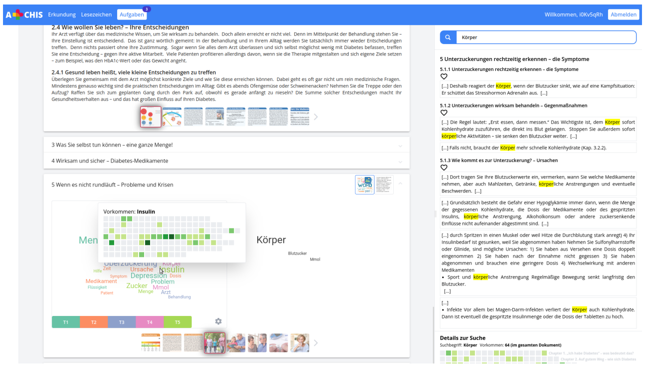
Study Design. The basis for the collection setup is a Consumer Health Information System (CHIS), which is used by study participants to process the posed information retrieval tasks. This Advanced interactive, Adaptive, personalized and visual CHIS (A^+CHIS) [SLM*25], developed within the eponymous funding project, comprises various components for presenting the textual and pictorial content of information sources for medical knowledge (Fig. 1). Text can be consumed in original format but also through various *distant reading* [Mor05] solutions, such as word clouds, topic modelings, heat maps, search functionality, etc., which allow to process the given information in non-linear fashion. Subsystems are in place to display pictorial content, including interactive infographics.

Table 1: *A breakdown of related datasets in terms of scale.*

Dataset	# Users	# Interactions	Observation duration†	Widget Information	Guided‡
Balabit	10	4.61 M	178.21 H		○
Chao Shen	28	79.69 M	499.00 H		○
Bogazici	24	72.83 M	539.62 H	✓	○
DFL	21	129.56 M	718.06 H		○
SapiMouse	120	1.18 M	8.22 H	N.A.	●
TAC	2,909	0.11 M	12.74 H	✓	◍
ReMouse	100	0.38 M	6.74 H	N.A.	●
AdSERP	47	1.17 M	16.39 H		◍
GUIDÆTA	253	2.39 M	49.61 H	✓	◉

† The accumulated duration over all sessions. As there is no uniform definition of 'session', we assume that a gap of more than 10 seconds in the recording indicates a session separator. Note, that this can result in durations that substantially differ from the authors' reports.

‡ Level of guidance as defined by Kahn et al. [KDMH24]: ● *Fixed static sequence of actions*, ◉ *App restricted continuous*, ◍ *App agnostic semi-controlled*, and ○ *Completely free*.



2.4 Wie wollen Sie leben? – Ihre Entscheidungen

Ihr Arzt verfügt über das medizinische Wissen, um Sie wirksam zu behandeln. Doch allein erreicht er nicht viel. Denn im Mittelpunkt der Behandlung stehen Sie – Ihre Einstellung ist entscheidend. Das ist ganz wörtlich gemeint: In der Behandlung und in Ihrem Alltag werden Sie tatsächlich immer wieder Entscheidungen treffen. Denn nichts passiert ohne Ihre Zustimmung. Sogar wenn Sie alles dem Arzt überlassen und sich selbst möglichst wenig mit Diabetes befassen, treffen Sie eine Entscheidung – gegen Ihre aktive Mitarbeit. Viele Patienten profitieren allerdings davon, wenn sie die Therapie mitgestalten und sich eigene Ziele setzen – zum Beispiel, was den HbA1c-Wert oder das Gewicht angeht.

2.4.1 Gesund leben heißt, viele kleine Entscheidungen zu treffen

Überlegen Sie gemeinsam mit dem Arzt möglichst konkrete Ziele und wie Sie diese erreichen können. Dabei geht es oft gar nicht um rein medizinische Fragen. Mindestens genauso wichtig sind die praktischen Entscheidungen im Alltag: Gibt es abends Ofengemüse oder Schweinenacken? Nehmen Sie die Treppe oder den Aufzug? Raffen Sie sich zum geplanten Gang durch den Park auf, obwohl es gerade anfängt zu nieseln? Die Summe solcher Entscheidungen macht Ihr Gesundheitsverhalten aus – und das hat großen Einfluss auf Ihren Diabetes.

3 Was Sie selbst tun können – eine ganze Menge!

4 Wirksam und sicher – Diabetes-Medikamente

5 Wenn es nicht rundläuft – Probleme und Krisen

5 Unterzuckerungen rechtzeitig erkennen – die Symptome

5.1.1 Unterzuckerungen rechtzeitig erkennen – die Symptome

[...] Deshalb reagiert der Körper, wenn der Blutzucker sinkt, wie auf eine Kampfsituation: Er schüttet das Stresshormon Adrenalin aus. [...]

5.1.2 Unterzuckerungen wirksam behandeln – Gegenmaßnahmen

[...] Die Regel lautet: „Erst essen, dann messen." Das Wichtigste ist, dem Körper sofort Kohlenhydrate zuzuführen, die direkt ins Blut gelangen. Stoppen Sie außerdem sofort körperliche Aktivitäten – sie senken den Blutzucker weiter. [...]

[...] Falls nicht, braucht der Körper mehr schnelle Kohlenhydrate (Kap. 3.2.2).

5.1.3 Wie kommt es zur Unterzuckerung? – Ursachen

[...] Dort tragen Sie Ihre Blutzuckerwerte ein, vermerken, wann Sie welche Medikamente nehmen, aber auch Mahlzeiten, Getränke, körperliche Anstrengungen und eventuelle Beschwerden. [...]

[...] Grundsätzlich besteht die Gefahr einer Hypoglykämie immer dann, wenn die Menge der gegessenen Kohlenhydrate, die Dosis der Medikamente oder des gespritzten Insulins, körperliche Anstrengung, Alkoholkonsum oder andere zuckersenkende Einflüsse nicht aufeinander abgestimmt sind. [...]

[...] durch Spritzen in einen Muskel oder weil Hitze die Durchblutung stark anregt) 4) Ihr Insulinbedarf ist gesunken, weil Sie abgenommen haben Nehmen Sie Sulfonylharnstoffe oder Glinide, sind mögliche Ursachen: 1) Sie haben aus Versehen eine Dosis doppelt eingenommen 2) Sie haben nach der Einnahme nicht gegessen 3) Sie haben abgenommen und brauchen eine geringere Dosis 4) Wechselwirkung mit anderen Medikamenten
- Sport und körperliche Anstrengung Regelmäßige Bewegung senkt langfristig den Blutzucker.

[...]

[...]
- Infekte Vor allem bei Magen-Darm-Infekten verliert der Körper auch Kohlenhydrate. Dann ist eventuell die gespritzte Insulinmenge oder die Dosis der Tabletten zu hoch.

Details zur Suche

Suchbegriff: Körper Vorkommen: 64 (im gesamten Dokument)

Figure 1: *Screenshot of the A^+CHIS, used in the data collection.*

To increase the outreach, the study was conducted in purely unsupervised fashion, by providing participants with a link to the website on which they conducted the tasks using their own hardware devices. To foster interest in participation, advertisements were placed on the project's website, posted on social media, emailed to students and those expressing interest in the project, and distributed in physical form as posters. Actual ambition to solve the posed IR tasks was incentivized with an unconditional expense allowance of €20 and the chance to get an additional €100 (which

Proceedings published by Eurographics - The European Association for Computer Graphics.

was distributed ten times by lot to all participants who were able to correctly answer all tasks). To ensure a smooth execution they had to go through a number of well-defined steps:

1. Initially, participants are greeted with a welcome page outlining the objectives of the study, clarifying the scope and what data will be collected. Only after agreeing to these terms and conditions they are forwarded to the next step.
2. Participants are presented with a form for filling out anonymous sociodemographic information and asked to complete a standardized knowledge-test on diabetes – the Revised Brief Diabetes Knowledge Test (DKT2) [FFA*16] – (the thematic domain of the A$^+$CHIS).
3. Next, they are presented with a click-through help wizard, educating them on the different components of the system through textual descriptions combined with visual examples.
4. Next, a task modal pops up, displaying the first of 3 tasks which are provided in randomized order:

   ***task A*:** *"What is the time frame after administration when the peak effect of regular insulin is reached?"*,
   ***task B*:** *"From what blood sugar level are acute measures necessary to prevent a risky hyperglycemia?"*, and
   ***task C*:** *"How often should the HbA1c value be checked by a doctor at a minimum?"*

   A participant can re-open this modal and the help wizard at any moment through distinct buttons. For completing a task, the modal has to be reopened and the answer filled in. After each task, participants are prompted with a form evaluating the experienced CL, with a CL questionnaire [KBR*23].
5. After completing all the tasks, a form for a System Usability Scale (SUS) [Bro96] and further details on compensation are presented to participants.

After the study part, participants were still able to use the system for further exploration. We deliberately decided to pose open questions, which had to be answered in a free-form text, to prevent users from guessing.

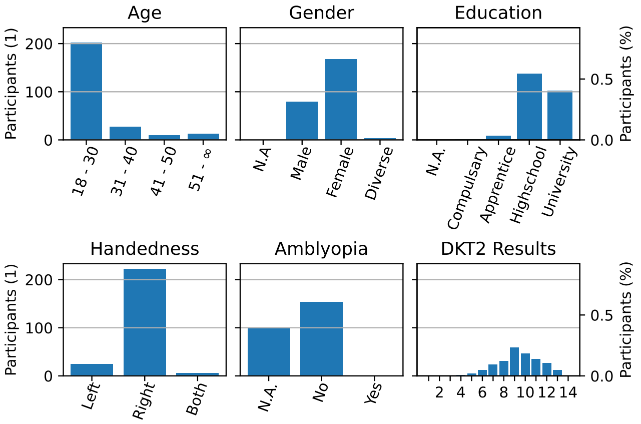


Figure 2: *The distribution of age, gender, education level, handedness, amblyopia, DKT2 results among the study participants.*

Technological Setup. For the study the A$^+$CHIS was striped down to contain only its relevant core parts. Yet, additional subsystems for the onboarding, task handout, questionnaires and help pages were developed to allow for a fully unsupervised execution. The most substantial adaptations pertain to the changes related to the comprehensive tracking of interactions. To this end, we employ various JavaScript event listeners[†], monitoring mouse events (click, scroll, move), keyboard events and window events (focus, blur, resize). To capture additional context, we enrich the A$^+$CHIS's Document Object Model (DOM) with custom tags labeling tools and content. This allows us, e.g., to reconstruct that a mouse click happened using a specific tool while exploring a specific information unit (e.g., a sentence, an image, etc.). All interactions are cached in the frontend and sent in batches [SV23] (every 10 seconds or when a session is finalized) to a django[‡] backend, where they are stored in a PostgreSQL[§] database.

Table 2: *Task related statistics such as duration or number of mouse moves. Durations are given as fractions of minutes. Relative values (e.g., mouse moves per second) are in parentheses.*

Prop.	Min	Max	Mean	StdDev	Median
task A					
Dur.	0.31 (—)	95.83 (—)	6.58 (—)	8.70 (—)	4.30 (—)
Moves	10 (0.09)	19.37K (95.74)	3.46K (11.83)	3.29K (9.84)	2.28K (9.93)
Clicks	2 (0.01)	238 (0.74)	31 (0.12)	31 (0.11)	20 (0.09)
Scrolls	0 (0.00)	13.13K (18.77)	1.03K (2.73)	2.02K (3.77)	254 (0.81)
Keys	0 (0.00)	1.64K (23.71)	72 (0.30)	153 (1.55)	38 (0.12)
task B					
Dur.	0.26 (—)	27.78 (—)	3.21 (—)	3.30 (—)	2.25 (—)
Moves	8 (0.11)	17.43K (86.46)	2.35K (13.80)	2.56K (9.18)	1.60K (11.61)
Clicks	0 (0.00)	229 (0.76)	21 (0.13)	27 (0.12)	13 (0.09)
Scrolls	0 (0.00)	4.47K (14.45)	454 (2.30)	795 (3.05)	111 (0.72)
Keys	0 (0.00)	6.41K (28.30)	48 (0.26)	414 (1.83)	11 (0.07)
task C					
Dur.	0.19 (—)	12.68 (—)	2.67 (—)	2.05 (—)	2.16 (—)
Moves	8 (0.21)	14.86K (57.99)	2.15K (14.09)	2.19K (8.93)	1.58K (12.25)
Clicks	1 (0.01)	167 (0.66)	22 (0.13)	31 (0.11)	12 (0.10)
Scrolls	0 (0.00)	3.64K (15.76)	350 (2.13)	562 (2.94)	114 (0.86)
Keys	0 (0.00)	7.01K (28.94)	44 (0.25)	453 (1.87)	7 (0.07)
$\Sigma^\dagger$					
Dur.	2.16 (—)	71.34 (—)	11.74 (—)	7.79 (—)	9.90 (—)
Moves	94 (0.14)	44.98K (81.00)	7.76K (12.31)	5.92K (8.64)	6.33K (10.63)
Clicks	9 (0.02)	423 (0.59)	74 (0.13)	57 (0.11)	58 (0.09)
Scrolls	0 (0.00)	16.10K (15.62)	1.88K (2.74)	2.78K (3.38)	512 (0.91)
Keys	0 (0.00)	14.84K (28.07)	171 (0.28)	986 (1.86)	74 (0.11)

[†] This includes only participants who completed all tasks.

Data Filtering. The database in its raw form is not fit for direct interaction analysis as it contains mostly system-related overhead

---

[†] `https://developer.mozilla.org/en-US/docs/Web/API/Event`

[‡] `https://www.djangoproject.com/`

[§] `https://www.postgresql.org/`

(content, logs, temporary data, settings, etc.), but all relevant interaction data was carefully extracted and transformed into an interoperable and reusable [WDA*16] format (Sec. 3.3). This was done mostly automatically, but the correctness of the users' given answers to the IR tasks was evaluated manually. In this migration process, we also filtered out incomplete, irrelevant, or erroneous data. That is, we removed all exploration sessions with a duration of one second or less, which likely resulted from window tabbing or similar interface actions. This was the case for 2% of sessions, affecting about 2‰ of interaction events. We similarly treated all sessions without a finalization context – i.e., cases where the connection was interrupted, the browser window closed unexpectedly, etc. Lastly, we discarded all user tasks with a duration of less than 10 seconds, as we reason that the task was not properly attempted in such cases. All data, orphaned by these filters (e.g., users without any tasks or mouse events without a session) are discarded as well.

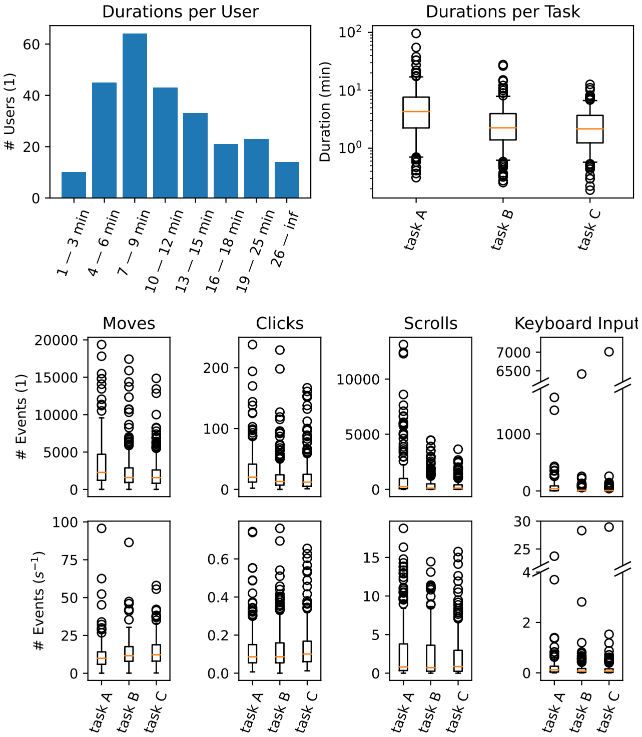


Figure 3: *Task-related statistics of overall durations (top) and absolute/relative counts of different events (bottom).*

3.2. Statistics

From 271 users (having completed 798 tasks), 253 (716) remained after the filtering. The remaining 716 completed tasks comprise a total of 2,911 separate sessions (see Sec. 3.3), containing a total of 2.39 M mouse and keyboard events. For all users, we also collected sociodemographic information regarding age, gender, education, handedness, and amblyopia, as shown in the break down in Fig. 2. The majority of participants belong to the youngest age group (18-30 years). Note that we required participants to (i) be at least 18 years of age; (ii) use a desktop setup (i.e., monitor with mouse and keyboard); and (iii) not suffer from a visual impairment. In terms of gender, we observe a 2:1 majority of female participants and the majority received a secondary or even tertiary education. On the conducted DKT2 test on Diabetes Mellitus on average participants were able to answer 9.42 out of 14 questions correctly ($SD = 1.95$), exhibiting a Normal distribution.

In terms of the duration required for solving the tasks, we observe a wide spread among subjects and tasks, as outlined in Table 2 and Fig. 3. The overall duration for solving all tasks follows a Poisson distribution (mean = 7.79 minutes, median = 9.90 minutes), with two outliers (user IDs 104 and 106) with a duration of 55 and 96 minutes for *task A*, respectively. An inspection of the respective interaction logs reveals that in both cases there are gaps in the data accounting for almost the whole duration. As the exploration view was active during this time, we reason that the respective subjects took a break from the study.

Table 3: *Fraction of correct answers to the IR tasks and the experienced ICL and ECL according to the questions by Krieglstein et al. [KBR*23]. Values in parentheses show the standard deviation.*

		task A	task B	task C	Σ
Correct		0.49	0.80	0.87	0.72
ICL	ICL1	2.30 (2.23)	1.95 (2.03)	1.64 (1.91)	1.96 (2.08)
	ICL2	2.24 (2.25)	1.79 (1.96)	1.45 (1.73)	1.83 (2.02)
	ICL3	3.00 (2.41)	2.95 (2.24)	2.55 (2.27)	2.83 (2.32)
	ICL4	3.52 (2.47)	3.21 (2.41)	2.86 (2.40)	3.20 (2.44)
	ICL5	2.18 (2.27)	1.97 (2.06)	1.46 (1.80)	1.87 (2.07)
	Σ	2.65 (2.39)	2.37 (2.23)	1.99 (2.12)	2.34 (2.26)
ECL	ECL1	3.76 (2.77)	2.88 (2.52)	2.57 (2.39)	3.07 (2.61)
	ECL2	3.21 (2.57)	2.37 (2.32)	2.14 (2.29)	2.57 (2.44)
	ECL3	3.26 (2.64)	2.51 (2.38)	2.27 (2.28)	2.68 (2.48)
	ECL4	4.09 (2.86)	2.76 (2.63)	2.49 (2.49)	3.11 (2.76)
	ECL5	3.09 (2.70)	2.63 (2.51)	2.39 (2.55)	2.70 (2.61)
	Σ	3.48 (2.74)	2.63 (2.48)	2.37 (2.41)	2.83 (2.59)

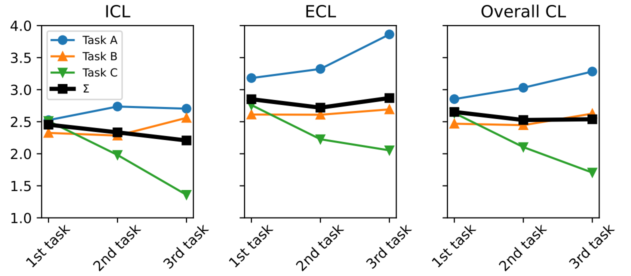


Figure 4: *The mean experienced ICL, ECL, and overall CL of all subjects over the course of the study.*

Comparing tasks, we observe that *task A* took substantially longer to answer (more than twice the time) than *task B* and *task C*. It appears that *task A* is, unintentionally, much harder to answer

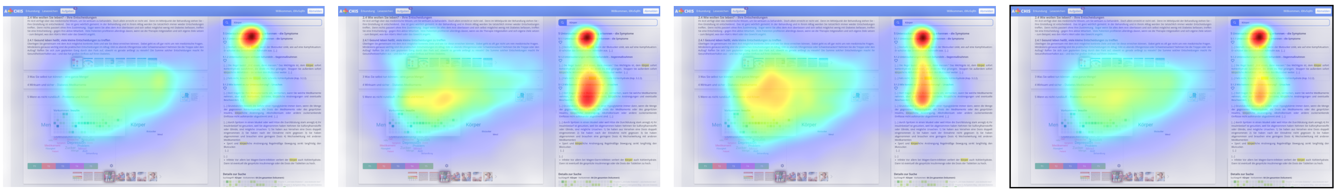

(a) *task A* (b) *task B* (c) *task C* (d) Combined

Figure 5: *Mouse cursor positions for all participants over the course of* task A–task C *respectively (a)–(c) and for all tasks combined (d). The heatmaps reveal two distinct intensity clusters over the center of the main exploration view and the searchbar with adjacent results list.*

than the other two. However, in terms of input activity, the tasks are largely balanced as reflected in Fig. 3, bottom. Although there appears to be slightly more activity for *task A* due to the generally longer durations, this imbalance disappears when looking at the time-normalized statistics. In terms of keyboard inputs, there are two users with substantially more events than all others. An inspection of the respective logs reveals that one subject (user ID 53) employs the arrow keys for scrolling, and another subject (user ID 153) has continuous and unperturbed input from the 0-key on the numpad, which we attribute to a hardware issue.

The assumption that *task A* was harder to solve is also backed-up by the CL scores and the correctness rate, reported in Table 3. We observe that only 49.37% of the participants were able to correctly answer *task A*, while rates were 80.33% and 87.40% for *task B* and *task C*, respectively. The same is true for the CL scores. Compared to the established questionnaire by Krieglstein et al. [KBR*23], we implemented two adjustments: First, the subscale on Germane CL was left out due to recent theoretical developments of the CL theory, which proposes dividing Germane CL into ICL and ECL, which in turn add up to a total CL score (for a review on recent developments of the CL theory see Duran et al. [DZS22]). Second, we slightly reworded some items. For example, some items include the terms ‘learning material’ or ‘learning content’, which could have been confusing for participants in our context. Consequently, we used the terms ‘material’ and ‘content’ (in German: ‘Material’ and ‘Inhalt’). Both the ICL and ECL are substantially higher for *task A* and appear to even increase over the course of the study (Fig. 4).

To get an estimate of the subjects’ general cursor patterns, we generate intensity heatmaps for the cursor positions during both the individual tasks (Fig. 5a–(c)) and overall (Fig. 5d). In order to combine mouse trajectories, we normalized all positions (c.f. Sec. 4) and scaled them to a common 16:9 aspect ratio and overlayed it to a characteristic view of the system. Note that the actually used aspect ratio depends on the subjects’ window setup and the appearance of the exploration view is dynamic as sections can be expanded/collapsed and representations changed. It is, nonetheless, a good approximation for the majority of participants. That is, for all tasks we can observe two intensity clusters reflecting the system’s bipartite layout: (1) a bulgy cluster at the center of the scrollable document view, which constitutes the “main” component, and (2) a strong i-shaped manifestation on the search interface on the right-hand side. The latter can be easily explained through the characteristics of a search process – i.e., a user rests the mouse over the input field while typing and subsequently moves the cursor downwards over the top results. One particularity we can observe is that there is more intense usage of the document view for *task C* which we attribute to that fact that the answer for this task can be found solely in the pictorial content, which is displayed there.

3.3. Data Structure and Organization

The presented dataset was drawn from the A$^+$CHIS’s database, removing all bloat which is not directly related to the behavioral data or relevant for further research. Further steps were taken to comply with the FAIR Guiding Principles for scientific data management and stewardship [WDA*16]. Those include a normalization of the used identifiers. As many of the raw data records were not part of the study or subjected to filtering, the thereof resulting fragmented ID ranges were pruned accordingly. All timestamps, which were mostly recorded in GMT+2 time zone, were transformed to UTF and converted to floating-point Unix timestamps. We organize the data in tabular scheme (Fig. 6), adhering to the second normal form of relational databases [Cod72].

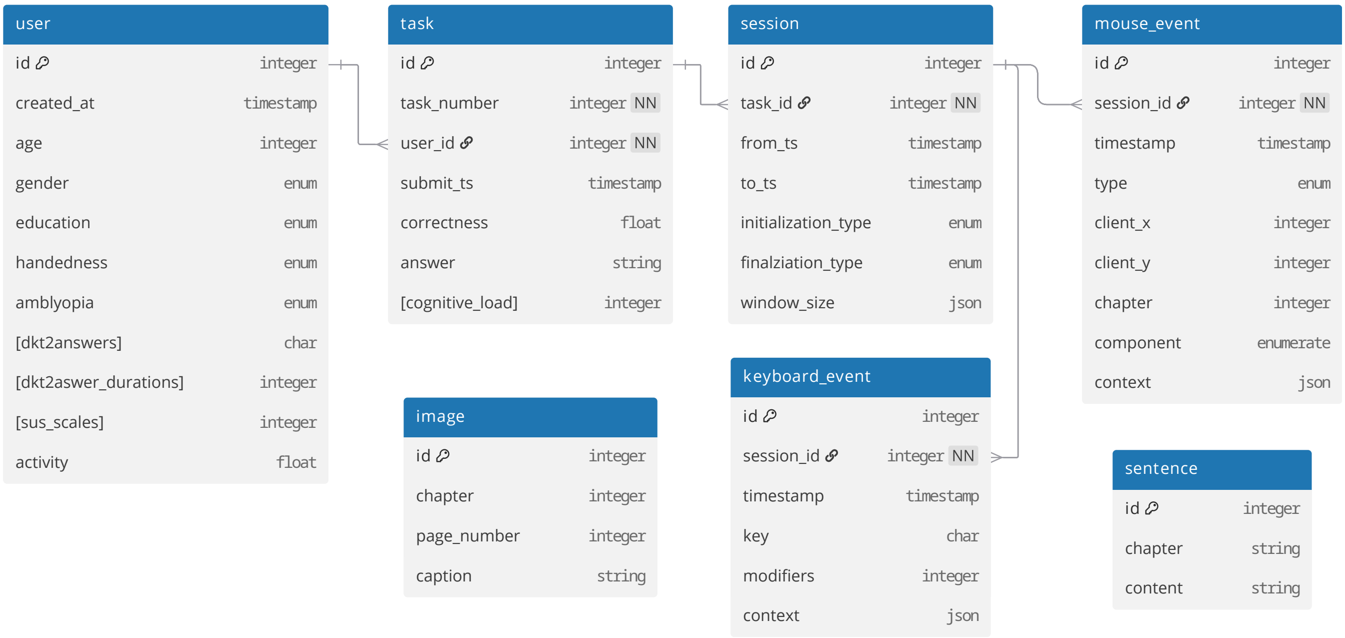
We differentiate between two types of inputs (henceforth referred to as **InputEvents**), triggered by the user:

KeyboardEvent is recorded upon observing a “keydown”[¶] event. Additional to the time of the event, we log the key which has been pressed and the modifiers which have been active during the keypress. The latter allows us to infer the usage of shortcuts. If a type-in element (e.g., a search bar) was in focus at the time of the event, we also log the current input string of this element.

MouseEvent describes a mouse click, -move, or -scroll event. Besides a timestamp, it comprises the cursor’s X/Y coordinates in viewport space[‖], the entire viewing area in which the document is presented in the browser. Note that the value ranges for these coordinates are different from user to user as they depend on the browsers’ window sizes on their respective monitors. Even for a certain user they can change at any moment if the window is resized or altered in another way. Consequently, we also store the viewport sizes at all times (see **Session**) and also provide software for normalizing the mouse trajectories (Sec. 4).

[¶] https://developer.mozilla.org/en-US/docs/Web/API/Element/keydown_event

[‖] https://developer.mozilla.org/en-US/docs/Web/CSS/CSSOM_view/Coordinate_systems



user		
id 🔑	integer	
created_at	timestamp	
age	integer	
gender	enum	
education	enum	
handedness	enum	
amblyopia	enum	
[dkt2answers]	char	
[dkt2aswer_durations]	integer	
[sus_scales]	integer	
activity	float	

task		
id 🔑	integer	
task_number	integer	NN
user_id 🔗	integer	NN
submit_ts	timestamp	
correctness	float	
answer	string	
[cognitive_load]	integer	

session		
id 🔑	integer	
task_id 🔗	integer	NN
from_ts	timestamp	
to_ts	timestamp	
initialization_type	enum	
finalziation_type	enum	
window_size	json	

mouse_event		
id 🔑	integer	
session_id 🔗	integer	NN
timestamp	timestamp	
type	enum	
client_x	integer	
client_y	integer	
chapter	integer	
component	enumerate	
context	json	

image	
id 🔑	integer
chapter	integer
page_number	integer
caption	string

keyboard_event		
id 🔑	integer	
session_id 🔗	integer	NN
timestamp	timestamp	
key	char	
modifiers	integer	
context	json	

sentence	
id 🔑	integer
chapter	string
content	string

Figure 6: *The underlying schema of the GUIDÆTA. Note, that* `image` *and* `sentence` *tables are connected to the* `keyboard_event`/`mouse_event` *tables via their* `context` *attribute in which the respective IDs appear.*

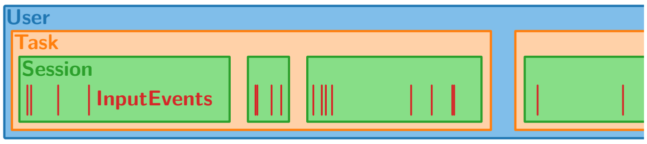


Figure 7: *All **InputEvents** are contained within three hierarchical containers, referring to a specific **User**, **Task**, and time frame of uninterrupted exploration (i.e., **Session**).*

All **InputEvents** are organized in three tiers of containers, as illustrated in Fig. 7, which allow to relate them to a specific **User**, **Task**, and **Session**:

User is the uppermost container and comprises all user-related information, such as sociodemographic parameters, SUS answers, DKT2 answers, or parameters describing the overall activity.

Task is the data frame for the posed IR tasks (up to three per user) and features the given answer, whether it was correct, and the CL indicated by the user through a filled-out questionnaire.

Session is the lowest container and comprises the actual **InputEvents**. This additional tier is necessary since the IR task can be interrupted – either by switching to another (browser) window or other predefined actions which indicate an interruption of the exploration process, i.e., opening the help wizard or task modal. Besides the start and end timestamp of a **Session**, we store its initialization- and finalization context, e.g., “window loosing focus” or “help wizard closed”. As the study was conducted purely unsupervised these session time frames together with their initialization/finalization contexts provide vital cues for detecting illicit behavior, such as using other sources for answering the posed tasks or revealing whether users experience issues with the platform’s components (indicated through frequent help wizard visits). That is, this information can be employed for filtering based on additional inclusion criteria. A **Session** also contains the window sizes as key value-pairs with timestamp, necessary for normalizing the cursor positions. Usually the window size remains constant over the course of a **Session**, but could change if a user, e.g., changes to full screen mode.

The schema in Fig. 6 also features a table **Sentence**, which contains an index list of sentences extracted from the data source [BV21], which we use in A$^+$CHIS. **Sentences** can be linked to some of the **InputEvents** if they contain a reference in their respective contexts (Sec. 3.3.1). Detailed descriptions of the additional attributes of the schema are documented in the software framework (Sec. 4), which we provide for further processing. All data records are publicly available on OSF (`https://osf.io/fhvbm/`). To ensure interoperability [WDA*16] without any custom software, we provide them as CSV files, with the following folder structure, relating to the naming conventions in Fig. 6:

```
/
├─ mouse_events
│  └─ me_<user>_<task>.csv
├─ keyboard_events
│  └─ ke_<user>_<task>.csv
├─ task_answers.csv
├─ sessions.csv
├─ users.csv
├─ sentences.csv
└─ images.csv
```

3.3.1. Event Contexts

Several **MouseEvents** (827 K, i.e., 35.13%) feature additional context information. Besides widget information (the reference to the currently used visual component) and content information (the reference to the currently displayed information unit), we also log specific types of interaction with the system. To this end, we predefined a set of 18 actions which are supported by our system. On a high-level they include things like enlarging content, different interactions with the system's visual components, navigation, bookmarking, and searching. Details on the different actions, which also store related context information (e.g., the particular word, which has been clicked in case of a click in a word cloud), are documented within our data handling software framework (Sec. 4). Fig. 8 shows a breakdown of the occurrences of the different actions. The, by far, most prevalent action is *Hover Sentence* – the event that a user hovers or clicks a particular sentence – followed by expanding text snippets, clicking or hovering a term in the word cloud, enlarging an image, etc. These context data allow to, e.g., (i) determine if users perform clicks, which are not supported by the system; (ii) identify reoccurring interaction cycles; or (iii) determine a preference for certain tools.

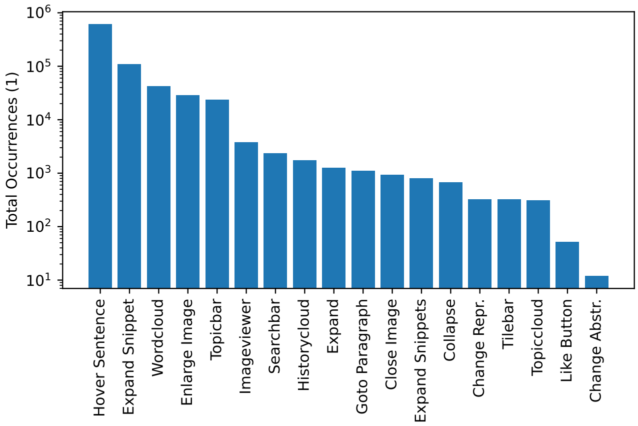


Figure 8: *The 18 predefined actions supported by our system, together with their respective prevalence in the GUIDÆTA dataset.*

4. Data Handling and Analysis

Alongside the described dataset, we also publish the software framework[**] used for processing the data. Written in Python, it allows to easily (i) load all the data records in Fig. 6 as class objects, (ii) filter for specific records (e.g., 'get all **MouseEvents** associated with *task C*'), and (iii) normalize positions. The latter is relevant for **MouseEvents** whose coordinates are stored in viewport space. Together with the `window_size` attribute in their belonging **Session** they can be mapped to a $[0,1)$ value domain (for X and Y independently, or in combined fashion).

Based on the `context` and widget (i.e., `component`) information in the **InputEvents**, we define two higher-level interaction types. First, a **Dwelling**, which constitutes an unperturbed interaction with a certain visual component for a period of time. As such, it has start- and end timestamp as well as information pertaining to the used widget and the displayed content. Second, a **Hovering**, which is an interval of an unperturbed hovering over a certain element. This differs from a **Dwelling** in the sense that it is a transient interaction which can be interrupted by clicks, scrolls, or keyboard inputs. **Hoverings** are thus often enclosed within a **Dwelling**. We note that these interactions merely serve as a starting point for behavior modeling or other purposes and can be adjusted or extended depending on given requirements.

In an effort to capture as much information as possible, our collected mouse positions stem from an event-driven sampling scheme – i.e., the HTML events triggered by the used browsers. This results in a very high sampling rate (up to 95 Hz, c.f. Table 2) in most cases. However, events are unevenly spaced in both time and space. For various trajectory analyses, the (re)sampling of the cursor positions constitutes a necessary prerequisite. In general, we differentiate between spatial-based or time-based normalization [KHW*19]. With the prior, the mouse trajectory is represented by a series of locations with uniform distances. Likewise, with the latter, the mouse trajectory is represented by a series of locations evenly spaced in time. The time-based normalization is beneficial for the comparison of short-time actions [SGK05] while the spatial-based normalization is advantageous for the spatial analysis of taken paths and long-term processes. Both approaches are implemented in our software framework (Fig. 9), in which the user is able to provide custom sample frequencies or distances. The framework was also employed to generate all statistics figures shown in this paper.

5. Potential Limitations and Future Work

We believe that the published dataset constitutes a valuable contribution for interaction-driven studies in different application domains. Nonetheless, we want to point out potential issues which potentially limit the usage of the data for specific purposes. Even though we strive for diversity in terms of gender, background, and ages in our set of participants, the subjects which actually took part in the study are predominantly from a young age group, female and of high education backgrounds (Fig. 2). This has to be considered when interpreting behavioral patterns and drawing conclusions regarding generalizability.

The unsupervised nature of the study setup enabled a broad outreach and allowed us to efficiently collect data from hundreds of participants. On the downside, this left us with limited control over the hardware and environment setup subjects used to conduct the study. While some properties, such as the window sizes over the duration of the study are closely monitored and logged, other factors, such as the mouse sensitivity, double-click threshold, dpi, and all properties pertaining to the look and feel of the system are beyond our control. That means that – although we log which components and contents were displayed at a specific time – we cannot faithfully reconstruct the actual representation seen by the user.

The context information (Sec. 3.3.1), associated to **InputEvents**, is strongly tailored to the A^{+}CHIS, with clearly defined event types such as 'expanding a chapter' or 'bookmarking a certain piece of

---

[**] `https://github.com/lenxn/apchis-guidaeta.git`

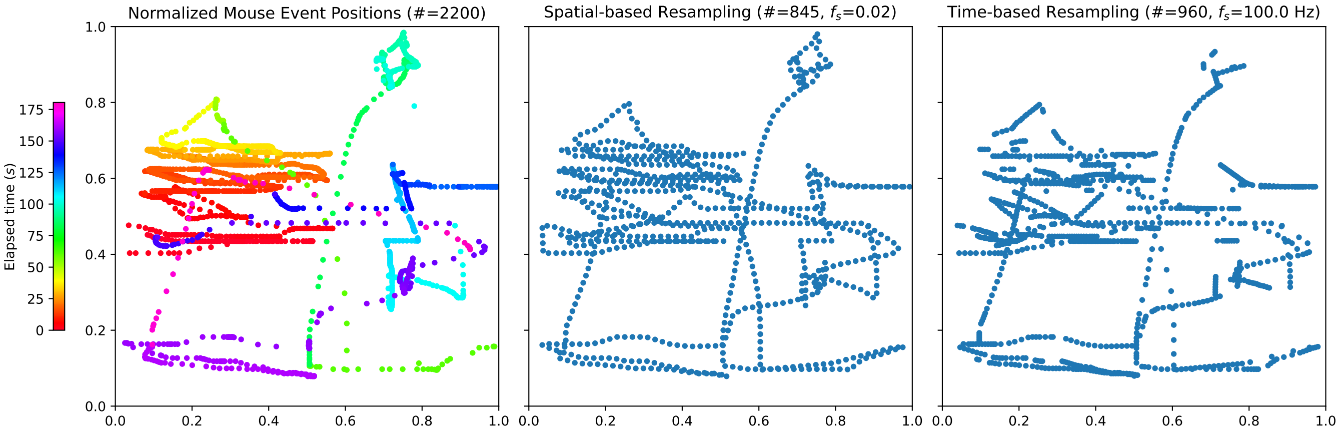


Figure 9: *Left: Normalized mouse positions for a randomly selected exploration session of approximately 3 minutes. Middle: The spatial-based resampling (with a sampling frequency f_s of 2% of the screen space's diagonal) is able to faithfully capture the overall trajectory while reducing the number of positions by more than half. Right: With time-based resampling ($f_s = 100$ Hz), some trajectory features could get lost, while it is still beneficial for some applications.*

information'. While this limits the generalizability of our interaction tracking (which was purposely developed for this study), it provides very detailed information on how an exploration process was conducted, painting a very clear picture of users' behavior. We can, e.g., evaluate the time a user spends on processing a patch of information as opposed to the time spent on finding new information patches – something that is studied within the *information foraging theory* [PC99].

Regarding future work, we also intend to use the dataset to study the effectiveness of different behavioral indicators for measuring cognitive load. Also, building upon previous work [LST*24], we develop and evaluate visual interfaces, which enable expert users to effectively review behavioral patterns and reveal between-subject similarities [XOW*20]. Beyond that, the dataset is well-suited for experimenting with user classification and development of CL over time. As we also report the correctness of the given task answers (reflecting users' success), it even allows to study which factors/interactions lead to either a successful completion or a failure.

6. Conclusion

With the GUIDÆTA, we present an encompassing dataset obtained from 253 individual users. The core strengths – besides the sheer volume – are the rich context and content information associated with most **InputEvents**, as well as various additional metadata pertaining to sociodemographic information about participants, answers from various feedback questionnaires, and experienced CL. We believe that, together with the tools for data handling and processing, GUIDÆTA can be employed in versatile fashion and constitutes thus a valuable contribution to different research domains.

Acknowledgments

This work was funded by the Austrian Science Fund (FWF) as part of the project 'Human-Centered Interactive Adaptive Visual Approaches in High-Quality Health Information' (A$^+$CHIS; Grant No. FG 11-B).

References

[ABF21] Antal M., Buza K., Fejer N.: SapiAgent: A Bot Based on Deep Learning to Generate Human-Like Mouse Trajectories. *IEEE Access 9* (2021), 124396–124408. doi:10.1109/ACCESS.2021.3111098. 2

[ADF19] Antal M., Denes-Fazakas L.: User Verification Based on Mouse Dynamics: a Comparison of Public Data Sets. In *2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (May 2019), pp. 143–148. doi:10.1109/SACI46893.2019.9111596. 1, 2

[AFB21] Antal M., Fejér N., Buza K.: SapiMouse: Mouse Dynamics-based User Authentication Using Deep Feature Learning. In *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (May 2021), pp. 61–66. doi:10.1109/SACI51354.2021.9465583. 2

[AMFVR22] Acien A., Morales A., Fierrez J., Vera-Rodriguez R.: BeCAPTCHA-mouse: Synthetic mouse trajectories and improved bot detection. *Pattern Recognition 127* (2022), 108643. Publisher: Elsevier. doi:10.1016/j.patcog.2022.108643. 1

[Bro96] Brooke J.: SUS – a quick and dirty usability scale. In *Usability Evaluation in Industry*. London, England, 1996, pp. 189–194. 4

[BV21] Baumgart J., Viegener U.: *Den Diabetes Im Griff: Ein Handbuch Für Patientinnen Und Patienten Mit Diabetes Mellitus Typ 2*. KomPart Verlagsgesellschaft mbH & Company KG, 2021. 7

[CLZM17] Chen Y., Liu Y., Zhang M., Ma S.: User Satisfaction Prediction with Mouse Movement Information in Heterogeneous Search Environment. *IEEE Transactions on Knowledge and Data Engineering 29*, 11 (Nov. 2017), 2470–2483. doi:10.1109/TKDE.2017.2739151. 2

[Cod72] Codd E. F.: Further normalization of the data base relational model. *Data base systems 6*, 1972 (1972), 33–64. 6

[DZS22] Duran R., Zavgorodniaia A., Sorva J.: Cognitive load theory in computing education research: A review. *ACM Transactions on Computing Education (TOCE) 22*, 4 (2022), 1–27. 6

[FEM*12] Feher C., Elovici Y., Moskovitch R., Rokach L., SCHCLAR A.: User identity verification via mouse dynamics. *Information Sciences 201* (Oct. 2012), 19–36. doi:10.1016/j.ins.2012.02.066. 2

[FFA*16] FITZGERALD J. T., FUNNELL M. M., ANDERSON R. M., NWANKWO R., STANSFIELD R. B., PIATT G. A.: Validation of the revised brief diabetes knowledge test (dkt2). *The Diabetes Educator 42*, 2 (2016), 178–187. doi:10.1177/0145721715624968. 4

[Fit54] FITTS P. M.: The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology 47*, 6 (1954), 381–391. doi:10.1037/h0055392. 1

[FKKWP16] FÜLÖP A., KOVÁCS L., KURICS T., WINDHAGER-POKOL E.: Balabit Mouse Challenge Data Set, 2016. URL: https://github.com/balabit/Mouse-Dynamics-Challenge. 2

[GA10] GUO Q., AGICHTEIN E.: Towards predicting web searcher gaze position from mouse movements. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta Georgia USA, Apr. 2010), ACM, pp. 3601–3606. doi:10.1145/1753846.1754025. 2

[GW09] GOTZ D., WEN Z.: Behavior-driven visualization recommendation. In *Proceedings of the 14th international conference on Intelligent user interfaces* (2009), IUI '09, Association for Computing Machinery, pp. 315–324. doi:10.1145/1502650.1502695. 1

[HAG15] HINBARJI Z., ALBATAL R., GURRIN C.: Dynamic User Authentication Based on Mouse Movements Curves. In *MultiMedia Modeling*, vol. 8936. Springer International Publishing, Cham, 2015, pp. 111–122. doi:10.1007/978-3-319-14442-9_10. 2

[Hic52] HICK W. E.: On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology 4*, 1 (Mar. 1952), 11–26. doi:10.1080/17470215208416600. 1

[HWB12] HUANG J., WHITE R., BUSCHER G.: User see, user point: Gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2012), CHI '12, Association for Computing Machinery, pp. 1341–1350. doi:10.1145/2207676.2208591. 2

[KBR*23] KRIEGLSTEIN F., BEEGE M., REY G. D., SANCHEZ-STOCKHAMMER C., SCHNEIDER S.: Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review 35*, 1 (2023), 9. doi:10.1007/s10648-023-09738-0. 4, 5, 6

[KDMH24] KHAN S., DEVLEN C., MANNO M., HOU D.: Mouse Dynamics Behavioral Biometrics: A Survey. *ACM Comput. Surv. 56*, 6 (Feb. 2024), 154:1–154:33. doi:10.1145/3640311. 1, 2, 3

[KFH*21] KHAN S., FRASER C., HOU D., BANAVAR M., SCHUCKERS S.: Authenticating facebook users based on widget interaction behavior. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)* (2021), pp. 1–8. ISSN: 2331-9860. doi:10.1109/CCNC49032.2021.9369604. 1, 2

[KHW*19] KIESLICH P. J., HENNINGER F., WULFF D. U., HASLBECK J. M., SCHULTE-MECKLENBECK M.: Mouse-tracking: A practical guide to implementation and analysis 1. In *A handbook of process tracing methods*. Routledge, 2019, pp. 111–130. 8

[KN18] KATERINA T., NICOLAOS P.: Mouse behavioral patterns and keystroke dynamics in End-User Development: What can they tell us about users' behavioral attributes? *Computers in Human Behavior 83* (June 2018), 288–305. doi:10.1016/j.chb.2018.02.012. 1

[KYA21] KILIÇ A. A., YILDIRIM M., ANARIM E.: Bogazici mouse dynamics dataset. *Data in Brief 36* (June 2021), 107094. doi:10.1016/j.dib.2021.107094. 2

[LA20] LEIVA L. A., ARAPAKIS I.: The attentive cursor dataset. *Frontiers in Human Neuroscience Volume 14 - 2020* (2020). doi:10.3389/fnhum.2020.565664. 1, 2

[LGL25] LATIFZADEH K., GWIZDKA J., LEIVA L. A.: A versatile dataset of mouse and eye movements on search engine results pages. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2025), pp. 3412–3421. 3

[LST*24] LENGAUER S., SHAO L., TYTARENKO M., KLAFFENBÖCK M., SCHRECK T.: Interaction Visualization for Analysing and Improving User Models. In *32nd ACM Conference on User Modeling, Adaptation and PersonalizationJune 2024: UMAP'2024* (2024). doi:10.1145/3631700.3664877. 9

[MKSY23] MEYER T., KIM A. D., SPIVEY M., YOSHIMI J.: Mouse tracking performance: A new approach to analyzing continuous mouse tracking data. *Behavior Research Methods 56*, 5 (Sept. 2023), 4682–4694. doi:10.3758/s13428-023-02210-5. 1, 2

[MLZM14] MAO J., LIU Y., ZHANG M., MA S.: Estimating Credibility of User Clicks with Mouse Movement and Eye-Tracking Information. In *Natural Language Processing and Chinese Computing*, vol. 496. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 263–274. doi:10.1007/978-3-662-45924-9_24. 2

[Mor05] MORETTI F.: *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005. 3

[MSK*90] MEYER D. E., SMITH J. E. K., KORNBLUM S., ABRAMS R. A., WRIGHT C. E.: Speed—Accuracy Tradeoffs in Aimed Movements: Toward a Theory of Rapid Voluntary Action. In *Attention and Performance Xiii*. Psychology Press, 1990. 1

[PC99] PIROLLI P., CARD S.: Information foraging. *Psychological Review 106*, 4 (1999), 643–675. doi:10.1037/0033-295X.106.4.643. 9

[SCG12] SHEN C., CAI Z., GUAN X.: Continuous authentication for mouse dynamics: A pattern-growth approach. In *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012)* (June 2012), pp. 1–12. ISSN: 2158-3927. doi:10.1109/DSN.2012.6263955. 1, 2

[SGK05] SPIVEY M. J., GROSJEAN M., KNOBLICH G.: Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences 102*, 29 (July 2005), 10393–10398. doi:10.1073/pnas.0503903102. 8

[SLM*25] SHAO L., LENGAUER S., MIRI H., BEDEK M., KUBICEK B., KUPFER C., ZANGL M., DIENSTBIER B., JEITLER K., KRENN C., SEMLITSCH T., ZIPP C., ALBERT D., SIEBENHOFER A., SCHRECK T.: Visual Document Exploration with Adaptive Level of Detail: Design, Implementation and Evaluation in the Health Information Domain. In *14th International Conference on Information Visualization Theory and Applications* (Sept. 2025), pp. 133–141. 1, 3

[SV23] SADEGHPOUR S., VLAJIC N.: Remouse dataset: On the efficacy of measuring the similarity of human-generated trajectories for the detection of session-replay bots. *Journal of Cybersecurity and Privacy 3*, 1 (2023), 95–117. doi:10.3390/jcp3010007. 1, 3, 4

[WDA*16] WILKINSON M. D., DUMONTIER M., AALBERSBERG I. J., APPLETON G., AXTON M., BAAK A., BLOMBERG N., BOITEN J.-W., DA SILVA SANTOS L. B., BOURNE P. E., ET AL.: The fair guiding principles for scientific data management and stewardship. *Scientific data 3*, 1 (2016), 1–9. 5, 6, 7

[WZC19] WEI A., ZHAO Y., CAI Z.: A deep learning approach to web bot detection using mouse behavioral biometrics. In *Biometric Recognition*, vol. 11818. Springer International Publishing, 2019, pp. 388–395. Series Title: Lecture Notes in Computer Science. doi:10.1007/978-3-030-31456-9_43. 1

[XOW*20] XU K., OTTLEY A., WALCHSHOFER C., STREIT M., CHANG R., WENSKOVITCH J.: Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum 39*, 3 (June 2020), 757–783. doi:10.1111/cgf.14035. 9

[YCON25] YANEZ F., CONATI C., OTTLEY A., NOBRE C.: The state of the art in user-adaptive visualizations. *Computer Graphics Forum 44*, 1 (2025), e15271. doi:10.1111/cgf.15271. 1

[ZPW16] ZHENG N., PALOSKI A., WANG H.: An Efficient User Verification System Using Angle-Based Mouse Movement Biometrics. *ACM Transactions on Information and System Security 18*, 3 (Apr. 2016), 1–27. doi:10.1145/2893185. 2