Attract & Engage Visitors for Tabular Data Maintenance: a Longitudinal Naturalistic Field Study

SHAUN WALLACE, University of Rhode Island, USA
NA KYOUNG LEE*, Brown University, USA
ZHENGYI PENG*, Brown University, USA
TALIE MASSACHI*, University of Pennsylvania, USA
LONG DO, Brown University, USA
SURBHI RATHORE, University of Rhode Island, USA
SARAH M. BROWN, University of Rhode Island, USA
JEFF HUANG, Brown University, USA

Maintaining an evolving tabular dataset over time requires attracting and engaging enough accurate, unpaid contributors to visit and maintain it. We developed Drafty, a large publicly accessible tabular dataset of Computer Science faculty profiles. We observed Drafty's visitors' edits to the data, the accuracy of those edits, and engagement with the system. In March 2022, we released new add-ons to create four "sources" to attract more visitors: organic, static, asking, and dynamic. Our dynamic sources create **automatic insights** from our evolving tabular dataset by automatically creating aggregate statistics and sharing them on CS Open Rankings and Twitter. Visits from automatic insights made 3 times more edits than Drafty's everyday organic visitors. Also, visits from our low-effort automatic dynamic sources and high-effort asking sources (i.e., friendsourcing) had similar accuracy. Our longitudinal study demonstrates that new features requiring minimal human effort (i.e., automatic insights) can help attract and sustain an anonymous community of visitors to maintain our large, evolving tabular dataset.

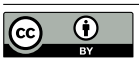
CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Computer supported cooperative work**; • **Information systems → Crowdsourcing**; *Data cleaning*; *Asynchronous editors*; Incomplete data.

Additional Key Words and Phrases: evolving tabular data, crowdsourcing, peer production, recommendation systems, sourcing

ACM Reference Format:
Shaun Wallace, Na Kyoung Lee, Zhengyi Peng, Talie Massachi, Long Do, Surbhi Rathore, Sarah M. Brown, and Jeff Huang. 2026. Attract & Engage Visitors for Tabular Data Maintenance: a Longitudinal Naturalistic Field Study. *Proc. ACM Hum.-Comput. Interact.* 10, 6, Article CSCW053 (October 2026), 41 pages. https://doi.org/10.1145/3816901

*Denotes Equal Contributions

Authors' Contact Information: Shaun Wallace, University of Rhode Island, Kingston, RI, USA, shaun.wallace@uri.edu; Na Kyoung Lee, Brown University, Providence, RI, USA, na_kyoung_lee@alumni.brown.edu; Zhengyi Peng, Brown University, Providence, RI, USA, zhengyi_peng@alumni.brow.edu; Talie Massachi, University of Pennsylvania, Philadelphia, PA, USA, taliem@seas.upenn.edu; Long Do, Brown University, Providence, RI, USA, long_do@brown.edu; Surbhi Rathore, University of Rhode Island, Kingston, RI, USA, surbhi_rathore@uri.edu; Sarah M. Brown, University of Rhode Island, Kingston, RI, USA, brownsarahm@uri.edu; Jeff Huang, Brown University, Providence, RI, USA, jeff_huang@brown.edu.



This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 2573-0142/2026/10-ARTCSCW053
https://doi.org/10.1145/3816901

1 Introduction

How can we motivate knowledgeable users to visit and contribute to a public tabular dataset with little or no effort from the data curators? Creating and maintaining accurate, large, evolving tabular datasets over short periods or multiple years requires sustained efforts from data curators and a community [127, 142].

Visitors to public systems often lurk by visiting and not contributing to the source data [10]. For example, a lack of editors can cause articles on Wikipedia to become out of date [27, 86]. An alternative, given sufficient financial resources, is paying crowdworkers to collect or maintain data [106, 142, 154]. However, this is not financially feasible for longitudinal data maintenance efforts [30, 81, 120]. In addition to financial constraints, recent research indicates that the quality of paid crowdworkers is declining due to low effort and shifting motivations [41, 53, 80, 92]. An automated alternative to paying crowdworkers is to use web scraping to create tabular datasets from publicly available sources. This is difficult to scale because web page design is inconsistent and changes over time [20]. Someone needs to continually update their web scraper (a challenging task [34]) to keep up with every evolving information source. One additional alternative is to rely on knowledgeable community members as unpaid contributors [141, 142], if we can sustainably attract and motivate these users to contribute. Other alternatives include asking others to fix data [13] or having systems nudge users to contribute [118, 143]. However, the extent to which these methods motivate contributions is short-lived, and some, such as nudges to contribute, are considered manipulative [118]. Thus, we examine how to design and develop automated systems that naturally encourage visitors to contribute to our public online spreadsheet system, Drafty.

Drafty provides a publicly editable spreadsheet of Computer Science faculty profiles that we have developed and maintained since 2014. Our system has attracted over a million visitors and has been widely used in the Computer Science community. The Drafty dataset is used by the NSF to study diversity gaps among Computer Science departments [102], and prospective students worldwide use Drafty to identify potential advisors and select universities to attend.

Our tabular dataset contains all publicly available information and requires frequent updates, as it evolves annually (e.g., new hires) and is featured in rankings (which motivate the community). We have recruited paid crowdworkers or employed nudges to maintain it, but these efforts proved unsustainable. Towards late 2019, we struggled to attract enough visitors and contributions to maintain it. Thus, we developed new systems informed by research on enhancing user agency and motivation in public systems [29, 44, 117, 150].

Diversifying the approach to attract visitors and studying the increase in unpaid contributions mirrors the recommendation to study natural behavior in social computing systems [12]. Previous work has called for continued efforts to study how to attract visitors [142]; we explore this using four different sourcing methods to attract visitors to our system:

- **Organic sources**: Our spreadsheet system Drafty, updated in 2020, includes search engine optimizations. It attracts organic visitors by offering thousands of rows of tabular search indexes, which are directly indexed and searchable through Google. Similar to peer production, for example, how Wikipedia's articles are search-indexed and attract organic visitors through search engines [6].
- **Static sources**: Posts on static websites, for example Reddit, including analyses of our tabular dataset at a single point in time, suggested by [143, 151]. These posts are static; their analyses do not update as the dataset changes.
- **Asking sources**: We ask CS community members on Twitter and over email (newly hired faculty and department chairs) to visit the dataset, using a call-to-action. Inspired by friend-sourcing [13, 14, 32].

- **Dynamic sources**: These public sources (CS Open Rankings and the Databaits API) automatically update with insights from Drafty's tabular dataset to create reciprocity [24]. The Databaits API posts insights on Twitter. Prior CSCW research [118, 142] suggests using these low-effort dynamic sources to attract visitors.

One of our dynamic sources, the Databaits API, powers Drafty's Did You Know feature, allowing visitors to freely create and share automated insights. This system design decision answers the call for future work made in [8, 29, 142], thereby supporting interactions within Drafty, for example, making edits, spending time on the system, and creating insights from its data.

Using these sourcing methods to attract visitors, we can conduct a naturalistic field study where everyday unpaid contributors freely visit and edit our evolving tabular dataset in Drafty. We ask the following research questions:

- **RQ1)** What methods (i.e., sources) of **attracting** visitors to Drafty increase behaviors that help maintain its evolving tabular dataset (measured by edits per visit, accuracy of edits, and engagement per visit)?
- **RQ2)** Is freely interacting with the Databaits API in Drafty associated with increased positive behaviors with Drafty (measured by edits per visit, accuracy of edits, and engagement per visit)?

In the context of this paper, edits per visit, conversion rate, and accuracy of edits per visit directly map to behaviors that help maintain the tabular dataset. We define engagement per visit as two metrics: number of interactions (see the list of interactions in Table 7), and session length. We show that our **low-effort automated systems** (dynamic sources that generate and share insights from our tabular dataset) **attract visitors who make accurate edits** comparable to or exceeding conventional approaches (organic discovery [142], static forum posts [143], and asking through friendsourcing [13]), without relying directly nudging [118]. We also find that visitors prefer **randomized insights to similar ones**, building on prior research [140]. We open-source our systems and datasets[1], (2) the CS professors tabular dataset, (3) CS Open Rankings [139], and (4) the Databaits API.

2 Related Work

2.1 Motivating Accurate Contributions and Engagement from Visitors to a Public Dataset

Building on peer production research, a public data system like Drafty requires a sustained flow of users to review data and contribute to maintain up-to-date information [113, 142]. Crowd contributors, for example, paid crowdworkers on Prolific, Amazon Mechanical Turk workers, and Wikipedia editors, choose whether to complete tasks for various reasons: payment (or lack thereof) [15, 38], time to complete [22], task difficulty [48, 60], and level of interest or helping their community [90, 91]. Our work primarily focuses on visitors and contributors to our real-world system from various sources who are interested in its tabular data and in improving it to benefit their community.

Previous research has explored the application of Wikipedia-style collaborative editing to enhance user agency on public platforms [6]. Similar to our system, Li et al. studied the trade-offs of bringing Wikipedia-style collaborative editing into Stack Overflow, traditionally a Q&A site [74]. They allowed users to collaboratively edit anyone's questions and answers, showing that it increased the quality of contributions. In studies on the Zhihu platform, He et al. show that collaborative editing for producing high-quality research-sensemaking may be challenging for non-experts [45]. They describe disparities using qualitative results and correlations between broad participation and the quality of contributions. Similar to our analysis, these two longitudinal studies do not make causal claims across years. Similar CSCW research on user contributions to Q&A sites indicates that time-limited promotional gamification does not generate lasting, accurate contributions and engagement across all users because of varying intrinsic motivations [44]. We build on these studies by developing intrinsic

[1] [https://github.com/drafty-research]: Drafty Systems (with our PolyMu library and the Databaits API) and Datasets

motivators that continuously run and update for our community to increase user engagement with automatic insights [95].

Intrinsic motivators can be better studied, when a public system removes money as an extrinsic motivator for participation, the motivations of organic unpaid contributors can be studied [11, 137]. While this kind of public deployment is a relatively high-risk scenario, its results will have high ecological validity [12]. Prior research shows that a system must be transparent and provide users with agency to earn and maintain their trust, motivating continued participation [8, 74, 93, 117]. Thus, this paper focuses on a public deployment and methods to motivate unpaid visitors to become contributors and engage with our systems.

We first consider user agency in systems maintained solely based on crowd contributions. Prior research shows that unpaid contributors can make accurate edits across data types requiring varying domain expertise [142]. However, Wallace et al. [143] also show that while a system asking users to fix data matching their interests produces highly accurate edits, users rarely completed the request. Similarly, prior research shows how to craft messages to nudge users to voluntarily contribute their time to altruistic causes [26, 118]. However, nudging can be considered manipulative, and users' contributions and participation are short-lived [118, 142]. This highlights the importance of agency in this context. Drafty needs its visitors to contribute, not just consume, to maintain its evolving data. Thus, Drafty requires a mechanism to motivate contributions without interrupting users during natural information-seeking tasks.

Our paper's systems and interactive features, such as Drafty, CS Open Rankings, and the Databaits API, preserve user agency by enabling transparent, automated advertising of our dataset on Twitter and the web. As in SuggestBot [26] and Botivist [118], prior research uses sentence templates to attract contributions of time to specific causes automatically. However, this prior research nudges people to contribute, rather than allowing the agency to contribute freely like in a peer production system [8]. Our systems (Drafty, CS Open Rankings, and the Databaits API) provide freedom to freely contribute and view source data, aligning them with peer production systems [83].

Prior work shows that paid crowdworkers produce higher-quality contributions when they can learn something new while participating [26, 111, 122]. While the Databaits API and Did You Know extend Drafty's functionality to study user engagement within the system, the automatic insights Databaits produces and tweets also serve as one method to attract new and existing users by allowing them to learn something new. Prior research has explored several methods to encourage voluntary contributions, including micro-volunteering [10], friendsourcing [13, 14], engaging in interesting tasks [90], and asking paid crowdworkers to contribute to the greater good [115]. Even popular peer production systems, such as Wikipedia, demonstrate that asking people to contribute can yield accurate contributions [32]. Prior work shows that extrinsic motivators do not increase the quality of intrinsically motivated users' contributions [58, 144]. Will visitors recruited by different methods lurk or contribute? This paper presents four methods to attract visitors to Drafty. This paper will evaluate these recruitment methods within the context of a real-world system.

In summary, when researchers deployed systems in the wild, they found it imperative to balance users' trust and agency with the system [117]. Users should be free to interact with a system, as in peer production, but their contribution tasks should be directed and controlled, as in crowdsourcing [8]. For Drafty, we develop add-ons and deploy them in a naturalistic scenario to attract anonymous visitors to make unpaid contributions to its evolving dataset.

2.2 Using Natural Language Descriptions of Data to Engage Visitors

Understanding how to translate tabular data into human-understandable text can be complex. Our research develops CS Open Rankings and the Databaits API to automate this task. We avoid the cold-start problem in user recommendation, a common challenge for other recommendation systems

like Movielens [97]. Prior work explores methods for translating tabular and structured data into human-readable insights using natural language [87, 148, 149]. Still, there has been little research in the CSCW community on their effects on human behavior in the wild.

Producing natural language from tabular data is a rich research area.Many systems can perform natural language generation tasks, creating text from non-textual data such as tables. Most prior systems use machine learning models. For example, Langkilde and Knight used a statistical language n-gram-based system [71], while Sutskever et al. used a neural net model [131]. Table question answering models respond to natural language queries using tabular data provided at query time to produce factually correct responses [46, 75, 108]. Zhang's system generates representative statements about tables using a neural network model [153]. Whereas in the Databaits API, the statements produced by Databaits are insights that draw on specific cells in the data. Parikh proposed a similar task of generating a single sentence from a table and selected columns, analogous to the methodology behind Databaits generation [107]. Mahapatra [78] proposes a knowledge-light, non-rule-based system to create relevant natural language phrases from input data.

Previous work has used a template-based system to generate fun facts from Wikipedia tables semi-automatically (Korn et al., [68]). Using the Databaits API, we aim to generate valuable insights from large tabular datasets with thousands of rows. While their study focuses on identifying the tables within an article that would most likely produce interesting facts, the Databaits API focuses on producing varied, interesting insights from one large tabular dataset. Both systems require manually creating sentence templates to regenerate different insights for users. Adeyanju [2] found that knowledge-intensive systems (using templates or rules) typically outperform knowledge-light systems.

Our Databaits API creates insights from Drafty's tabular dataset to act as boundary objects to appeal to our community [62]. By focusing on one large dataset of thousands of rows, our efforts build on prior research that links small related tables on Wikipedia [36]. SuggestBot also shows how to structure fact-based sentences to maximize their ability to motivate people to interact using "High-Involvement" or personally relevant examples [26]. Research shows this style increases long-term retention among paid crowdworkers participating in conversational micro-tasks [111]. Similar to prior work, the Databaits API aims to automatically generate insights from tabular data to attract, engage, and motivate users to visit and contribute to Drafty.

3 A Public Platform for Maintaining Evolving Tabular Data: Drafty an Interactive Web-based Spreadsheet, CS Open Rankings, and the Databaits API

Since 2016, we have iteratively built and maintained a public spreadsheet system to motivate anonymous visitors to help maintain an evolving tabular dataset of CS faculty from the US and Canada. Figure 1 shows a timeline of Drafty's development. The authors of this paper have led various efforts to maintain this evolving tabular dataset over 10 years.

Drafty provides free access and open editing to large tabular datasets (thousands of rows) via a custom spreadsheet interface. Drafty hosts a tabular dataset of tenure-track Computer Science profiles from top Universities in the US and Canada. Each profile includes their full name, the university at which they are employed, the year they joined the university as a tenure-track professor, their research area of expertise (i.e., subfield), and the names of the institutions that awarded their Bachelor's and Doctorate degrees.

We have developed add-ons to Drafty that use its tabular dataset to create rankings and insights (CS Open Rankings and the Databaits API). CS Open Rankings is a meta-ranker of Computer Science departments in the US and Canada that includes four rankings that users can toggle to create their own ranking. The Databaits API uses Drafty's tabular data to generate insights by using the column names and values to create statistics and plug them into sentence templates.

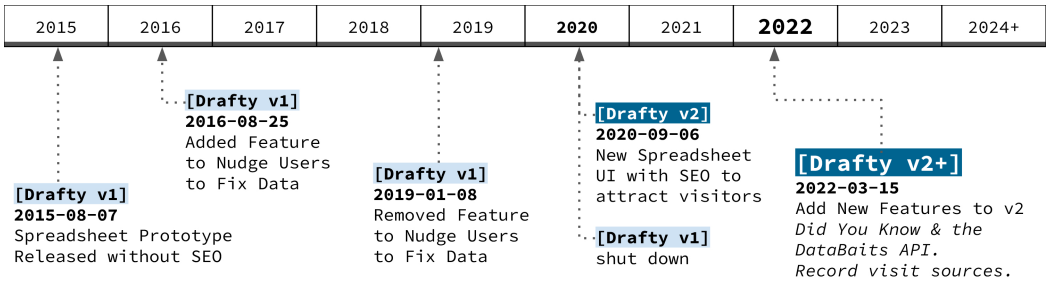


Fig. 1. The timeline for developing different versions of Drafty. We primarily use Drafty v2+ for our analysis in this study. From 2020 to early March 2021, we introduced several editing features in v2, including the ability to add or delete rows.

On March 15th, 2022, we integrated the Databaits API within Drafty's spreadsheet interface as the new feature Did You Know. Anyone can freely see and generate insights from its CS professor profiles using Did You Know's Modal with the Databaits API and have the system post them to our Twitter account.

3.1 Drafty: a Web-based Spreadsheet for Public Tabular Data Maintenance

Organic visitors can freely visit our public web application Drafty using search engines, typing the URL in the browser, or by people sharing a link. Before our efforts described in this paper, we maintained the tabular dataset in a different web-based spreadsheet from 2014 to 2020. However, this system became difficult to maintain and was not easily discoverable on search engines. We developed a new version for this paper, Drafty v2 (see Figure 1 for a timeline). We released the prototype in 2020, replacing the old version. From 2020 to 2021, we added new features, including the ability to add and delete rows. We released add-ons in 2022; see the interface in Figure 2

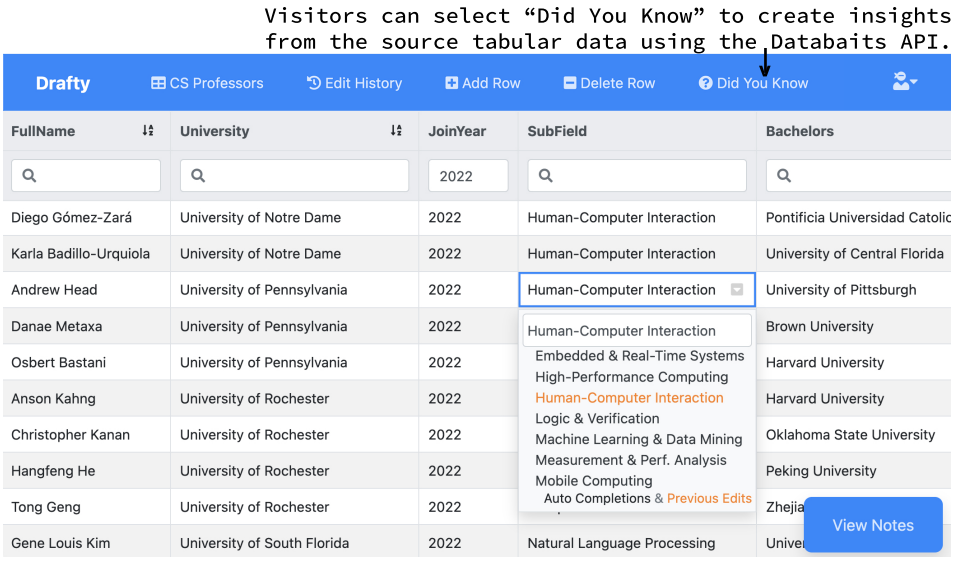


Fig. 2. Drafty v2 spreadsheet interface, we developed to support low-effort public edits, infinite scrolling, and thousands of rows of search-indexed client-side data. Please see Appendix section A for Table 7 to see the interactions Drafty anonymously records.

3.2 CS Open Rankings: a Meta-Ranker from Drafty's Computer Science Professor Profiles

We developed CS Open Rankings as an alternative approach to computer science department rankings (see Appendix Figure 7). Rankings are an ideology, and each can be biased. CS Open Rankings is a meta ranker using real-time data from CS Professors (Drafty) as one of its sources. Its rankings and data are updated dynamically every five minutes. In CS Open Rankings, users freely choose and combine existing rankings to generate their preferred meta-ranking for computer science programs in the United States and Canada. In CS Open Rankings, each university lists its current tenure-track Computer Science faculty and alumni who are tenure-track Computer Science faculty. Each faculty listing has a link to view/edit their profile on Drafty. The four ranking sources are (see Appendix B for full descriptions):

(1) reputation (U.S. News)
(2) faculty publications (counting papers published from top conferences)
(3) academic placement (placement rank)
(4) recognition (best paper awards)

3.3 Databaits API: Automatically Extracting Insights from Tabular Data

Our Databaits API (a public RESTful API) creates insights from a tabular dataset. Each Databait type (10 total) is a sentence template that can be filled in with values and/or computed statistics derived from any tabular dataset. Figure 3 shows how the Databaits API generates an insight using Drafty's Computer Science faculty dataset.

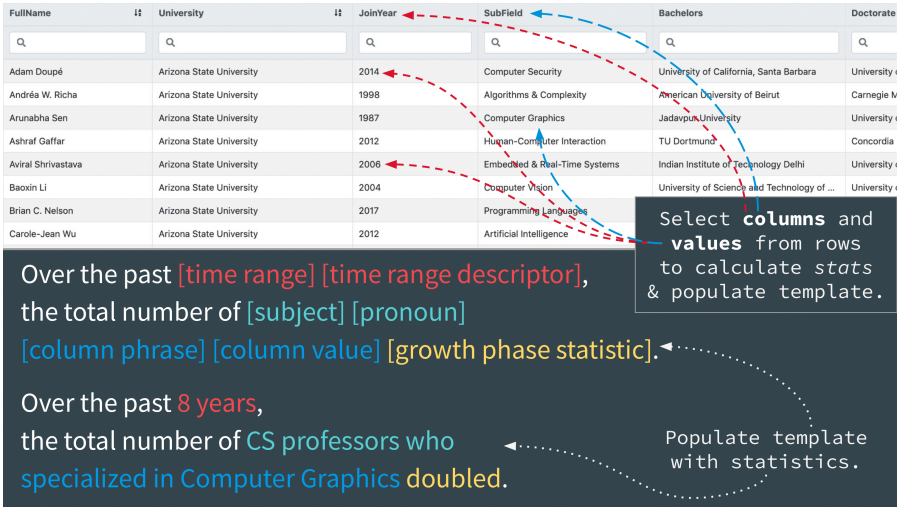


Fig. 3. Insights from a tabular dataset (i.e., CSV) are created by providing column names and column values as part of the input to the Databaits API. The Databaits API computes statistics using these inputs and merges the statistics with a randomly or intentionally chosen sentence template. We designed this method to work best with large tabular datasets (i.e., CSVs) with thousands of rows.

3.3.1 Databaits API: Design Considerations for Creating Insights from Tabular Data. We developed the Databaits API to generate insights from any tabular dataset in two main steps. First, it uses row values and column labels from a tabular dataset to compute aggregate statistics. Second, it merges these values, labels, and statistics into sentence templates. Each Databait type (10 total) has a unique statistic and a unique sentence template. [2] for how the Databaits API generates statistics and merges them with templates.

[2]View the pseudo-code in our GitHub repository: https://github.com/drafty-research

Databaits API: Creating the Insights Overview. First, the Databaits API takes in a list of column names and/or column values the user has interacted with. For example, on a video games sales dataset, a user clicks the "Publisher" column with the value "Nintendo"; our system uses these as inputs. Then, it calculates statistics about these column names and/or column values. These statistics might be the average value of a column or the rate of change of a column value over time.

The Databaits API can calculate 10 types, each combining a unique statistic with a corresponding template. For instance, one type describes how the frequency of a value changes over time (e.g., a rise in Sony-published games using a video game sales dataset). See Figure 8 for a full list of examples. For example, one template describes how a column value's frequency changes over time: 'Over the past 25 years, the number of Sony-published games increased 97 times.'

3.3.2 Databaits API: Pilot Study. In the age of misinformation, we believe it is responsible to conduct a preliminary study to pilot the insights created by the Databaits API on two different tabular datasets. Overall, people found the insights from the Databaits API interesting 77% of the time, believable 91% of the time, and easy to understand 93% of the time. Overall, people rated Databaits API's insights as interesting (77%), believable (91%), and easy to understand (93%). Notably, people wanted to see the insight's source data 30% of the time. Prior research shows relationships between someone's prior interest in a topic and increased engagement [25, 90], thus the results of this validation study gave us the confidence to integrate the Databaits API into Drafty to create the new Did You Know feature. Appendix section F presents the full validation study.

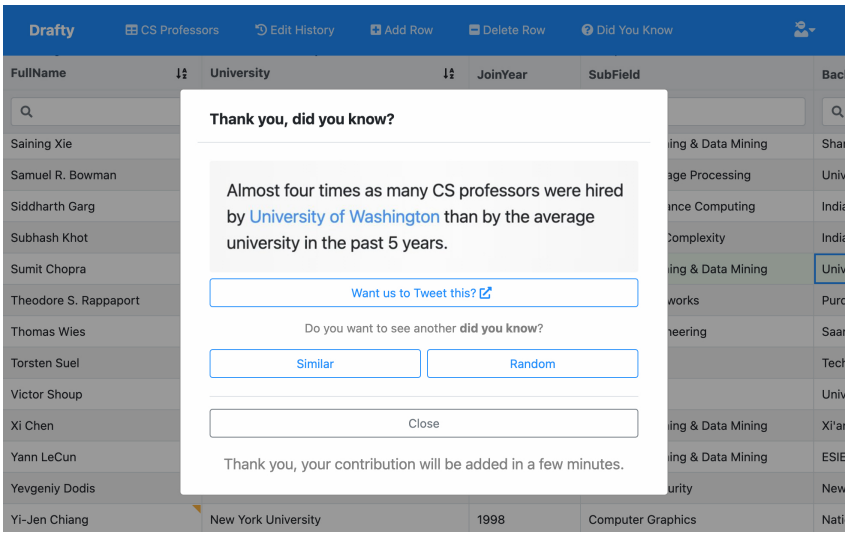


Fig. 4. **The Did You Know Modal** displayed post-edit or when a visitor creates an insight in Drafty. Actions: **tweet** to our Twitter, create a **similar/random** insight, click **highlighted text** to filter matching rows, or return to editing (X/ESC/"Go back to editing").

3.3.3 Did You Know: Providing Visitors Interactive Insights from the Databaits API within Drafty. We developed the Did You Know feature as an interactive modal in Drafty's spreadsheet interface to show visitors insights from the Databaits API. In real time, anyone can interact with insights derived from Drafty's evolving tabular data in the modal (see Figure 4). To build on prior research [26, 118], Drafty and its Twitter account present insights created using the Databaits API to current and potential visitors as a Did You Know.

In Drafty, anyone can create an insight from the Databaits API using one of the five methods:

(1) Right-click on a cell within the spreadsheet and select "See a Did You Know".
(2) Select "Did You Know?" in the top menu bar.
(3) Every time a user edits data (changing a cell's value, adding a new row, or deleting a row), the system generates a random Did You Know based on that row's information.
(4) In the Did You Know Modal, a user chooses to see another "Similar" or "Random" insight.

See Appendix section C for more details on how each method creates an automatic insight.

4 Methods: How to Attract and Engage Everyday Visitors in the Wild on Drafty

Our research combines descriptive and comparative analyses of anonymous visitor behaviors in Drafty. Our procedures and analysis follow a naturalistic field study design [72]. Three new recruitment methods were introduced in 2022 (our static, asking, and dynamic sources). Our comparisons before and after the introduction of these new recruitment methods highlight possible shifts in visitor behaviors. However, we note that this study cannot be interpreted as causal inference, much like prior work on Wikipedia-style editing behaviors on Stack Overflow [74].

Our visitor populations are not fixed across years (some may overlap, others may not); some may visit the system only once, without our knowledge, because of our system's anonymous user tracking; and external temporal factors may also influence their visits. We aim to understand our visitors' editing behaviors as they happened in our real-world peer production system. Our approach emphasizes ecological validity over strict causal effects.

With our local Institutional Review Board (IRB), we reviewed our systems, methods, metrics, and proposed analysis multiple times. Their decision was based on Drafty not recording visitors' personally identifiable information. Although our efforts do not fall within the categories of IRB exempt, expedited, or full review, we implemented data ethics and privacy measures in Drafty. Visitors to Drafty can request deletion of their previous interaction data and edits with a single button click. We note that, since Drafty uses only browser sessions to track visitors, this deletion applies only to data and edits made on the user's current browser. Per our IRB's recommendations, Drafty informs every first-time visitor that:

> *Drafty is a user-centric platform to edit and maintain tabular data. All interactions are captured and used anonymously for studies. Cookies are used to track per-browser usage. To edit data, double-click a cell.*

4.1 Methods RQ1: Attract (Source) Visitors to Drafty: Organic, Static, Asking, and Dynamic

Visitors to Drafty came from four sources: organic, static, asking, and dynamic. Drafty's back-end web services that track visitors from specific URLs per browser do not track their IP address. Most links are in publicly available websites, such as CS Open Rankings, Twitter, or Reddit.

Organic: Everyday Visitors Discovering and Visiting Drafty. "Organic" visitors can freely visit the Drafty web app through various means, such as search engines, typing the URL in the browser, or other people sharing a link to Drafty's public web app. We developed our PolyMu client-side JavaScript library to make Drafty's thousands of rows of data easier to search index by search engines (see Appendix A). Prior research on publicly available tabular datasets shows that they can rely on everyday visitors' natural usage (i.e, via URLs or search engines) as contributors [61, 83, 142].

Static: Computer Science Open Data, a Static Website with Analysis from Drafty's CS Professor Profiles, and Posting on Forums about Drafty. The first static sourcing method is Computer Science Open Data[3]; a web page with four analyses of Drafty's Computer Science professor dataset created in early 2022. Creating these websites and analysis takes considerable effort over

[3] https://jeffhuang.com/computer-science-open-data/

multiple months. We treat both as static because the content is manually authored and does not update from the dataset in real time; both primarily serve as inbound links to the spreadsheet.

(1) CS Faculty Composition and Hiring Trends
(2) Bias in Computer Science Rankings
(3) Who Wins CS Best Paper Awards?
(4) Verified Computer Science Ph.D. Stipends

Its static web-based articles can answer questions like "what percentage of professors get their bachelor's and PhD from the same institution?" or "what computer science areas have been growing?". These mirror questions from Drafty users that we have observed over time.

The second static sourcing method are posts about Drafty's CS professor dataset on public forums like Reddit. This visitor-recruitment method is similar to that used in many prior studies [143]. An example title used on Reddit was "Some Interesting Insights on CS Depts and Faculty in the US/Canada". See an example Reddit post in Appendix G.

Asking: Providing a Personal Touch by Emailing and Tweeting People to Visit, Review, and Contribute to Drafty's CS Professor Dataset. Inspired by prior work showing that direct requests and friendsourcing can elicit high-quality contributions [13, 118], we recruited contributors through asking: targeted outreach by a member of the CS community. This approach draws on social capital and network-tie accounts of participation [64, 126], and on social influence mechanisms whereby a clear call to action can increase participation [13].

First, using their personal Twitter account to contact people who recently announced they had accepted a tenure-track faculty offer in a Computer Science department included in Drafty, for example:

> "Congrats <first_name>! We have been building a public dataset of computer science professors to help students to find potential advisors. You should add/update yourself as a prof! :)"

We publicly replied to 25 people on Twitter over the Spring and Summer of 2022. Tweeting requires sustained effort. These tweets serve as a call to action for a community member to contribute. Public tweets directly link to Drafty.

We also emailed department chairs and directors of the Computer Science departments listed in Drafty. The emails contain links to Drafty and include a call to action for a department to review its faculty and alumni data. Since we emailed multiple people per department, anyone could share the email or choose to contribute. Please see Appendix section H for the template.

Dynamic: CS Open Rankings and Tweets with Insights from the Databaits API. Dynamic visitors can arrive at Drafty by either selecting links on CS Open Rankings or tweets from Drafty's Twitter account. Information and insights presented in CS Open Rankings and Twitter are dynamically created or regenerated from Drafty's tabular dataset. They evolve as Drafty's data changes. The idea is that people would visit this public information and then select a link to view and possibly edit the source data on Drafty.

CS Open Rankings provides a link back to Drafty for each professor or alumnus listed at a university. Drafty will automatically search its spreadsheet to display that professor's complete academic profile to the user.

Dynamic sourced visitors can visit Drafty from a tweet on its public Twitter; see Figure 5. These tweets are generated: 1) once daily by Drafty itself. 2) Visitors within the Did You Know Modal can have Drafty tweet the insight to its Twitter account. Each tweet uses a consistent template that supports all Databait types. See Figure 5 for examples. These templates avoid explicit imperative calls to action, consistent with Savage et al.'s recommendations [118]. The goal is to be transparent and communicate the data source that generated the automatic insight to potential visitors.

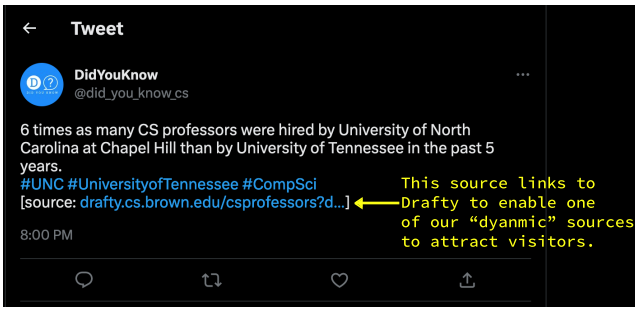


Fig. 5. Drafty's public Twitter account (@did_you_know_cs). Each tweet contains an insight created from the Databaits API. People can select a link per tweet to anonymously visit Drafty's CS Professor tabular dataset.

4.2 Methods

RQ2: Engage Visitors within Drafty using the Databaits API and Did You Know

To encourage Drafty's visitors to continuously make contributions to its evolving data and interact with the system, Drafty displays an insight to users via the Did You Know Modal interface (see Figure 4) in two ways. First, visitors may freely create an insight within Drafty at any time. Second, they can see an insight post-edit as part of a formal *"thank you"* and confirmation message. This occurs after they edit a cell or add or delete a row. Visitors can freely generate, view, and interact with as many insights as they want. After seeing the Did You Know Modal, visitors could perform one of six interactions (these interactions are included as part of our engagement metric) within the modal:

- Tweet Insight
- Create a Similar Insight
- Create a Random Insight
- Leave Drafty
- Close Modal
- Select Value in Insight to Search Drafty

In the modal, they can choose to create an insight that is either **similar** to or **random** relative to the current insight. A **similar** insight uses the same input values as the current insight (ex., university or research area). A **random** insight uses different inputs and randomly selects a different Databait type. Notably, the modal always presents the button for a similar insight first, thereby influencing visitors to select it [19]. Prior research on recommender systems shows that users prefer similar recommendations, insights, or examples [42, 49, 155]. Prior research also shows gratitude can increase unpaid contributions [94, 150]. These methods enable analysis to determine whether users' next actions benefit the data's health. For example, do users make accurate contributions, remain engaged, or create new insights?

5 Results: Naturalistic Field Study

Our approach and methods follow recommendations from Human-Computer Interaction researchers conducting field studies in the wild [72, 104]. On March 15th, 2022, we added the Did You Know feature to Drafty. Thus, our main analysis uses data collected over 196 days from March 15th, 2022 to September 28th, 2022. Notably, our data collection period *pre-dates widespread LLM usage*, as ChatGPT was released in November 2022 [69].

5.1 Metrics and Analysis Overview

Our analysis covers how our sources attract visitors (organic, static, dynamic, asking) and how visitors interact with **Did You Know** in Drafty. We compute the metrics below per visit. Our server

uses cookies to anonymously identify browsers; a visit is a unique browser with interactions no more than 20 minutes apart. Below are the metrics we use throughout the results.

- *Edits:* The number of contributed data points (i.e., edits per cell) per visit.
- *Conversion Rate:* Did the visit make an edit?
- *Accuracy:* The number of contributed data points marked as correct over the total number of contributions checked per visit. For example, a new row contributes six data points, while editing a single cell is one data point.
- *Engagement (Number of Interactions):* Within a visit, the total number of recorded interactions within Drafty.
- *Engagement (Session Length):* Per visit, the seconds between the first and last recorded interaction with Drafty.

Throughout this paper, we discuss edits that modify the tabular dataset, conversion rate, and accuracy separately from engagement with the system. For RQ1, we measure engagement using the number of recorded interactions and session length per visit. For RQ2, we also examine the types of interactions visitors take within Drafty, including edits and other interactions that do not directly modify the tabular dataset. Please see the appendix Table 7 for a list of interactions that Drafty anonymously records.

To answer RQ1, our analysis splits visits into groups based on our independent variables (organic, static, dynamic, or asking) and other anonymous visitor behaviors we can track:

(1) Visits from the four types of "sources" (organic, static, asking, and dynamic). (see Section 5.3.1)
(2) First-time visits versus return visits. (see Section 5.3.2)
(3) Visits from users who only made accurate edits versus those who did not. (see Appendix 5.3.3)

To answer RQ2, our analysis in sections 5.4.1 and 5.4.2 examines visits that freely create at least one insight in the Did You Know Modal. Section 5.4.2 examines if visits create new, similar, or random insights in the modal. This will answer Wallace et al.'s call for future research to study user preferences for random recommendations from a cohesive data source [140].

When comparing conditions for RQ1 and RQ2, we interpret effects primarily from within-2022 data, and we verify stability using month-stratified conversion rates (see Section 5.2.2). We also provide descriptive trends of visitor behaviors (see section 5.5). Our visitor populations are not fixed from year to year, and unknown external temporal factors may influence their behavior; thus, these trends contextualize Drafty's results and cannot be interpreted as causal effects.

5.1.1 Statistical Analysis Overview. Our analysis uses the D'Agostino-Pearson test when the sample size exceeds 5,000 to assess whether each group's data are normally distributed. Otherwise, we use the Shapiro-Wilk test. For post hoc pairwise comparisons following a chi-square test or Kruskal-Wallis H-test, we control the false discovery rate using the Benjamini-Hochberg procedure [134]. For all other families of multiple comparisons, we apply Holm–Bonferroni correction [1, 123]. Our data contains user interaction logs. Prior research on online user behavior consistently shows non-normal distributions with long tails for various metrics [7, 103]. Our data follows these familiar patterns. Hence, our accuracy analysis reports the median and interquartile range in Figure 6, along with other standard metrics.

5.2 Data Overview

In this subsection, we confirm the consistency and robustness of our data. While the findings here are not themselves impactful, we use them to confirm that our results are not unduly affected by the outlier events we can account for.

All users included in this analysis made at least one interaction event within Drafty (i.e., click, search, edit, etc.). This inclusion criterion ensures users in the analysis are not bots. After applying

these inclusion criteria, our dataset from March to September 2022 comprises 57,700 visits, 290,416 interactions, and 3,438 edits.

To compute accuracy per visit, we used stratified sampling to select edits to manually label as correct or incorrect. We used faculty web pages, department web pages, CVs, and LinkedIn profiles to verify that the edit was correct at the time it was made. This labeling strategy mirrors prior related papers [106, 141, 142]. We reviewed and labeled 1,709 edits from 227 visits for 2022. The average accuracy per visit in 2022 was 95%. We only used visits with at least one edit, which we checked by hand, to compute average accuracy per visit. Figure 6 presents an overview of the average accuracy per visit, with visits grouped by our independent variables.

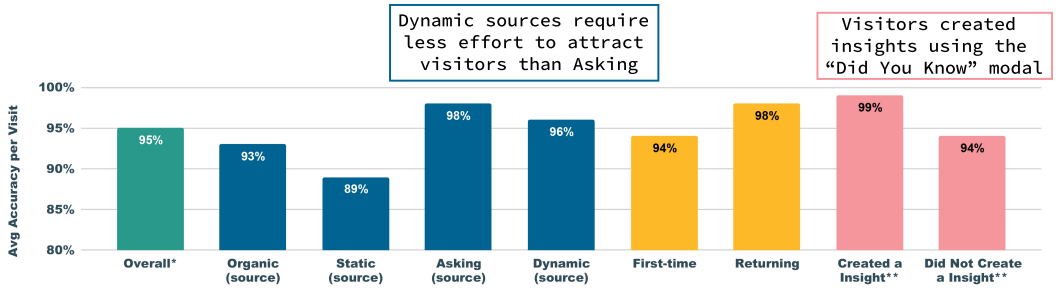


Fig. 6. This Figure shows the average accuracy per visit across different groups. Each color grouping maps to a group we compared in Results. The Median and IQR average accuracy per visit was 100% and (Q1 100%, Q3 100%) for each group. The majority of visits were 100% accurate. This pattern aligns with the system's long-term popularity and visitors' motivation to contribute accurate information for the benefit of others [90]. *Is the average for visits in our data from 2022. **Comparing visits that did and did not freely create an insight in Drafty, a Mann-Whitney U test shows that the increase in average accuracy per visit is statistically significant.

5.2.1 In Drafty Accuracy Comes from the Community, not a Few Power Contributors. Public data systems like ours often rely on a small group of users to maintain data quality [59]. For instance, research on Wikipedia shows that a few power contributors are responsible for most accurate contributions to its unstructured data [47, 70, 74, 113]. In contrast, in our data, accuracy appears to be distributed across many contributors and does not correlate with their total number of edits (r = 0.07), total number of visits (r = 0.06), total number of interactions (r = 0.09), or average session length (r = 0.11). This mirrors prior results on peer production efforts on datasets covering a single cohesive topic [142].

5.2.2 Verifying Data Consistency: Robustness Checks for Month-to-Month Variations. We apply robustness checks to the two main conversion effects in RQ1–RQ2. To ensure results are not driven by a single unusual month, we recompute the conversion rate (a visit made ≥1 edit) per month from April to September 2022. These checks do not establish causality; they only assess whether effect directions are stable over time.

For RQ1, we use organic as a baseline. *Asking* has a higher conversion rate in April-August (5/6 months), but less than organic in September, coinciding with the start of the Fall semester and a decrease in new faculty hires. *Static* has a lower conversion rate than organic in April-July, but slightly higher in August and September. *Dynamic* is lower than organic in April-June, similar in August, and higher in July and September. We observed anecdotally (from Reddit and email conversations) that people use our dynamic sources during admissions/application seasons to select universities to attend or apply to. In summary, relative to organic, asking's monthly change in conversion rate ranged from +1.2 to +9.0 percentage points, static (-4.3 to +0.5 points), and dynamic (-6.9 to +3.1 percentage points). This month-stratified pattern aligns with one of our main takeaways: asking,

while requiring more manual effort, is a relatively consistent high-conversion sourcing method in our data, mirroring prior research [13].

For RQ2, we check the conversion rate across months for visits that created an insight. The effect is directionally consistent in every month we observe: visitors who *freely created* a Did You Know insight have higher conversion in all six months (April-September). We observed that this difference tends to be higher in June-August and smaller in September. Thus mirroring a seasonal shift in visitor behaviors as the semester begins. Future data curators may wish to observe seasonal shifts in their users over time.

5.3 Results RQ1: Attract (Source) Visitors to Drafty: Organic, Static, Asking, and Dynamic

Our analysis first compares visitor behaviors by sources to attract visitors (i.e., recruitment method): organic, static, asking, and dynamic. Then, we motivate separate analyses, splitting our data on confounds from similar studies [142].

Source	Visits	Edits	Visits with ≥ 1 Edit	Conv. Rate	All Visits Edits per Visit	Visits ≥ 1 Edit Edits per Visit	Accuracy per Visit	Min. Acc.
Organic	3,408	1,509	198	5.8%	0.44	7.62	94.8%	0%
Static	1,785	278	84	4.7%	0.16	3.31	88.0%	0%
Asking	387	429	51	13.2%	1.11	8.41	99.6%	86%
Dynamic	1,402	1,220	72	5.1%	0.87	16.94	99.1%	80%

Table 1. In our data, visits from asking sources had the highest observed conversion rate (Conv. Rate = made ≥ 1 edit). While this mirrors prior research [13], manually contacting individuals to contribute is time-intensive and unsustainable for data curators. Visitors recruited through **dynamic or asking** sources were over 99% accurate, with minimum accuracy per visit no lower than 80%. Our static sources showed fewer edits and lower average accuracy.

5.3.1 The Highest Observed Accuracy per Visit was Among Asking and Dynamic Sources. How do we source or attract new visitors, and hopefully keep them editing and interacting with the system? Previous research shows that friendsourcing [13, 51] and personalized interventions can effectively motivate contributions to peer production systems like Wikipedia by leveraging social relationships and targeted incentives [91, 118]. However, research shows automated methods to nudge users to contribute work in the short term, but does not elicit sustained long-term contributions [118].

Visitor Source Did Not Impact Edit Accuracy. While accuracy per source varies (see Table 1), a Kruskal-Wallis H-test shows they are not statistically significant. While visits sourced from asking were the most accurate (mirroring prior research on friendsourcing [13]), they were similar to dynamic, the 2nd highest accuracy. Notably, the static sourcing method achieved the lowest accuracy and the fewest edits among the four sourcing methods.

While the Asking Source was Associated with Higher Conversation Rates, Trends show Editors from the Dynamic Sources Made More Edits per Visit. Table 1 shows the conversion rate and average edits per source (organic, static, asking, and dynamic). A chi-squared test shows the conversion rates were significantly different, $X^2(3, N = 6982) = 43.6$, $p < 0.001$, Cramer's $V = 0.08$. See Table 3 for the post-hoc comparisons. These results are consistent with prior findings that directly asking community members can be a high-conversion but labor-intensive strategy [13, 32]. A Kruskal-Wallis H-Test shows that the average edits per visit between sources were significantly different, $X^2(3, N = 6982) = 45.5$, $p < 0.001$, $\varepsilon^2 = 0.006$. See Table 3 for the post-hoc comparisons. However, to provide additional context, because many visits have no edits, we also analyze edits per visit only for visits with ≥ 1 edit. A Kruskal-Wallis H test indicates that sources differ in edits per visit,

$X^2(3, N = 405) = 47.0$, $p < 0.0001$, $\varepsilon^2 = 0.11$. Notably, among editors from our dynamic sources had 2.22 more edit per visit than organic, and 5.12 more than asking; see Table 2 for the post-hoc comparisons.

Visitor Source Affected Session Length and Interactions per Visit. In crowd-powered and peer production systems, sustained visitor participation and growing a community are essential to maintaining accurate data [8, 17, 18]. Table 4 shows that while visits from the dynamic source had the lowest average session length and interactions per visit, they were equally engaged when comparing these values per second. Given the high conversion rates and accuracy from the dynamic and asking sources, one possible interpretation is that some of these visits likely arrived at Drafty, quickly edited data, and left rather than lurk and explore.

When comparing average interactions per visit across sources, a Kruskal-Wallis H-test indicates they were significantly different, $X^2(3, N = 6982) = 671.9$, $p < 0.001$, $\varepsilon^2 = 0.10$. Also, for the average session length per visit across sources, a Kruskal-Wallis H-Test shows they were significantly different, $X^2(3, N = 6982) = 372.8$, $p < 0.001$, $\varepsilon^2 = 0.05$. Table 4 reports effect sizes and shows that these differences were significant across different sources. When splitting visits by source, Pearson's Correlation shows no effect of accuracy on average session length or average interactions per visit. Although session lengths varied across sources, we did not observe a relationship between session length and edit accuracy, mirroring prior work on a similar dataset [142].

5.3.2 First-Time Visits versus Return Visits. We compare first-time and return visits. A return visit is someone who revisits Drafty using the same browser. We study these two groups because related peer production research on Wikipedia and WikiData shows that returning visitors are more likely to contribute, and motivating first-time visitors to contribute is challenging [6]. Notably, Drafty does not require special onboarding like other peer production efforts [6, 96].

First Time Visitors and Returning Visitors are Similarly Accurate. While there is a 4% increase in average accuracy per visit when comparing first-time visits (M = 0.94%, SD = 0.24, N = 136) and return visits (M = 0.98%, SD = 0.12, N = 91), a Mann-Whitney U test shows this increase is not statistically significant. This result replicates previous research on a similar dataset of CS professors, showing that first-time and return visits make accurate edits [142].

First Time Visitors and Returning Visitors Contribute Edits at Similar Rates. For first-time visits, 259/4,562 made ≥ 1 edit (5.68%); for return visits, 147/2,421 did so (6.07%). A χ^2 test shows no significant difference in conversion rate. Return visits made more edits per visit (M = 0.72, SD = 6.73, N = 2,421) than first-time visits (M = 0.37, SD = 3.58, N = 4,562), but this difference was also not significant by Mann-Whitney U, despite prior work suggesting return visitors contribute more [143].

Returning Visitors Have Longer Sessions, but Similar Numbers of Interactions as First Time Visitors. The average interactions per visit were similar (only around three interactions less)

	(Dynamic vs. Source) *Editors Avg Edits/Visit* Post-hoc*				*(Asking vs. Source)* *Editors Avg Edits/Visit* Post-hoc*			
Source	**Increase**	U	p	r	**Increase**	U	p	r
Organic	2.22×	4486	<0.001	0.28	1.10×	3437	<0.001	0.22
Static	5.12×	4582	<0.001	0.44	2.54×	3214	<0.001	0.42
Asking	2.01×	2166	$=0.078$	0.15	–	–	–	–
Dynamic	–	–	–	–	0.5×	2166	$=0.078$	0.15

Table 2. Post-hoc comparisons of ***average edits per visit** for editors only (visits with ≥ 1 edit). Standard deviations for editors only: organic (SD = 17.84), static (SD = 3.82), asking (SD = 11.24), and dynamic (SD = 29.52). See Table 1 for N and M for visits ≥ 1 edit. Please note, increases described here are increases of dynamic/asking, respectively, over the given source (leftmost column). *U is for the Mann–Whitney U test.

	(Asking vs. Source) Conversion Rate Post-hoc					*(Asking vs. Source) Edits per Visit Post-hoc*			
Source	**Increase**	**df**	χ^2	p	V	**Increase**	U	p	r
Organic	2.3×	1	29.59	<0.001	0.09	2.5×	609245	<0.001	0.04
Static	2.8×	1	37.73	<0.001	0.13	7.1×	375733	<0.001	0.06
Dynamic	2.6×	1	29.40	<0.001	0.13	+21.5%	249798	<0.001	0.06

Table 3. Post-hoc comparisons for **asking** versus other sources. Left: conversion rate (made ≥ 1 edit) using chi-squared tests and Cramer's *V*. Right: edits per visit using Mann-Whitney U tests. Asking summary: M = 1.11, SD = 5, N = 387. χ^2 tests use the combined sample size of Asking + the comparison source (Organic+Asking $N = 3{,}795$; Static+Asking $N = 2{,}172$; Dynamic+Asking $N = 1{,}789$), whereas the edits-per-visit tests report per-source N (Organic $N = 3{,}408$; Static $N = 1{,}785$; Dynamic $N = 1{,}402$). *U is for the Mann–Whitney U test.

comparing first-time (M = 33, SD = 65, N = 4562) to return visits (M = 36, SD = 74, N = 2421). A Mann-Whitney U test shows this 9% decrease is not significant. However, the average seconds per visit for first-time visits (M = 214, SD = 503, N = 4562) versus return visits (M = 575, SD = 1213, N = 2421) was 6 minutes and 1 second shorter. A Mann-Whitney U test shows this 63% decrease in session length is statistically significant, $U = 5305198$, $p = 0.007$, but the standardized effect size is minimal $r = 0.03$. Return visits' standard deviation for session length is more than double that of first-time visits, indicating a greater variation in session lengths.

5.3.3 Behaviors of Highly Accurate Visitors versus Lurkers. This section analyses visitor behaviors of accurate visits (i.e., visits that made only accurate edits) and visits with no edits (i.e., lurkers). In peer production, understanding the transition from lurker to active contributor is complex: some lurkers often remain passive observers [65, 146].

Common Mistakes of Highly Accurate Visits. Only 8% (19 total visits) made at least one error, while 92% of visits made only accurate edits. This sample size, which made at least one error, is too small to conduct hypothesis testing. Thus, we will examine their behaviors, starting with common editing errors:

(1) Accidentally selecting the first value from the modal edit drop-down. There were cases in which visitors made this error but immediately corrected it. These mistakes were not included in the analysis.
(2) Sharing information that was not yet public or true (i.e., someone leaving their university).
(3) Entering a research area that was a professor's secondary research area, not their primary.

	Average Session Length per Visit					*Average Interactions per Visit*				
Source	**Avg.**	**Org.**	**Stat.**	**Ask.**	**Dyn.**	**Avg.**	**Org.**	**Stat.**	**Ask.**	**Dyn.**
Organic	460 sec	-	1.9x ↑	2.0x ↑	2.4x ↑	35.3	-	8.1% ↓	11.4% ↓	48.5% ↑
Static	247 sec	46.4% ↓	-	*4.4% ↓	21.4% ↑	39.8	12.8% ↑	-	3.7% ↑	1.7x ↑
Asking	236 sec	48.7% ↓	*4.6% ↑	-	17.8% ↑	38.4	8.8% ↑	3.5% ↓	-	1.6x ↑
Dynamic	194 sec	57.8% ↓	21.4% ↓	17.8% ↓	-	23.8	32.7% ↓	40.3% ↓	38.1% ↓	-

Table 4. This table shows the effect sizes for the **average session length in seconds** and **average interactions** per visit across sources. The reported effect sizes compare the left-most column with the source from the top row. Mann-Whitney U tests reveal all reported differences *(* except the average session length for asking versus static)* are statistically significant, $p < 0.001$ in all comparisons. See Table 8 in the Appendix I for the standardized effect sizes. They range from $r = 0.5$ to $r = 0.46$. * Table abbreviations are Org. = Organic, Stat. = Static, Ask. = Asking, Dyn. = Dynamic

(4) The most common error was Bachelor's degree. There was no obvious pattern to errors. The overall accuracy for all visitors' edits to Bachelor's degrees in 2022 was 96%.

Edits per visit and Accuracy are not Correlated Pearson's Correlation shows no effect between accuracy and the number of edits per visit ($r = 0.08$, $p = 0.25$). This result replicates previous research showing no relationship between the number of edits and accuracy [142]. The lack of a relationship between accuracy and the number of edits is inconsistent with results from Wikipedia, where a few users contribute the bulk of accurate edits in its unstructured data [47, 70]. While Drafty has grown from thousands to millions of visitors, it does not attract as many visitors as Wikipedia/WikiData. These results might shift if Drafty's users vandalized data (i.e., making intentional frequent errors).

Highly Accurate Visits were More Engaged than Lurkers. The average number of interactions per visit for accurate visits (M = 63, SD = 72, N = 208) compared to visits with no edits (M = 32, SD = 65, N = 6577) was 1 minute and 48 seconds longer. A Mann-Whitney U test shows this 47% increase in session length is statistically significant, $U = 1009201$, $p < 0.001$, $r = 0.14$. The average session length per visit for accurate visits (M = 707, SD = 72, N = 208) compared to visits with no edits (M = 324, SD = 65, N = 6577) was 6 minutes and 23 seconds longer. A Mann-Whitney U test shows that the average session length for accurate visits is 2.2 times longer than visits with no edits, is statistically significant, $U = 1040255$, $p < 0.001$, $r = 0.16$.

5.4 Results RQ2: Engage Visitors within Drafty using the Databaits API and Did You Know

5.4.1 Visitors who Created Insights using Did You Know versus those Who Did Not. Prior research shows contributors have complex motivations beyond simple altruism. Creating insights from data represents creating something new, while potentially creating reciprocity between contributors and the source data [4, 140]. Providing visitors with agency to generate insights affords freedom while also enabling them to create a small artifact or boundary object from Drafty's source data that appeals to their community [62].

Visits who Created Insights were More Accurate. The average accuracy for visits that created an insight (M = 99%, SD = 0.12, N = 73) versus visits that did not (M = 94%, SD = 0.23, N = 154) was 5 points higher. A Mann-Whitney U test shows this 5% increase in average accuracy per visit is statistically significant, $U = 6091$, $p = 0.03$, $r = 0.07$. One in three visitors who edited data on Drafty also freely chose to create insights in the Did You Know Modal.

Visits that Created an Insight were More Likely to Contribute. The conversion rate for visits that created an insight, is 27.6% (105/380 visitors made at least one edit). Visits that did not create an insight, their conversion rate is 4.6% (301/6,603 visitors made at least one edit). Notably, chi-squared test shows this 6× increase was statistically significant, $X^2(1, N = 6983) = 345.1$, $p < 0.001$, $\varepsilon^2 = 0.22$. Among all our results, visits that created insights had the highest conversion rate. On average, visits that created an insight (M = 2.81, SD = 11.66, N = 380) made 7.8 more edits per visit than visits that did not (M = 0.36, SD = 4.17, N = 6603). A Mann-Whitney U test shows this increase is statistically significant, $U = 1546049$, $p < 0.001$, $r = 0.09$.

Visits who Created Insights made more Interactions. Spreadsheets typically offer limited support for users to engage, interact, with something personally relevant [112]. Our Did You Know helps address this by presenting insights from Drafty's data. Visits that created an insight had 25 more interactions (M = 58, SD = 100, N = 380) than those that did not (M = 33, SD = 65, N = 6603). A Mann-Whitney U test indicates that this 1.8 times increase is statistically significant, $U = 1619791$, $p < 0.001$, $r = 0.11$. The average session length per visit for visits that generated an insight (M = 545, SD = 852, N = 380) was approximately 3 minutes and 38 seconds longer than for visits that did not (M = 328, SD = 838, N = 6603). A Mann-Whitney U test shows this 2.2 times increase in duration is statistically significant, $U = 1750462$, $p < 0.001$, $r = 0.16$.

5.4.2 After Creating or Seeing an Insight in the Did You Know Modal: What were Visitors' Next Actions? We do not consider creating an insight in the Did You Know Modal to be an edit to the tabular dataset. Instead, we examine visitors' next actions within the modal as subsequent actions after creating or seeing insights, including both actions related to maintaining the dataset and other interactions with the feature. Visitors saw an insight 1,542 times by freely creating one, and post-edit 902 times. After a visitor sees the Did You Know Modal, there is a 46% chance they create another insight within the modal. Theoretically, this number cannot reach 100%, since visitors will eventually have to close the modal. Likewise, if all visitors generated one additional insight and then closed the modal, then this percentage would be 50%. Thus, we believe that a value of 46% indicates meaningful use of the feature. This analysis and approach mirrors prior recommendations from peer production systems in CSCW [140].

Drafty's Visitors Prefer Randomized Insights. Visitors were 6 times more likely to create a random insight than a similar insight. A one proportion z-test shows that visitors creating a random (vs. similar) insight 86% of the time is statistically significant ($z = 23.86$, $p < 0.001$), with a large effect size (Cohen's $h = 0.79$). This shows Drafty's visitors preferred to generate randomized insights from the source tabular dataset, demonstrating reciprocity between visitors and the dataset [23, 85] and contradicts prior research in recommendation systems [42].

Visitors who Freely Create Insights are More Likely to Contribute. We analyze visitors' next actions using two-tailed McNemar's tests with an odds ratios (ψ) [110], like prior CSCW research [140]. We apply continuity correction when testing paired proportions, and separately examine actions *within* the Did You Know Modal and actions *after* the modal is closed. These analyses compare visitors who freely choose to *create* an insight with visitors who *see* an insight post-edit.

In both conditions, visitors disproportionately chose to create a randomized (vs. similar) insight as their next action (see Table 5), with large standardized effects. Visitors who **freely create an insight** were 6.6× more likely to create a randomized (vs. similar) insight ($z = 23.13$, $p < 0.001$, $h = 0.79$). Likewise, visitors who **see an insight post-edit** were 6.6× more likely to create a randomized (vs. similar) insight ($z = 7.56$, $p < 0.001$, $h = 0.83$). Both comparisons have a large effect size (Cohen's h)

Within the modal, visitors who created an insight had a 66.2% chance of creating another insight as their next action, which is 5.6× higher than when visitors saw an insight post-edit. A McNemar test shows this increase was statistically significant ($\chi^2(1) = 106$, $p < 0.001$, $\psi = 0.07$). Conversely, when visitors saw an insight post-edit, there was a 78.8% chance they would close the Did You Know Modal as their next action, which is 3.8× higher than for visitors who created an insight; this difference was statistically significant ($\chi^2(1) = 711$, $p < 0.001$, $\psi = 14.05$).

After a visitor closes the modal, their next actions are one of three outcomes listed in Table 6. While visitors were similarly likely to leave Drafty after freely creating an insight versus being shown one

		Created an Insight					
Freely chose to create the Insight	**Tweet Insight**	**Total***	*Create Similar*	*Create Random*	**Leave Drafty**	**Close Modal**	**Search Drafty**
Yes	1.0%	66.2%	9.7%	56.5%	11.3%	20.9%	-
No	1.7%	11.8%	1.6%	10.5%	7.8%	78.8%	-
Likelihood	0.6	5.6	6.2	5.5	1.5	0.3	-

Table 5. Visitors who freely chose to create an insight were most likely to create another insight as their next action. The most common action for visitors who saw an insight post-edit was to close the modal. McNemar tests show all relationships were statistically significant $p < 0.001$. No visitors searched the dataset using the modal, this feature may have been hard to discover. Likelihood compares next actions between visitors who voluntarily created an insight and those who encountered one after an edit.

post-edit, visitors who did not freely choose to create an insight were 2× more likely to not contribute within that same visit. A McNemar test shows this decrease was statistically significant ($\chi^2(1) = 182$, $p < 0.001$, $\psi = 2.21$).

	Interactions after closing the Did You Know Modal		
Freely chose to create a Insight	**Left Drafty**	**No Contribution**	**Made a Contribution**
Yes	11.4%	10.2%	78.3%
No	11.1%	20.2%	68.7%
Likelihood	1.0	0.5	1.1

Table 6. Visitors who did not freely choose to create an insight were twice as likely not to make a contribution to Drafty after the Did You Know Modal. Likelihood compares the next actions of visitors who freely chose to create an insight with those who did not (i.e., they saw the Did You Know Modal post-edit). A McNemar tests show all relationships were statistically significant $p < 0.001$.

5.5 Cross-Year Descriptive Trends for Contextualizing Findings

We also report a brief cross-year descriptive comparison (2021 vs. 2022) to contextualize our 2022 findings; we do not interpret cross-year differences as causal. Our primary research questions are based on 2022-only data. This section does not directly answer our research questions. Instead, it reports trends in visitor behavior from before we introduced our new systems and features, providing descriptive context for interpreting the 2022 analyses and informing future research. We compare the same date range in each year (March 15th through September 28th). We treat all cross-year differences as descriptive context only because 1) our visitors are anonymous and not fixed across years, and 2) because external conditions may differ per year. We do not collect IP addresses or other personally identifiable information, so we cannot use visitors' physical location or demographics. The system timeline in Figure 1 summarizes the major changes introduced in 2022. During the 2021 window, we made no updates to Drafty. In preparing the 2021 data, we followed the same procedures from section 5.1.1. Across the 2021 window, 2,536 total visits produced 86,950 interactions and 265 edits.

Accuracy per visit trended higher in 2022. The average accuracy per visit in 2022 (M = 95%, SD = 0.20, N = 227) compared to 2021 (M = 73%, SD = 0.42, N = 55) is around 22 points higher (30.5% relative difference). This difference occurs during the period when the 2022 updates were deployed (see Figure 1), but our analysis cannot attribute the change to any single factor. For additional context, prior longitudinal deployments of a similar dataset report accuracy *per edit*. In 2022, the average accuracy per edit (M = 98%) exceeds earlier reports from 2016–2017 (M = 76%) [143] and 2017–2019 (M = 89%) [142]. Those studies also found higher accuracy when nudging users to edit content aligned with their interests (2016–2017: M = 88%; 2016–2019: M = 95%) [142, 143].

Editing was more common and edits per visit trended higher in 2022. As descriptive longitudinal context, the conversion rate for visits was higher in 2022 (5.8%) than in 2021 (2.5%). Edits per visit were also higher in 2022 (M = 0.49, SD = 4.91, N = 6983) than in 2021 (M = 0.13, SD = 2.00, N = 2536), a difference of 0.36 edits per visit. We report these cross-year contrasts as context only and do not interpret them as feature-driven or causal effects because visitors are anonymous and external conditions may differ across years.

Session length trended higher in 2022. The average interactions per visit were similar in 2022 (M = 34, SD = 68, N = 6983) and 2021 (M = 34, SD = 93, N = 2536). However, the average seconds per visit in 2022 (M = 340, SD = 840, N = 6983) compared to 2021 (M = 232, SD = 580, N = 2536) was 1 minute and 48 seconds longer (47% relative difference). Overall, these descriptive trends provide context for our 2022 findings, while remaining consistent with our non-causal interpretation.

6 Key Takeaways

The Effectiveness of Dynamic and Organic Sources to Attract Visitors. Dynamic sources (e.g., CS Open Rankings, Twitter via the Databaits API) attracted visits who made accurate edits without direct prompts to friends (i.e., friendsourcing) [14]. While our asking (or friendsourcing) method did show an increase in conversion rates across months like prior work [118], it generated fewer edits than our dynamic or organic sourcing methods. Organic visitors contributed the most edits with comparable or better accuracy than in earlier studies [142, 143]. While our dynamic sourcing methods may offer a scalable approach to support continuous contributions [13], developers have to spend initial time creating them.

First-Time vs. Returning Visitors. First-time visitors made accurate contributions at rates similar to those of returning visitors. This diverges from findings in Wikipedia, where return users tend to be more accurate [47, 70], and new users benefit from structured gamification [95]. First-time visitors were also just as likely to contribute, suggesting that systems focusing on a dataset with a singular topic could aim to support a broad range of contributors, not just high-frequency power contributors [154]. These observations may also be a result of our system reducing the initial effort to contribute (i.e., no logins or vetting processes), compared to platforms (Wikipedia, WikiData, MovieLens, or Zooniverse) that require creating user accounts or performing sequential steps to contribute [6, 105, 154].

Insights and Engagement via Did You Know Modal. The Databaits API and Did You Know were deployed in Drafty in the Spring of 2022. This is in response to years of people creating static insights from Drafty's data. When Drafty's visitors create an insight in Drafty, they tend to be more likely to contribute without another prompt from the system. This observation builds on prior research suggesting that a system should not ask people to contribute [118, 143]. This further indicates that providing Did You Know as a feature offered another way for visitors to freely interact with Drafty, similar to prior work focusing on user agency to seek randomized insights [140].

Post Did You Know Automatic Insight Visitor Behavior. Visitors behaved similarly whether they generated an insight or were shown one as a post-edit thank-you. Many chose randomized insights rather than similar ones, showing initial results in one system (Drafty) that contradicts collaborative filtering approaches over datasets with more diverse content [67]. This aligns with prior calls to explore randomized example generation in public data systems [140]; and mirrors prior results on how gratitude messages after contributions may further support user contributions and engagement with a system [94, 150].

Visitor Accuracy and Engagement Trended Higher, Emerging Alongside the Introduction of New Features in Drafty. Our new additions were designed to grant users agency over their interactions [111]. We observed high levels of accuracy and engagement in 2022, without using demotivating tactics such as nudging [118, 142]. We observed accuracy levels that met or exceeded those reported in similar studies of unpaid contributors [142, 143]. Our Did You Know feature provides automated insights derived from user interactions with data, thereby building on Self-Determination Theory's recommendations to use personalized messages to motivate contributions [43].

7 Discussion

Our efforts designing Drafty support understanding of attracting and retaining highly accurate anonymous contributors. We address the challenge of sustaining participation identified [38, 142]. Visits from dynamic sources that generate automated insights were associated with higher edits per visit, accuracy, and engagement (RQ1). Also, visits where users freely generated their own insights using Drafty's Did You Know Modal were associated with higher engagement and accuracy (RQ2).

Conceptual Model for Did You Know and Engagement: We interpret our Did You Know results as an engagement loop that links (1) immediate value from the dataset (via insights), (2)

user agency to choose what to explore or create, and (3) a low-friction next action back into the spreadsheet (e.g., search/edit). Visits that *freely create* an insight show the strongest co-occurrence with contribution behaviors, consistent with agency as an important driver, whereas insights shown post-edit are less strongly associated with additional contributions.

7.1 Our Interpretations of Visitor Behaviors and Motivation per Source

Our discussion below contextualizes behavioral differences across sources based on our observations and conversations with the community, not to claim definitive visitor motivations or intent.

Organic visitors likely have the most varied motivations, arriving via search engines and links. We believe they may seek specific information (e.g., a professor or university) and may fix missing or outdated entries.

Static-source visitors come from specific CS sub-communities and contexts (e.g., graduate admissions forums, general CS subreddits, or one-time analyses on PhD stipends). They may visit to review the source data or locate specific departments or individuals mentioned in posts. Our static sources offer less reciprocity: edits do not update the source, thus potentially reducing the perceived payoff from making contributions.

Visitors from **Asking** sources likely arrive with a clear social prompt and personal relevance to the data (e.g., being newly hired). They may contribute to being helpful, to reciprocate, or to ensure that their profile or their new department's public representation is accurate. These motivations are consistent with friendsourcing and social obligation [13, 32]. This source may produce higher conversion rates because the visit begins with an explicit reason to contribute.

Dynamic sources may attract visitors through *decision-making* (e.g., comparing programs for applications/admissions or exploring subfields) and possibly curiosity influenced by automatic insights. Because these motivations could be explicitly derived from Drafty's source data, they could foster stronger reciprocity and feedback loops. For CS Open Rankings, edits can also change downstream representations (e.g., your department's rankings), thus creating real-world value for community members and their departments. This source might attract short task-oriented visits that: arrive, verify/edit, and leave. This aligns with our observation that dynamic-sourced visits tended to have long browsing sessions.

7.2 Designing Systems for a Single Community: Enabling Divergent Design Choices

Small, interest-aligned communities enable different design tradeoffs than large platforms and can sustain high-quality contributions [3]. Our observations are suggestive rather than comprehensive, and motivate future work on sustaining motivation and accuracy as contexts, users, and datasets evolve [31, 53, 54].

7.2.1 Recommendation Systems for Small Communities Should Enable Randomized Recommendations. Collaborative filtering has excelled in past research on recommendation systems because it leverages user data to personalize recommendations [66, 67]. In contrast, our visitors freely chose random insights (rather than similar ones) from Drafty's data. This preference may stem from the dataset and community's shared context: even random entries remain relevant because of our single dataset's narrow scope [140]. For example, moving from a University of Chicago row to one about the University of Washington still aligns with users' academic interests, unlike reading two unrelated Wikipedia topics. These findings offer a simple, effective alternative to traditional recommendation techniques for systems focused on small communities or singular topics like ours.

7.2.2 Common Themes in a Single Dataset can Foster Intrinsic Motivation. By centering on a singular dataset of computer science faculty, we built complementary systems like CS Open Rankings that generate community-relevant insights. This design choice might encourage contributions by enabling

users to directly support and improve resources meaningful to themselves or their peers [90, 91]. They also could possibly be motivated by social comparisons of rankings and insights [21]. Public datasets with a unifying theme can attract contributors who share common goals or interests [142]. For example, CS Open Rankings instantly reflects updates after a user edits Drafty, offering contributors an immediate, community-aligned reward. These intrinsic rewards—timely, personally relevant, and socially beneficial—can be more motivating than gamified incentives like badges or points [44, 57, 84]. Prior research supports this, showing that contributors are more engaged when their efforts visibly benefit a shared cause [6].

7.3 Future Research Directions and Recommendations for Interaction Design

We integrate our system-building perspective and history with several of Norman's design principles for Human-Computer Interfaces [101] and the future generative AI applications outlined by Weisz et al. [147] to discuss our system's features. We discuss the potential design implications for both existing applications and those integrating generative AI. We hope that future system designers can learn from our experiences and identify which recommendations and opinions will work for their systems and user bases.

7.3.1 Increase Agency with Publicly Accessible Features and Data. We designed our platform to connect and make visible the tabular dataset of Computer Science faculty that empowers each system: Drafty, CS Open Rankings, and the Databaits API. Drafty makes thousands of rows of raw tabular data fully visible and editable client-side, ensuring users have equal access and agency to explore and edit all data.

Our systems operate as anonymous collaborative environments that combine structured task models from crowdsourcing with the autonomy of peer production [8]. We believe this hybrid design supports accurate contributions while enabling individual freedom. We believe that the anonymity of edits might help reduce social loafing (where individuals exert less effort in groups) by shifting the focus from social dynamics to individual motivation and empowerment [24]. Our results on anonymous editing build on related social computing research adopting Wikipedia-style collaborative editing requiring user accounts [28, 74].

7.3.2 Increasing Reciprocity through Dynamic Sources with Insights from the Source Data. Norman emphasizes meaningful feedback as a core usability principle [101]. Our systems operationalize this through Cialdini's theory of reciprocity [23], which posits that users are more motivated to contribute when they receive something valuable in return. We applied Fogg's Behavioral Model [39] by designing simple, rewarding ways to contribute. For example, CS Open Rankings and Did You Know provide automatic insights from the shared dataset, thereby creating intrinsic value for contributors. We found that these feedback loops might help sustain visitor participation through these shared boundary objects [85]. In contrast to prior research on Wikipedia by Klein et al. [63], we show that increases in editors or visitors did not decrease accuracy or data quality for our specific dataset. This is likely because our single dataset avoids the horizontal organization and heterogeneous user-interest issues that Wikipedia editors face. For example, faculty themselves (or those who know them) might be motivated to contribute data to enhance their professional identity or ensure their department's reputation and ranking are accurate in CS Open Rankings. Compared with automated methods on Wikipedia for detecting quality issues [28], our results show that trends and longitudinal automated methods might help positively attract quality editors to interact with and contribute to our system. We believe future work should study how boundary objects that might influence personal and group reputation across datasets could affect visitors' motivations to contribute.

7.3.3 Design for the Generative Variability of Automatic Insights. The Did You Know Modal's automatic insights rely on deterministic, template-based sentence structures, similar to prior systems such as SuggestBot [26], Botivist [118], and others [68]. While effective, these static templates might become repetitive or be interpreted differently across cultures and grammatical structures. As large language models (LLMs) become more capable, future systems could experiment with generating more natural, diverse variations. LLMs could also help translate or simplify content for non-English speakers. However, the risks of hallucination remain [52]. We recommend future work to explore an agentic approach balancing human understanding with the factual accuracy of insights created from tabular data. Future work could also compare user-controlled insights with those generated by LLMs, thereby addressing the call for user agency during a generative process [147].

7.3.4 Future Directions Towards Drafty and WikiData Co-Creation. To extend our community of collaborators beyond Drafty to more widely used systems, future work can explore integrating Drafty with broader platforms such as Wikidata. Though WikiData hosts more users, its interface may be too complex for new contributors [98, 132]. Translating Drafty's tabular data into graph-based formats might bridge the gap between low-effort and expert workflows [88]. Simultaneous longitudinal studies of contributors across multiple datasets and both platforms could reveal design insights to support different skill levels and increase diversity, particularly through recent efforts to support multilingual and bilingual editing [56].

7.3.5 Design for Co-Creation between Humans and AI. Our systems provide users with access and control over contributions to the tabular dataset, per peer production principles [6, 8], participatory crowdsourcing [10, 11], and collaborative editing [74, 142]. As generative AI integrates into social computing and data systems [89], we argue that AI should augment, not replace, human contributions. What happens if people stop contributing to online information sources? LLMs are reshaping how communities create and maintain knowledge [69]. While they provide new affordances for collective intelligence, studies show they can also displace human effort. Empirical work highlights declines in participation after generative AI adoption [16] and warns that Wikipedia and similar efforts risk "death by AI" as volunteer labor erodes [136, 156].

Online communities are actively negotiating these risks. Reddit forums have introduced policies to regulate AI-generated content [77], while user reactions to AI companions show ambivalence and concern over authenticity [93]. Research demonstrates opportunities for human-AI collaboration: LLMs can scaffold workers' sensemaking [55], support culturally aware moderation [121], and aid content generation when paired with structured interaction design [37]. However, users express concerns about trust, over-reliance, and their agency despite reporting that AI helps them work faster [125, 130]. This is all while unpaid contributors prefer non-AI collaborative editing tools to paid crowdworkers [141].

Future CSCW systems and research could build on how we evaluated Databaits and Did You Know to help users opt-in to AI co-creation features to further study open-research questions in trust, safety, and responsibility in AI [69, 89]. Thus, we implore future researchers to consider data lineage. Can AI-created or modified content be watermarked?

8 Limitations and Future Work

We believe our longitudinal observations serve as a foundation for future controlled studies to build on our limitations to explore analytical methods such as propensity score matching with AI-influenced data maintenance efforts.

A more thorough qualitative study conducted from April to September 2022 would have enabled additional interpretations of our results at the time. This also helps reveal more details about visitors' individual motivations, for example, directly mapping their domain expertise or professional

affiliation with their edits. We chose not to conduct a qualitative study at that time because we did not want to introduce an additional confound that could interrupt visitors' interactions and flow within our system. This would make it difficult to answer RQ2 about visitor engagement.

The initial Databait sentence templates are in English. Future researchers should consider translating sentence templates into other languages [128, 129]. We conducted pilot studies to develop easy-to-understand sentence templates. Translating these sentence templates will require careful consideration and, in some cases, the development of new templates to account for differences in grammar and connotation across languages and neurdivergent reader profiles [50, 76, 100, 138].

While we validated Databaits API with multiple datasets, our main study focuses on one tabular dataset of Computer Science professors from the US and Canada. We believe that future work should extend to studying other large, evolving tabular datasets and international universities, and to considering the interplay between information understanding, insights, and tabular data [99, 145]. We focused solely on a single large dataset because we had an existing community of editors that had grown over nearly a decade of effort.

Confounding Variables and External Events. Like with many longitudinal studies in the wild, we cannot control for or predict every situation that might affect individual users' behaviors [82]. We believe that studies like ours are valuable, responding to the call for more research on users in their natural environments over time [5].

COVID-19: We acknowledge that our data was collected during COVID. Based on our interactions with the community of our dataset, our visitors come from a diverse range of physical locations and demographic groups. Given that our visitors are anonymous, our analysis cannot account for the random effects of COVID-19 across geographic locations.

Chat-GPT and LLMs: Our data collection stopped in September of 2022, before the wide adoption of LLMs [69]. ChatGPT was publicly released in November 2022. Thus, our findings relate more to non-LLM-assisted edits. Future work could examine how interest in a data source influences an individual's use of an LLM when editing and maintaining that same data.

Rise in CS admissions: An increased interest in Computer Science degrees in the US from 2005 to 2023 might contribute to the increase in visitors and engagement we see in our data [135].

Geographic Location and External Events: Our system does not collect personally identifiable information, nor does it require user accounts to edit data. Therefore, we cannot determine with 100% certainty if the same visitor appears across multiple years. We can only identify repeat visitors if they use the same browser on the same device. As such, we cannot claim to track fixed groups of visitors longitudinally. Among visitors who used the same browser/device, 6 made edits in 2021 and 2022, and 19 visited our system in both years. Some visitors may visit only once; it is impossible to determine whether this is true due to our system's anonymous visitor tracking.

Visitor Privacy and Data Ethics: Future work should balance preserving user anonymity (shown to positively influence engagement [109, 141]) with implementing fixed user tracking and account creation, as in Wikipedia. If enough covariates can be identified, such as geography and local events from IP Addresses, and visitor demographics (age, prospective student, current student, faculty, staff, etc.), then analysis methods such as propensity score matching could be applied to study new features in a public system over time. There is recent research applying propensity score matching with platforms requiring user accounts or IPs to contribute, such as Reddit [35] and Wikipedia [152].

9 Conclusion

We first developed a new version of our public data system Drafty to attract organic visitors to become contributors and help them continuously engage and interact with our systems. We developed two Dynamic sources that require low effort to maintain: CS Open Rankings and the Databaits API. They generate automated insights from Drafty's tabular data on computer science faculty profiles in the

US and Canada. These systems were designed to automatically and continuously attract and engage visitors who make accurate contributions and interact, building on prior work in Reciprocity [23, 114], Self-Determination Theory [43], and Boundary Objects [73] to intrinsically motivate contributions to the dataset's community [33, 124].

Visitors from our automated dynamic sources were just as accurate as those we asked (i.e., friend-sourcing [13]). Our dynamic sources require far less effort from data curators. Visitors who created automatic insights in Drafty using the Did You Know feature made more contributions and had higher engagement over time than those who did not. Notably, visitors preferred to see a random Did You Know compared to a similar one. Thus, future recommendation systems could employ randomization to appeal to the preferences of anonymous communities sharing similar interests. Thus, appealing to user interests aligns with critical thinking in AI co-creation tasks [37].

Our approach to building systems for groups of like-minded users allows you to attract everyday visitors to become unpaid contributors in ways that popular platforms cannot. Future work can compare the efforts of unpaid everyday users in smaller systems with those in large platforms like Wikipedia and WikiData. What if we expand public data systems to integrate AI and LLMs into the user-contribution loop to augment human behavior further? As data evolves, engaging the right user to maintain recent data will be necessary, as the volume of data society generates and organizes never stops. Without human contributions to public information, what will the balance become between human and AI-curated information?

Acknowledgments

We especially thank Alice Marbach for helping with CS Open Rankings. We also thank Zoya Bylinskii, John F Hughes, and Malte Schwarzkopf for their timely feedback and intellectual support. Lastly, we thank the thousands of visitors to Drafty and Computer Science Open Rankings for contributing their time and effort to maintaining Drafty's evolving tabular dataset of Computer Science Faculty profiles over the past 10 years.

References

[1] Hervé Abdi. 2010. Holm's sequential Bonferroni procedure. *Encyclopedia of research design* 1, 8 (2010), 1–8.
[2] Ibrahim Adeyanju. 2015. Generating weather forecast texts with case based reasoning. *arXiv preprint arXiv:1509.01023* (2015).
[3] Bryan Alexander. 2020. The little spreadsheet that could, and did: crowdsourcing COVID-19, higher education, data, and stories. https://bryanalexander.org/research-topics/the-little-spreadsheet-that-could-and-did-crowdsourcing-covid-19-higher-education-data-and-stories/. (Accessed on 04/12/2021).
[4] Yann Algan, Yochai Benkler, Mayo Fuster Morell, and Jérôme Hergueux. 2013. Cooperation in a Peer Production Economy Experimental Evidence from Wikipedia. *Available at SSRN 2843518* (2013).
[5] Aitor Apaolaza, Simon Harper, and Caroline Jay. 2013. Understanding users in the wild. In *Proceedings of the 10th international cross-disciplinary conference on web accessibility*. 1–4.
[6] Martina Balestra, Coye Cheshire, Ofer Arazy, and Oded Nov. 2017. Investigating the motivational paths of peer production newcomers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6381–6385.
[7] Albert-Laszlo Barabasi. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435, 7039 (2005), 207–211.
[8] Yochai Benkler. 2016. Peer production and cooperation. In *Handbook on the Economics of the Internet*. Edward Elgar Publishing, Cheltenham, United Kingdom.
[9] Emery D. Berger. 2020. CSRankings. Retrieved November 1, 2021 from http://csrankings.org/
[10] Michael Bernstein, Mike Bright, Ed Cutrell, Steven Dow, Elizabeth Gerber, Anupam Jain, and Anand Kulkarni. 2013. Micro-volunteering: helping the helpers in development. In *Proceedings of the 2013 CSCW Conference on Computer Supported Cooperative Work Companion*. ACM, New York, NY, USA, 85–88.
[11] Michael S Bernstein. 2013. Crowd-powered systems. *KI-Künstliche Intelligenz* 27, 1 (2013), 69–73.
[12] Michael S Bernstein, Mark S Ackerman, Ed H Chi, and Robert C Miller. 2011. The trouble with social computing systems research. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 389–398.
[13] Erin Brady, Meredith Ringel Morris, and Jeffrey P Bigham. 2014. Social microvolunteering: Donating access to your friends for charitable microwork. In *Second AAAI Conference on Human Computation and Crowdsourcing*. AAAI, Menlo

Park, CA, USA, 8–9.
[14] Erin Brady, Meredith Ringel Morris, and Jeffrey P Bigham. 2015. Gauging receptiveness to social microvolunteering. In *Proceedings of the 33rd Annual ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1055–1064.
[15] Gordon Burtch, Yili Hong, Ravi Bapna, and Vladas Griskevicius. 2018. Stimulating online reviews by combining financial incentives and social norms. *Management Science* 64, 5 (2018), 2065–2082.
[16] Gordon Burtch, Dokyun Lee, and Zhichen Chen. 2024. The consequences of generative AI for online knowledge communities. *Scientific Reports* 14, 1 (2024), 10413.
[17] L Cacciapuoti, VB Kostov, M Kuchner, EV Quintana, KD Colon, J Brande, SE Mullally, Q Chance, JL Christiansen, JP Ahlers, et al. 2025. VizieR Online Data Catalog: TESS Triple-9 Catalog (Cacciapuoti+, 2022). *VizieR Online Data Catalog* 751 (2025), J–MNRAS.
[18] Luca Cacciapuoti, Veselin B Kostov, Marc Kuchner, Elisa V Quintana, Knicole D Colón, Jonathan Brande, Susan E Mullally, Quadry Chance, Jessie L Christiansen, John P Ahlers, et al. 2022. The TESS Triple-9 Catalog: 999 uniformly vetted exoplanet candidates. *Monthly Notices of the Royal Astronomical Society* 513, 1 (2022), 102–116.
[19] Dana R Carney and Mahzarin R Banaji. 2012. First is best. *PloS one* 7, 6 (2012), e35088.
[20] Sarah E Chasins, Maria Mueller, and Rastislav Bodik. 2018. Rousillon: Scraping Distributed Hierarchical Web Data. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 963–975.
[21] Yan Chen, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li. 2010. Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review* 100, 4 (2010), 1358–1398.
[22] Justin Cheng, Jaime Teevan, and Michael S Bernstein. 2015. Measuring crowdsourcing effort with error-time curves. In *Proceedings of the 33rd Annual ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1365–1374.
[23] Robert B Cialdini. 2001. The science of persuasion. *Scientific American* 284, 2 (2001), 76–81.
[24] Robert B Cialdini and Robert B Cialdini. 2007. *Influence: The psychology of persuasion*. Vol. 55. Collins New York.
[25] Aaron Clauset, Samuel Arbesman, and Daniel B Larremore. 2015. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances* 1, 1 (2015), e1400005.
[26] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (Honolulu, Hawaii, USA) *(IUI '07)*. ACM, New York, NY, USA, 32–41. https://doi.org/10.1145/1216295.1216309
[27] Paramita Das, Bhanu Prakash Reddy Guda, Debajit Chakraborty, Soumya Sarkar, and Animesh Mukherjee. 2021. When expertise gone missing: Uncovering the loss of prolific contributors in Wikipedia. *arXiv preprint arXiv:2109.09979* (2021).
[28] Paramita Das, Bhanu Prakash Reddy Guda, Sasi Bhushan Seelaboyina, Soumya Sarkar, and Animesh Mukherjee. 2022. Quality change: Norm or exception? Measurement, analysis and detection of quality change in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–36.
[29] Nediyana Daskalova, Eindra Kyi, Kevin Ouyang, Arthur Borem, Sally Chen, Sung Hyun Park, Nicole Nugent, and Jeff Huang. 2021. Self-E: Smartphone-Supported Guidance for Customizable Self-Experimentation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13.
[30] Victor De Boer, Michiel Hildebrand, Lora Aroyo, Pieter De Leenheer, Chris Dijkshoorn, Binyam Tesfa, and Guus Schreiber. 2012. Nichesourcing: harnessing the power of crowds of experts. In *International Conference on Knowledge Engineering and Knowledge Management*. Springer, New York, NY, USA, 16–20.
[31] Esra Cemre Su de Groot and Ujwal Gadiraju. 2024. " Are we all in the same boat?" Customizable and Evolving Avatars to Improve Worker Engagement and Foster a Sense of Community in Online Crowd Work. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26.
[32] Dong Deng, Raul Castro Fernandez, Ziawasch Abedjan, Sibo Wang, Michael Stonebraker, Ahmed K Elmagarmid, Ihab F Ilyas, Samuel Madden, Mourad Ouzzani, and Nan Tang. 2017. The Data Civilizer System.. In *CIDR*.
[33] Xuefei Nancy Deng and Kshiti D Joshi. 2016. Why individuals participate in micro-task crowdsourcing work environment: Revealing crowdworkers' perceptions. *Journal of the Association for Information Systems* 17, 10 (2016), 3.
[34] Rabiyatou Diouf, Edouard Ngor Sarr, Ousmane Sall, Babiga Birregah, Mamadou Bousso, and Sény Ndiaye Mbaye. 2019. Web scraping: state-of-the-art and areas of application. In *2019 IEEE international conference on big data (big data)*. IEEE, 6040–6042.
[35] Anna Fang, Wenjie Yang, and Haiyi Zhu. 2025. Shaping Online Dialogue: Examining Community Rules and Discussion Structures on Reddit. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–10.
[36] Besnik Fetahu, Avishek Anand, and Maria Koutraki. 2019. Tablenet: An approach for determining fine-grained relations for wikipedia tables. In *The World Wide Web Conference*. ACM, New York, NY, USA, 2736–2742.

[37] Pablo Flores Romero, Kin Nok Nicholas Fung, Guang Rong, and Benjamin Ultan Cowley. 2025. Structured human-LLM interaction design reveals exploration and exploitation dynamics in higher education content generation. *npj Science of Learning* 10, 1 (2025), 40.

[38] Claudia Flores-Saviaga, Ricardo Granados, Liliana Savage, Lizbeth Escobedo, and Saiph Savage. 2020. Understanding the complementary nature of paid and volunteer crowds for content creation. *Avances en Interacción Humano-Computadora* 1, 1 (2020), 37–44.

[39] Brian J Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*. 1–7.

[40] Maksym Gabielkov, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout. 2016. Social clicks: What and who gets read on Twitter?. In *Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science*. 179–192.

[41] Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze, and Gianluca Demartini. 2015. Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1631–1640. https://doi.org/10.1145/2702123.2702443

[42] Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. Advances and challenges in conversational recommender systems: A survey. *AI Open* 2 (2021), 100–126.

[43] Paul Grau, Babak Naderi, and Juho Kim. 2018. Personalized motivation-supportive messages for increasing participation in crowd-civic systems. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–22.

[44] Reza Hadi Mogavi, Ehsan-Ul Haq, Sujit Gujar, Pan Hui, and Xiaojuan Ma. 2022. More gamification is not always better: A case study of promotional gamification in a question answering website. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–32.

[45] Changyang He, Yue Deng, Lu He, Qingyu Guo, Yu Zhang, Zhicong Lu, and Bo Li. 2024. Engage wider audience or facilitate quality answers? a mixed-methods analysis of questioning strategies for research sensemaking on a community Q&A Site. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–31.

[46] Jonathan Herzig, Pawel Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Eisenschlos. 2020. TaPas: Weakly Supervised Table Parsing via Pre-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

[47] Benjamin Mako Hill and Aaron Shaw. 2020. Wikipedia and the End of Open Collaboration. *Wikipedia* 20 (2020).

[48] Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. 2015. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, New York, NY, USA, 419–429.

[49] Reid Holmes, Robert J Walker, and Gail C Murphy. 2006. Approximate structural context matching: An approach to recommend relevant examples. *IEEE Transactions on Software Engineering* 32, 12 (2006), 952–970.

[50] Maedeh Hosseinpour, Borano Llana, Yusra Suhail, Kushas Khadka, Grant Garcia, and Shaun Wallace. 2026. One Reader, Two Reading Profiles: Language-Dependent Dyslexia and Reading Behaviors in English-Spanish Readers. In *CHI 2026 Workshop on Science and Technology for Augmenting Reading (STAR)* (2026-04-16). Barcelona, Spain. https://chi-star-workshop.github.io/src/assets/pdf/papers/One%20Reader%20Two%20Reading%20Profiles_Hosseinpour%20-%20maedeh%20hp.pdf Workshop paper.

[51] Joey Chiao-Yin Hsiao, Mei-Hua Pan, Hao-Chuan Wang, and Jane Yung-Jen Hsu. 2015. The price of the priceless: Understanding estimated costs of work in friendsourcing. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 1893–1898.

[52] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2023. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems* (2023).

[53] Hasan Humayun, Masitah Ghazali, and Mohammad Noman Malik. 2023. Is sustained participation a myth in crowdsourcing? A review. *European Journal of Innovation Management* (2023).

[54] Hasan Humayun, Mohammad Nauman Malik, and Masitah Ghazali. 2024. Analysis of Motivational Theories in Crowdsourcing Using Long Tail Theory: A Systematic Literature Review. *International Journal of Crowd Science* 8, 1 (2024), 10–27.

[55] Kashif Imteyaz, Claudia Flores-Saviaga, and Saiph Savage. 2024. GigSense: An LLM-Infused Tool for Workers Collective Intelligence. *arXiv preprint arXiv:2405.02528* (2024).

[56] Lucie-Aimée Kaffee, Kemele M Endris, and Elena Simperl. 2019. When humans and machines collaborate: cross-lingual label editing in wikidata. In *Proceedings of the 15th International Symposium on Open Collaboration*. 1–9.

[57] Keiko Katsuragawa, Qi Shu, and Edward Lank. 2019. PledgeWork: Online volunteering through crowdwork. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–11.

[58] Warut Khern-am nuai, Karthik Kannan, and Hossein Ghasemkhani. 2018. Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research* 29, 4 (2018), 871–892.

[59] Aniket Kittur, Ed Chi, Bryan A Pendleton, Bongwon Suh, and Todd Mytkowicz. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World wide web* 1, 2 (2007), 19.

[60] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 453–456. https://doi.org/10.1145/1357054.1357127

[61] Aniket Kittur and Robert E Kraut. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM CSCW Conference on Computer Supported Cooperative Work*. ACM, 37–46.

[62] Aniket Kittur, Bryan Pendleton, and Robert E Kraut. 2009. Herding the cats: the influence of groups in coordinating peer production. In *Proceedings of the 5th international Symposium on Wikis and Open Collaboration*. 1–9.

[63] Maximilian Klein, Thomas Maillart, and John Chuang. 2015. The virtuous circle of Wikipedia: recursive measures of collaboration structures. In *Proceedings of the 18th ACM CSCW Conference on Computer Supported Cooperative Work & Social Computing*. 1106–1115.

[64] Michael Koch, Gerhard Schwabe, and Robert O Briggs. 2015. CSCW and social computing: the past and the future. *Business & Information Systems Engineering* 57, 3 (2015), 149–153.

[65] Marios Kokkodis, Theodoros Lappas, and Sam Ransbotham. 2020. From lurkers to workers: Predicting voluntary contribution and community welfare. *Information Systems Research* 31, 2 (2020), 607–626.

[66] Yehuda Koren and Robert Bell. 2015. Advances in collaborative filtering. *Recommender systems handbook* (2015), 77–118.

[67] Yehuda Koren, Steffen Rendle, and Robert Bell. 2021. Advances in collaborative filtering. *Recommender systems handbook* (2021), 91–142.

[68] Flip Korn, Xuezhi Wang, You Wu, and Cong Yu. 2019. Automatically generating interesting facts from wikipedia tables. In *Proceedings of the 2019 International Conference on Management of Data*. ACM, New York, NY, USA, 349–361.

[69] Chinmay Kulkarni, Tongshuang Wu, Kenneth Holstein, Q Vera Liao, Min Kyung Lee, Mina Lee, and Hariharan Subramonyam. 2023. LLMs and the Infrastructure of CSCW. In *Companion publication of the 2023 CSCW Conference on Computer Supported Cooperative Work and Social Computing*. 408–410.

[70] Natalie Kupferberg and Bridget McCrate Protus. 2011. Accuracy and completeness of drug information in Wikipedia: an assessment. *Journal of the Medical Library Association: JMLA* 99, 4 (2011), 310.

[71] Irene Langkilde and Kevin Knight. 1998. Generation that exploits corpus-based statistical knowledge. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

[72] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.

[73] Charlotte P Lee. 2007. Boundary negotiating artifacts: Unbinding the routine of boundary objects and embracing chaos in collaborative work. *Computer Supported Cooperative Work (CSCW)* 16, 3 (2007), 307–339.

[74] Guo Li, Haiyi Zhu, Tun Lu, Xianghua Ding, and Ning Gu. 2015. Is it good to be like Wikipedia? Exploring the trade-offs of introducing collaborative editing model to Q&A sites. In *Proceedings of the 18th ACM CSCW Conference on Computer Supported Cooperative Work & Social Computing*. 1080–1091.

[75] Qian Liu, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, and Jian-Guang Lou. [n. d.]. TAPEX: Table Pre-training via Learning a Neural SQL Executor. In *International Conference on Learning Representations*.

[76] Borano Llana, Alisa Baron, Haihan Yu, Maedeh Hosseinpour, Yusra Suhail, Sean Chin, Kushas Khadka, and Shaun Wallace. 2026. Reading with Diversity in Mind: Pupillometry and Typography Towards Inclusive Design for ADHD Readers. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–7.

[77] Travis Lloyd, Jennah Gosciak, Tung Nguyen, and Mor Naaman. 2025. AI Rules? Characterizing Reddit Community Policies Towards AI-Generated Content. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[78] Joy Mahapatra, Sudip Kumar Naskar, and Sivaji Bandyopadhyay. 2016. Statistical natural language generation from tabular non-textual data. In *Proceedings of the 9th International Natural Language Generation conference*. 143–152.

[79] Andrew Mao, Ece Kamar, Yiling Chen, Eric Horvitz, Megan E Schwamb, Chris J Lintott, and Arfon M Smith. 2013. Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. In *First AAAI conference on human computation and crowdsourcing*. AAAI, Menlo Park, CA, USA, 94–102.

[80] Catherine C Marshall, Partha SR Goguladinne, Mudit Maheshwari, Apoorva Sathe, and Frank M Shipman. 2023. Who broke Amazon Mechanical Turk? An analysis of crowdsourcing data quality over time. In *Proceedings of the 15th ACM Web Science Conference 2023*. 335–345.

[81] Winter Mason and Duncan J. Watts. 2009. Financial Incentives and the "Performance of Crowds". *ACM SigKDD Explorations Newsletter* 11, 2 (2009), 77–85. https://doi.org/10.1145/1600150.1600175

[82] Talie Massachi, John Roy, Lauren Choi, Gabriela Hoefer, Shaun Wallace, and Jeff Huang. 2024. Chirp: The Impact of Private Online Self-Disclosure on Perceived Social Support. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–25.

[83] Connor McMahon, Isaac Johnson, and Brent Hecht. 2017. The substantial interdependence of Wikipedia and Google: A case study on the relationship between peer production communities and information technologies. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11. 142–151.

[84] Arpit Merchant, Daksh Shah, Gurpreet Singh Bhatia, Anurag Ghosh, and Ponnurangam Kumaraguru. 2019. Signals matter: understanding popularity and impact of users on stack overflow. In *The World Wide Web Conference*. ACM, New York, NY, USA, 3086–3092.

[85] Paweł Mikołajczak and Piotr Bajak. 2021. Does NGOs' commercialization affect volunteer work? The crowding out or crowding in Effect. *Public Organization Review* 21, 1 (2021), 103–118.

[86] Marc Miquel-Ribé, Cristian Consonni, and David Laniado. 2021. Wikipedia Editor Drop-Off. *Wiki Workshop at The Web Conference 2021* (2021).

[87] Abhijit Mishra, Anirban Laha, Karthik Sankaranarayanan, Parag Jain, and Saravanan Krishnan. 2019. Storytelling from structured data and knowledge graphs: An NLG perspective. In *Proceedings of the 57th annual meeting of the association for computational linguistics: Tutorial Abstracts*. 43–48.

[88] Marçal Mora-Cantallops, Salvador Sánchez-Alonso, and Elena García-Barriocanal. 2019. A systematic literature review on Wikidata. *Data Technologies and Applications* 53, 3 (2019), 250–268.

[89] Meredith Ringel Morris, Michael S Bernstein, Jeffrey P Bigham, Amy S Bruckman, and Andrés Monroy-Hernández. 2024. Is Human-AI Interaction CSCW?. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*. 95–97.

[90] Meredith Ringel Morris, Jeffrey P. Bigham, Robin Brewer, Jonathan Bragg, Anand Kulkarni, Jessie Li, and Saiph Savage. 2017. Subcontracting Microwork. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 1867–1876. https://doi.org/10.1145/3025453.3025687

[91] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What Do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 1739–1748. https://doi.org/10.1145/1753326.1753587

[92] Babak Naderi. 2018. *Motivation of workers on microtask crowdsourcing platforms*. Springer, New York, NY, USA.

[93] Mohammad Namvarpour and Afsaneh Razi. 2024. Uncovering contradictions in human-AI interactions: Lessons learned from user reviews of replika. In *Companion Publication of the 2024 CSCW Conference on Computer-Supported Cooperative Work and Social Computing*. 579–586.

[94] Khushnood Z Naqshbandi, Chunfeng Liu, Silas Taylor, Renee Lim, Naseem Ahmadpour, and Rafael Calvo. 2020. "I Am Most Grateful." Using Gratitude to Improve the Sense of Relatedness and Motivation for Online Volunteerism. *International Journal of Human–Computer Interaction* 36, 14 (2020), 1325–1341.

[95] Sneha Narayan, Jake Orlowitz, Jonathan Morgan, Benjamin Mako Hill, and Aaron Shaw. 2017. The Wikipedia Adventure: Field Evaluation of an Interactive Tutorial for New Users. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1785–1799. https://doi.org/10.1145/2998181.2998307

[96] Sneha Narayan, Jake Orlowitz, Jonathan T Morgan, and Aaron Shaw. 2015. Effects of a Wikipedia Orientation Game on New User Edits. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. 263–266.

[97] Senthilselvan Natarajan, Subramaniyaswamy Vairavasundaram, Sivaramakrishnan Natarajan, and Amir H Gandomi. 2020. Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data. *Expert Systems with Applications* 149 (2020), 113248.

[98] Phuc Nguyen, Ikuya Yamada, Natthawut Kertkeidkachorn, Ryutaro Ichise, and Hideaki Takeda. 2020. MTab4Wikidata at SemTab 2020: Tabular Data Annotation with Wikidata.. In *SemTab@ ISWC*. ACM, New York, NY, USA, 86–95.

[99] ThaoAn Lily Nguyen, Borano Llana, Jonathan Piccoli, Yusra Suhail, Maedeh Hosseinpour, and Shaun Wallace. 2026. REARS: An Interface for Browser-Based Eye-Tracking in Remote Readability Studies. In *CHI 2026 Workshop on Science and Technology for Augmenting Reading (STAR)* (2026-04-16). Barcelona, Spain. https://chi-star-workshop.github.io/src/assets/pdf/papers/CHI_STAR_REARS_FINAL%20-%20Borano%20Llana.pdf Workshop paper.

[100] Aleena Gertrudes Niklaus, Tianyuan Cai, Zoya Bylinskii, and Shaun Wallace. 2023. Digital reading rulers: Evaluating inclusively designed rulers for readers with dyslexia and without. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–17.

[101] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

[102] DOE Office of Science. 2022. Argonne Computer Scientist Works on Power and Processor Efficiency While Diversifying Workforce. https://www.hpcwire.com/off-the-wire/argonne-computer-scientist-works-on-power-and-processor-efficiency-while-diversifying-workforce/. (Accessed on 11/10/2025).

[103] Katherine O'Toole. 2023. Collaborative creativity in TikTok music duets. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[104] Antti Oulasvirta. 2008. Field experiments in HCI: promises and challenges. In *Future interaction design II*. Springer, 87–116.

[105] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*. 51–60.

[106] Alexandra Papoutsaki, Hua Guo, Danae Metaxa-Kakavouli, Connor Gramazio, Jeff Rasley, Wenting Xie, Guan Wang, and Jeff Huang. 2015. Crowdsourcing from Scratch: A Pragmatic Experiment in Data Collection by Novice Requesters. In *Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*. AAAI, Menlo Park, CA, USA, 140–149.

[107] Ankur P Parikh, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. 2020. Totto: A controlled table-to-text generation dataset. *arXiv preprint arXiv:2004.14373* (2020).

[108] Panupong Pasupat and Percy Liang. 2015. Compositional Semantic Parsing on Semi-Structured Tables. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics.

[109] Marcin Pawlukiewicz and Shaun Wallace. 2026. SEA-u-lator: Privacy Preserving Social Engineering Attack Simulator for Gmail. In *Proceedings of the 2026 Conference on Human Information Interaction and Retrieval*. 340–344.

[110] Matilda QR Pembury Smith and Graeme D Ruxton. 2020. Effective use of the McNemar test. *Behavioral Ecology and Sociobiology* 74, 11 (2020), 133.

[111] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Improving worker engagement through conversational microtask crowdsourcing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12.

[112] Sajjadur Rahman, Kelly Mack, Mangesh Bendre, Ruilin Zhang, Karrie Karahalios, and Aditya Parameswaran. 2020. Benchmarking spreadsheet systems. In *Proceedings of the 2020 acm sigmod international conference on management of data*. 1589–1599.

[113] Sam Ransbotham and Gerald C Kane. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *Mis Quarterly* 35 (2011), 613–627.

[114] Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. ACM, New York, NY, USA, 1364–1378.

[115] Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. 2011. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *Fifth International AAAI Conference on Weblogs and Social Media*. AAAI, Menlo Park, CA, USA, 321–328.

[116] Leah R Rosenzweig, Bence Bago, Adam J Berinsky, and David G Rand. 2021. Happiness and surprise are associated with worse truth discernment of COVID-19 headlines among social media users in Nigeria. *Harvard Kennedy School Misinformation Review* 1 (2021), 1–37.

[117] Niloufar Salehi and Michael S Bernstein. 2018. Ink: Increasing worker agency to reduce friction in hiring crowd workers. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 2 (2018), 1–17.

[118] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. 2016. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM CSCW Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 813–822.

[119] Ben Schmidt. 2020. Ranking CS Graduate programs. http://benschmidt.org/post/2020-07-30/2020-07-30-cs-rankings/ Accessed on December 20, 2022.

[120] Linda See, Alexis Comber, Carl Salk, Steffen Fritz, Marijn van der Velde, Christoph Perger, Christian Schill, Ian McCallum, Florian Kraxner, and Michael Obersteiner. 2013. Comparing the quality of crowdsourced data contributed by expert and non-experts. *PloS one* 8, 7 (2013), e69958.

[121] Farhana Shahid. 2024. Human-AI Collaboration to Facilitate Culturally-Aware Content Moderation. In *Companion Publication of the 2024 CSCW Conference on Computer-Supported Cooperative Work and Social Computing (CSCW Companion'24). Association for Computing Machinery, New York, NY, USA*. 9–12.

[122] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. 2020. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 453–463.

[123] J Sinclair, Paul J Taylor, and Sarah Jane Hobbs. 2013. Alpha level adjustments for multiple dependent variable analyses and their applicability–a review. *International Journal of Sports Science and Engineering* 7, 1 (2013), 17–20.

[124] Sebastian Spaeth and Sven Niederhöfer. 2020. User motivations in peer production. *The handbook of peer production* (2020), 123–136.

[125] Namrata Srivastava, Jennifer Healey, Rajiv Jain, Guanli Liu, Ying Ma, Borano Llana, Dragan Gasevic, Tilman Dingler, and Shaun Wallace. 2025. Priming at scale: An evaluation of using ai to generate primes for mobile readers. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–10.

[126] Michael Stefanone, Jeffery Hancock, Geri Gay, and Anthony Ingraffea. 2004. Emergent networks, locus of control, and the pursuit of social capital. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 592–595.

[127] Benjamin Stubbs, Polina Shpilker, Michael Sayers, Shaun Wallace, Noah M Daniels, and Alva Couch. 2024. AnVILMEDFORD: Documenting Cloud-Based Analysis Projects Using the MEDFORD Metadata Language. In *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 3158–3162.

[128] Yusra Suhail, Kushas Khadka, Borano Llana, Maedeh Hosseinpour, Emory Salaun, and Shaun Wallace. 2026. Different Font Attributes Influence English Readers' Perception of Font Size Across Age Groups. In *CHI 2026 Workshop on Science and Technology for Augmenting Reading (STAR)* (2026-04-16). Barcelona, Spain. https://chi-star-workshop.github.io/src/assets/pdf/papers/CHI_Workshop___Font_Size_Normalization_by_Age-5%20-%20Yusra%20Suhail.pdf Workshop paper.

[129] Yusra Suhail, Kushas Khadka, Borano Llana, Maedeh Hosseinpour, and Shaun Wallace. 2026. Personalizing Font Size Normalization: Linguistic Background Shapes Which Font Attributes Matter. In *Proceedings of the 34th ACM Conference on User Modeling, Adaptation and Personalization*. 203–212.

[130] Lu Sun, Stone Tao, Junjie Hu, and Steven P Dow. 2024. Metawriter: Exploring the potential and perils of ai writing support in scientific peer review. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–32.

[131] Ilya Sutskever, James Martens, and Geoffrey E Hinton. 2011. Generating text with recurrent neural networks. In *Proceedings of the 28th international conference on machine learning (ICML-11)*. 1017–1024.

[132] Pedro A Szekely, Daniel Garijo, Jay Pujara, Divij Bhatia, and Jiasheng Wu. 2019. T2WML: A Cell-Based Language to Map Tables into Wikidata Records.. In *ISWC Satellites*. ACM, New York, NY, USA, 45–48.

[133] Christopher Tensmeyer, Zoya Bylinski, Tianyuan Cai, Dave Miller, Ani Nenkova, Aleena Niklaus, and Shaun Wallace. 2023. Web Table Formatting Affects Readability on Mobile Devices. In *Proceedings of the ACM Web Conference 2023*. 1334–1344.

[134] David Thissen, Lynne Steinberg, and Daniel Kuang. 2002. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of educational and behavioral statistics* 27, 1 (2002), 77–83.

[135] Jodi L Tims, Cindy S Tucker, Mark A Weiss, and Stuart H Zweben. 2024. Student enrollment and retention in 2022–23 US undergraduate computing programs. *ACM inroads* 15, 4 (2024), 20–46.

[136] Christian Wagner and Ling Jiang. 2025. Death by AI: Will large language models diminish Wikipedia? *Journal of the Association for Information Science and Technology* 76, 5 (2025), 743–751.

[137] Shaun Wallace. 2022. Asynchronous Collaboration Systems for Evolving Information. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–5.

[138] Shaun Wallace, Zoya Bylinskii, Jonathan Dobres, Bernard Kerr, Sam Berlow, Rick Treitman, Nirmal Kumawat, Kathleen Arpin, Dave B Miller, Jeff Huang, et al. 2022. Towards Individuated Reading Experiences: Different Fonts Increase Reading Speed for Different Individuals. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–56.

[139] Shaun Wallace, Long Do, Alice Marbach, and Jeff Huang. 2020. *[Anon_Rankings]*. Retrieved June 28, 2026 from https://drafty.cs.brown.edu/csopenrankings/

[140] Shaun Wallace, Brendan Le, Luis A Leiva, Aman Haq, Ari Kintisch, Gabrielle Bufrem, Linda Chang, and Jeff Huang. 2020. Sketchy: Drawing inspiration from the crowd. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–27.

[141] Shaun Wallace, Talie Massachi, Jiaqi Su, Dave B Miller, and Jeff Huang. 2025. Towards Fair and Equitable Incentives to Motivate Paid and Unpaid Crowd Contributions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–32.

[142] Shaun Wallace, Alexandra Papoutsaki, Neilly H Tan, Hua Guo, and Jeff Huang. 2021. Case Studies on the Motivation and Performance of Contributors Who Verify and Maintain In-Flux Tabular Datasets. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.

[143] Shaun Wallace, Lucy Van Kleunen, Marianne Aubin-Le Quere, Abraham Peterkin, Yirui Huang, and Jeff Huang. 2017. Drafty: Enlisting Users to be Editors who Maintain Structured Data. In *Proceedings of the Fifth AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*. AAAI, Menlo Park, CA, USA, 187–196.

[144] Jing Wang, Gen Li, and Kai-Lung Hui. 2018. Do Monetary Incentives Create a Spillover Effect on Free Knowledge Contribution? Evidence from a Natural Experiment. *Evidence from a Natural Experiment (June 25, 2018)* (2018), 1–24.

[145] Liangwei Wang, Zhengxuan Zhang, Yi-Fan Cao, Fugee Tsung, and Yuyu Luo. 2026. TableTale: Reviving the Narrative Interplay Between Data Tables and Text in Scientific Papers. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–17.

[146] Morten Warncke-Wang, Vladislav R Ayukaev, Brent Hecht, and Loren G Terveen. 2015. The success and failure of quality improvement projects in peer production communities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 743–756.

[147] Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.

[148] Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, Maya Ramanath, Volker Tresp, and Gerhard Weikum. 2012. Natural language questions for the web of data. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. 379–390.

[149] Pengcheng Yin, Graham Neubig, Wen-tau Yih, and Sebastian Riedel. 2020. TaBERT: Pretraining for joint understanding of textual and tabular data. *arXiv preprint arXiv:2005.08314* (2020).

[150] Khushnood Z. Naqshbandi, Silas Taylor, Ajit G. Pillai, and Naseem Ahmadpour. 2021. Labour of Love: Volunteer Perceptions on Building Relatedness in Online Volunteering Communities. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–6.

[151] Ian Zapcic, Megan Fabbri, and Sharvari Karandikar. 2023. Using Reddit as a source for recruiting participants for in-depth and phenomenological research. *International Journal of Qualitative Methods* 22 (2023), 16094069231162674.

[152] Ark Fangzhou Zhang, Danielle Livneh, Ceren Budak, Lionel P Robert Jr, and Daniel M Romero. 2017. Crowd development: The interplay between crowd evaluation and collaborative dynamics in wikipedia. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–21.

[153] Shuo Zhang, Zhuyun Dai, Krisztian Balog, and Jamie Callan. 2020. Summarizing and exploring tabular data in conversational search. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1537–1540.

[154] Qian Zhao, Zihong Huang, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2016. Precision crowdsourcing: closing the loop to turn information consumers into information contributors. In *Proceedings of the 19th ACM CSCW Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 1615–1625.

[155] Yong Zheng. 2022. Context-aware collaborative filtering using context similarity: an empirical comparison. *Information* 13, 1 (2022), 42.

[156] Moyan Zhou, Soobin Cho, and Loren Terveen. 2025. LLMs in Wikipedia: Investigating How LLMs Impact Participation in Knowledge Communities. *arXiv preprint arXiv:2509.07819* (2025).

A Developing Drafty: Reducing Effort to Contribute & Infinite Scrolling Large Tabular Datasets

A key design focus for this new version of Drafty v2 is to reduce the effort to edit cells, add new rows, and delete existing rows. Reducing the time and effort required to edit data should increase visitors' motivation to contribute by lowering a barrier [22, 113]. When editing a single cell, a drop-down with autocomplete appears directly below the cell. This interface provides suggested values and highlights any previously contributed values for that row of data in orange. The previous version of Drafty v1 (pre-2020) required a lengthy process to add a new row via a complex wizard modal. Previous research shows that prior visitors were less likely to submit new rows of data because of the effort required to navigate this step-by-step interface [142]. We also optimized the interface for large-screen displays, given users' negative views of interacting with tables and spreadsheets on mobile devices [133]. The new version of Drafty v2 released in 2020, allows visitors to add a new row of data using an inline form embedded in the spreadsheet. When deleting a row, a visitor must provide a reason for the deletion. For example, it could be a duplicate, or the professor could have retired.

Drafty v1's (pre-2020) performed poorly in search engine results because it pulled rows to display from a server and did not store all data client-side. Our new version of Drafty v2, can handle near-infinite data client-side while providing a similar user experience with scrolling and search actions. We developed the pure client-side JavaScript library PolyMu[4] to handle and present hundreds of thousands of rows of data to the user. This data is also search-indexed. The spreadsheet is generated statically on the server every couple of minutes. The data is stored in a search-indexable HTML template element. This enables the page to load quickly, regardless of table size. PolyMu stores all table data in memory from the HTML template element. Then, PolyMu dynamically renders the correct row of data based on a user's scroll position, column sorting, and column search inputs.

B CS Open Rankings Overview and User Interface

CS Open Rankings is a dynamically generated web page that visualizes department-level rankings using Drafty's evolving dataset of Computer Science professor profiles. Visitors can explore interactive rankings and select the "D" icon to navigate to a filtered view of that professor's profile within Drafty. Rankings are based on four metrics: **Reputation (U.S. News)**, which reflects academic peer assessments; **Faculty Publications**, based on publication counts in selected top conferences since 2010, weighted by institutional authorship; **Placement Rank**, derived using a PageRank algorithm on faculty hiring networks, where degree-granting institutions are treated as sources and hiring institutions as destinations [119]; and **Recognition (Best Paper Awards)**, which assigns points based

[4] https://github.com/drafty-research

Type of Interaction	Interactions
Modify tabular data	edit cell; add new row; delete a row; visit from a source
Interact with cells	click; double-click; select multiple; copy; paste value
Interact with columns	sort by column; search (single/multiple); completed search (single/multiple); clear column search; copy all values from column
Interact with notes	add a note; vote thumbs up/down; deselect vote; open notes modal
Create an insight	create via row menu; create *similar*/*random* (Did You Know); tweet an insight; select insight value to auto-search Drafty
Other interactions	search Google for row data; view a different web page on Drafty

Table 7. The interactions that Drafty anonymously records user interactions. All interactions within the engagement metric fall under one of these seven interaction types.

on author affiliation and author order using a decreasing exponential scale across 30 top conferences. While CS Open Rankings uses the same geometric mean formula as `csrankings.org` [9], it differs in implementation, subfield selection, and faculty data source.

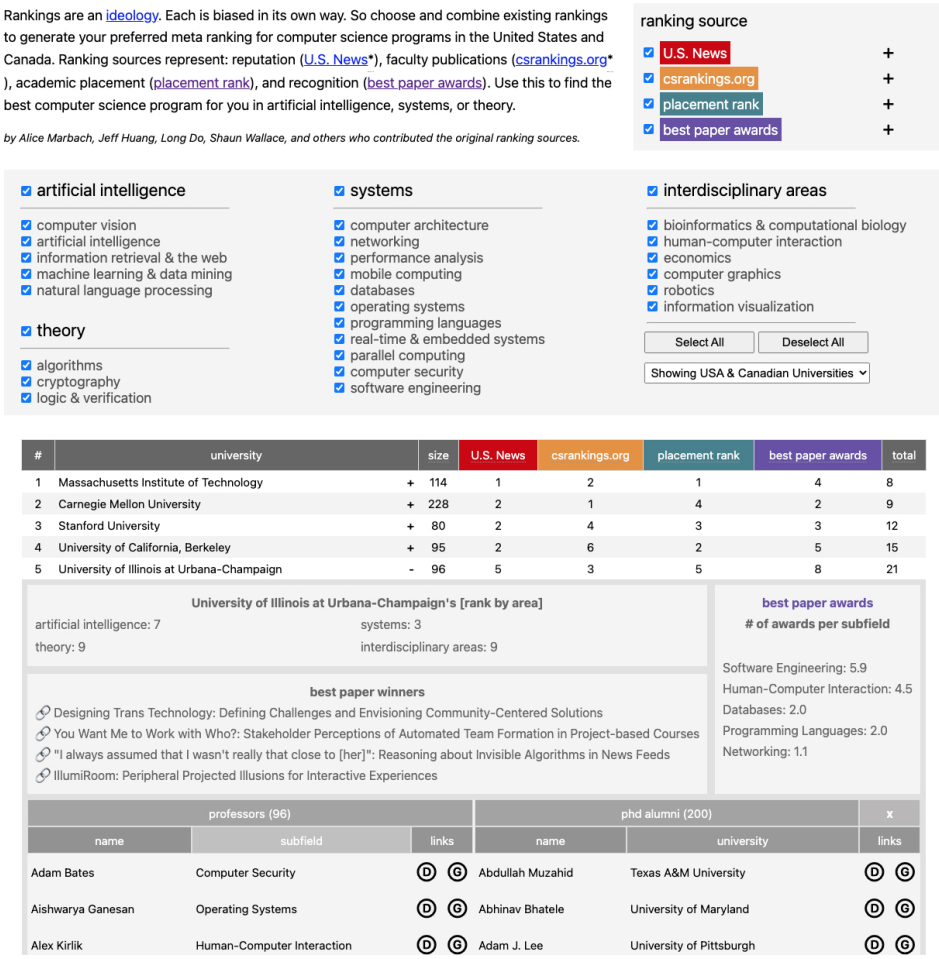


Fig. 7. This figure shows the user interface for CS Open Rankings. It is developed using HTML and JavaScript, allowing users agency over creating their own rankings using the various UI elements.

C How do Users Create an automatic insight in the Did You Know Modal?

Users can take these actions within the system's spreadsheet interface to generate an automatic insight from its tabular data.

(1) Right-click on a cell within the spreadsheet and select "See a Did You Know" to generate a random insight using the selected row's information. This method lets users generate a Did You Know related to specific rows.
(2) Select Did You Know? in the top menu bar. This generates a random insight by selecting random columns and values from the tabular dataset. A user might not have a cell selected, so this method allows them to generate a random insight without having first to edit data.
(3) Every time a user edits data (changing a cell's value, adding a new row, or deleting a row), the system generates a random Did You Know based on that row's information. This gives the user

a small reward for making an edit and another method to help them discover and interact with a Did You Know.

(4) When viewing the Did You Know Modal, a user can choose to see another similar insight. This generates a random type of insight using the columns and input values used to generate the previous insight. For example, if the previous insight used the values "Databases" and "McGill University", the next insight would create an insight using at least one of these values.

(5) When viewing the Did You Know Modal, a user can choose to see a random insight. This generates a random insight using the same method as Menu-Item-Random. This method allows users to continuously see a random Did You Know created from all of Drafty's tabular data.

D Databaits Validation Study: 16 Column Values to Generate Automatic Insights

The 16 subtopics (i.e., column values) that are featured across the insights in the Databaits validation study. These are selected from Drafty's Computer Science Professor dataset and a dataset of video game sales.

Topics: colleges and universities; video games; Harvard University; Carnegie Mellon University; Artificial Intelligence; Human-Computer Interaction; MIT; Brown; Princeton; Tsinghua; Nintendo; Sony Computer Entertainment; Sports video games; Action video games; Shooter video games; Simulation video games.

E Databaits API: Configuration Overview

Databaits API: Configuring System Behaviors to Create Insights. When the Databaits API creates an insight, it can take in as optional inputs the type of Databait to create, multiple column names, multiple values per column, and any of the other configuration parameters below:

- subject: the entity being described by each row in the table (e.g., "CS professors").
- pronoun: the pronoun that represents the subject (e.g., "who").
- column names(s): Often the column name that describes a column's information (e.g., "University").
- column phrase(s): a term or phrase that describes what the column's name does (e.g., "specialized in" for Subfield).
- column value(s): a unique set of values from a column.
- comparison statistic descriptor: a phrase that describes statistics that compare one column value to another (e.g., "twice as many").
- growth phase statistic descriptor: a phrase that describes statistics that track the change over time of one column value (e.g., "doubled").
- time range: the distance between the minimum and maximum for a time-based column value.
- time range descriptor: a unit of measurement for the time range (e.g., "years").
- historical event(s): historical events and they year/s they took place for Databait Type 7 (e.g., "the iPhone was released" and "2007").

These can be provided by passing a configuration file, or the Databaits API can select any of these at random.

For example, Figure 3's sentence template consists of static and variable parts. The variable parts are colored for distinction. These parts change based on which column names or column values the Databaits API used to generate each statistic. The static portions describe information inherent to the Databait type. These are shared across Databaits of the same type.

By allowing this type of configuration per dataset it expands the variety of data types and datasets that can be plugged into the Databaits API to generate insights.

Databait Type		Templates & then Samples computer from the CS Professor Tabular Dataset per Databait Type
Type 1 Single column Δ	Template	Over the past [time range] [time range descriptor], the total number of [subject] [pronoun] [column descriptor] [column value] [growth phase statistic] [statistic].
	Sample	Over the past 25 years, the total number of CS professors who specialized in Human-Computer Interaction more than quadrupled.
Type 2 Compare two values from one column	Template	[Statistic] [comparison statistic phrase] [subject] [column descriptor] [1st column value] than by [2nd column value] in the past [time range].
	Sample	6 times as many CS professors were hired by Carnegie Mellon University than by Brown University in the past 25 years.
Type 3 Compare two column frequencies	Template	Over the past [time range], the total number of [subject] [pronoun] [1st column descriptor] [1st column value] and [2nd column descriptor] [2nd column value] [growth phase statistic].
	Sample	Over the past 25 years, the total number of CS professors who were hired by Carnegie Mellon University and specialized in Artificial Intelligence increased 7 times.
Type 4 Compare column 2 frequency given shared column 1	Template	[Statistic] [comparison statistic descriptor] [subject] [pronoun] [1st column descriptor] [shared column value] [2nd column descriptor] [1st column value] than by [2nd column value] in the past [time range].
	Sample	Twice as many CS professors who specialized in Databases were hired by Brown University than by Carnegie Mellon University in the past 20 years.
Type 5 Compare against average value over time	Template	[Statistic] [comparison statistic descriptor] [subject] [column descriptor] [column value] than in the average [column name] in the past [time range].
	Sample	5 times as many CS professors specialized in Machine Learning and Data Mining than in the average SubField in the past 25 years.
Type 6 Single column proportion	Template	[Statistic] percent of all [subject] from the past [time range] [column descriptor] [column value].
	Sample	16 percent of all new CS professors from the past 5 years specialized in Machine Learning and Data Mining.
Type 7 Historical event column Δ	Template	The total number of [subject] [pronoun] [column descriptor] [column value] [change over time statistic descriptor] since [user-specified historic event].
	Sample	The total number of CS professors who specialized in Artificial Intelligence has nearly quadrupled since 1997, when IBM's Deep Blue beat Garry Kasparov.
Type 8 Proportional shared value	Template	[Statistic] percent of [subject] [1st column descriptor] and [2nd column descriptor] the same [column name].
	Sample	13 percent of CS professors got their bachelor's degrees from and got their doctorate degrees from the same university.
Type 9 Column value proportion Δ between two times	Template	[Statistic] percent of [subject] [column descriptor] [column value] in [year], compared to [raw statistic] percent in [year].
	Sample	11 percent of CS professors specialized in Algorithms & Complexity in 2020, compared to 18 percent in 1991.
Type 10 Compare Δ of two column values at two times	Template	[Less OR More] [subject] [column descriptor] [1st column value] in [year] than in [year], while [Less OR More] did in [2nd column value].
	Sample	Less CS professors specialized in High-Performance Computing in 2020 than in 2001, while more did in Robotics.

Fig. 8. This figure contains a short summary, template, and example of each Databait type (types 1-10). The template colors map to red for time, blue for columns, teal for the tabular dataset, and yellow for statistics. Examples for all 10 Databait types are in the Appendix; see Figure 8 in section E.

Databaits API: Calculating Statistics for Databaits. When the Databaits API is called, it either uses the predefined Databait type provided in the call or randomly selects a suitable Databait type capable of handling the number of columns and values provided. If the Databaits API cannot compute a statistic given the provided inputs, it will select random column names and values and create a random Databait type. After successfully creating the aggregate statistic, it will attempt to plug it into the sentence template. If this last step succeeds, the Databaits API will return the final insight via its RESTful API.

To ensure diverse insights and statistics, the Databaits API does not compute statistics for the largest effect sizes. For example, in the Computer Science professor dataset, every insight comparing the number of faculty per department would likely select Carnegie Mellon University because it has two to ten times as many professors as other departments. Future work could optimize these effect sizes to find the right balance between variety and usefulness per insight.

Databaits API: Generating Sentences. Once a statistic is calculated, it is converted into an English sentence. Each Databait type has a unique template comprising static and variable parts. The variable parts depend on the inputs used to generate the statistic.

Some variables, such as the subject and time range, can be plugged into sentence templates without modification. While, column names such as “University”, “SubField”, or perhaps something more

syntactically complicated such as "global_sales", must be translated into plain English before they can blend into sentences. To do so, the user specifies a dictionary of *column descriptors* that maps each column name to an English phrase describing its function. For example, the "SubField" column could be mapped to the phrase "specialize in." Combined with the *subject* and *pronoun* variables, column descriptors create a natural phrase that depicts the kind of data represented by each column. In this case, the full phrase would be "CS professors who specialize in" followed by a column value of the "SubField" column such as "Human-Computer Interaction."

Statistics are also converted into plain English to make information more digestible. If the insight compares two column values, the figure will be translated into a *comparison statistic descriptor*, such as "twice as many" or "nearly 10 times that of.". Suppose the insight is following the increase or decrease of a value over time. In that case, the figure will be converted into a *change over time statistic descriptor*, such as "doubled" or "grew nearly 10 times," followed by a *time range* variable, such as "over the past five years."

F Preliminary Databaits API Validation Study on Two Tabular Datasets

All insights presented to participants in this preliminary study are true based on the tabular data given to the Databaits API. No personally identifiable information was collected from participants.

Preliminary Databaits Validation Study: Method. We first introduced insights created by the Databaits API to the public through a preliminary study, which presented participants with insights created from two different tabular datasets (computer science professors and video game sales), then measured their reactions. Each participant read 24 different insights and then answered six follow-up questions per Databait. These questions measured whether participants found each insight interesting, believable, surprising, or easy to understand and if they wanted to learn more. Below is a sample set of questions:

Sample Databait: Over the past 25 years, the total number of games that were published by Nintendo and were classified as Sports nearly quadrupled.

- How interesting is this information? [Likert scale, 1 - "Not at all" to 6 - "Very"]
- How believable is this information? [Likert scale, 1 - "Not at all" to 6 - "Very"]
- How surprising is this information? [Likert scale, 1 - "Not at all" to 6 - "Very"]
- Is this sentence easy to understand? [Likert scale, 1 - "Not at all" to 6 - "Very"]
- Do you want to learn more about this information? [1 - "No" or 2 - "Yes"]
- Please share any thoughts or feedback on this sentence. [Open-ended]

This survey used a Likert scale with an even number of choices (1 to 6) so participants had to choose explicitly negative (1 to 3) or positive (4 to 6) answers. In addition, participants could only answer whether they wanted to learn more with a "Yes" or a "No." This binary choice best reflects whether users would choose to learn more in the real world. If an insight from the Databaits API were posted on Twitter, followed by a link to the source dataset it was created from, someone would either click on the link to learn more or decide not to click.

This survey also measured if participants' prior interest in the topics covered by the two datasets would react differently than those who did not. In other words, it sought to measure whether prior interest affected user reactions. Before participants saw any insights generated by the Databaits API, they answered 16 Likert scale questions with an even number of choices (1 to 6), indicating how interested they were in each of the 16 subtopics that later appeared in each insight in the survey. For example, one of Databaits' generated insights in the survey mentioned subtopics such as video games, Nintendo, and the Sports genre. Here are example questions participants answered:

- How interested are you in video games? [Likert scale, 1 - "Not at all" to 6 - "Very"]
- How interested are you in Nintendo? [Likert scale, 1 - "Not at all" to 6 - "Very"]

- How interested are you in Sports video games? [Likert scale, 1 - "Not at all" to 6 - "Very"]

We designed the survey to determine if some Databait types are more effective than others. Each Databait type appeared twice using the two different tabular datasets. The first tabular dataset is Drafty's Computer Science professors. Each row contains a CS professor's name, their subfield of expertise, the university where they teach, the year they were hired at that university, and the universities where they received their Bachelor's and Doctorate degrees. The second dataset is a table from Kaggle detailing historical video game sales. Each row contains the name of the video game, its publisher, the platform for which it was designed, its genre, the year it was published, and its sales globally: in North America, Europe, and Japan.

Preliminary Databaits Validation Study: Data Overview. In total, the survey received 57 responses from paid crowdworkers on Amazon Mechanical Turk and 33 responses from those on Prolific. All paid crowdworkers were paid $2.25 per task (around $9 per hour). The median completion time for paid crowdworkers on Amazon Mechanical Turk was 16 minutes and 22 seconds. The minimum qualifications for paid crowdworkers on Amazon Mechanical Turk were an acceptance rate of 99% or above and the completion of at least 500 tasks. The median completion time for paid crowdworkers on Prolific was 17 minutes and 32 seconds. At the time we conducted this preliminary study, Prolific did not allow us to restrict availability by number of tasks completed or prior success rate per crowdworker.

We recruited paid crowdworkers from two different sources because the initial responses from Amazon Mechanical Turk were lower quality than we initially expected. Some Amazon Mechanical Turk responses were most likely written by bots, as evidenced by nonsensical answers to open-ended questions or no variance in Like scale answers (e.g., providing the same answer every time). After excluding participants who displayed these behaviors, there were 41 responses from Amazon Mechanical Turk and 32 from Prolific. These 73 responses are used in the following sections. We found no difference in the responses from Amazon Mechanical Turk and Prolific used in the final analysis.

Preliminary Databaits Validation Study: Results & Discussion. Results from the preliminary study indicate that insights generated by the Databaits API are interesting, believable, and could be effective in attracting attention online and motivating people to engage with the source datasets. Participants responded positively to insights (answered four or above on the 6-point scale) on Likert scale questions asking if they found the insight interesting 77% of the time, believable 91% of the time, and easy to understand 93% of the time. Participants only found the insight surprising 45% of the time. Participants indicated they did not want to learn more about the insight around 70% of the time. While this response might be less enthusiastic compared to the other questions, it still represents a high level of user interest compared to the attention most links receive on Twitter, where only 60% of links on Twitter receive a single click [40].

If the Databaits API could create tweets that were believable and lead people to learn more about the source, this could motivate integrating the Databaits API into Drafty to attract more visitors. Pearson's Correlation shows medium to large effects between a participant wanting to learn more about the insight and feeling it is interesting ($r = 0.71$, $p < 0.05$), believable ($r = 0.46$, $p < 0.05$), and easy to understand ($r = 0.57$, $p < 0.05$). These results indicate that the insight alone might drive people to visit the source data. Pearson's Correlation shows no effect between a participant feeling surprised by the insight and wanting to learn more ($r = -0.09$). Databait's insights do not intend to motivate engagement or clicks by surprising people. Prior research has shown a correlation between clicks and people feeling surprised by misinformation campaigns [116]. The strongest Pearson's Correlation between all pairs of questions is between interesting and believability ($r = 0.80$, $p < 0.05$).

Previous research shows relationships between someone's prior interest in a topic and increased engagement [25, 90]. This survey measures prior interest by participants answering 6-point Likert

scale questions on their interest in the specific 16 subtopics (i.e., column values) featured across the insights. The 16 subtopics are listed in Appendix section D.

The answers to these Likert scale questions can be compared against how participants felt about individual insights they answered questions about later on in the survey. Pearson's Correlation shows prior interest in the column values was strongly correlated with whether the reader found the insight interesting ($r = 0.80$, $p < 0.05$). Prior interest also had moderate correlations with whether participants found the insights believable ($r = 0.67$, $p < 0.05$) and if they wanted to learn more ($r = 0.42$, $p < 0.05$). Pearson's Correlation shows a small effect between someone's prior interest and surprise had a slightly negative correlation ($r = 0.20$, $p < 0.05$), indicating that someone's prior knowledge about specific information makes insights about that information less surprising. The insight being "easy to understand" showed a medium effect ($r = 0.42$, $p < 0.05$) with prior interest. This might be because the insights with simple structures remained straightforward regardless of what the insight was about. The correlation between someone's prior interest and their wanting to know more about the insight should help attract visitors to the system that stores the insight's source data. For example, Drafty is a system that hosts a source.

We also examined whether some Databait types are more effective than others. If one Databait type was found not believable or interesting, it should not be used to generate and share insights in the wild. To determine this we used the interquartile range method (IQR) to compare the average responses to questions across all Databait types. First, it appears users prefer types with simple sentence structures over more complex ones (i.e., lower word counts and less punctuation). For example, Type 1 had high averages across all participants for whether participants found the insight interesting, believable, easy to understand, and wanted to know more. Type 1 is the simplest in structure and has only one column and one column value. It had the highest averages across all 10 types for ease of understanding (5.2) and for making users want to learn more (1.4). Moreover, it had the second-highest overall averages for interest (4.0) and believability (4.6). On the other hand, Type 4 is the most complex, involving two columns and three column values. It had the lowest average scores for being interesting (3.6) and believable (4.1), and the second-lowest score for being easy to understand (4.6). In summary, insights generated by the Databaits API with simpler sentence structures were more likely to attract users. Along with Type 4, Type 10 is another complex Databait type that also showed relatively weaker results. It had the lowest average score for being easy to understand (4.6), as well as the second lowest for being interesting (3.6) and having users want to learn more (1.2). However, these scores from the responses still fall within 1.5 times their respective interquartile ranges (IQR) from the first quartiles. Therefore, we decided that the relative weaknesses of Types 4 and 10 did not outweigh the diversity in sentence templates they provided and chose not to eliminate them from integrating the Databaits API into Drafty to create the new Did You Know feature.

G Reddit Post to Recruit People to Visit Drafty

An example title used on Reddit was "Some Interesting Insights on CS Depts and Faculty in the US/Canada" and an example post sharing Drafty contained the text shown in the Appendix section G.

> "Hi, I just wanted to share a resource for the CS community we have developed at the Brown HCI group. We have helped upkeep a dataset of tenure-track CS faculty in the US/Canada for 8 years.
> The system (and users) generate automatic insights from the data, for example: Over the past 10 years, the total number of CS professors who were hired by Stanford University and specialized in Computer Graphics more than doubled."

H Email Template for Asking Chairs, Directors, and Faculty of Computer Science Departments to Visit Drafty

Any text in this template enclosed in <text> is replaced with actual numbers and names from the email.

> <number_of_faculty> faculty from <university_name> are listed in this public dataset of CS professors' data, a resource we've been compiling for the CS community.
>
> CS Open Rankings uses this data, which features your department.
>
> According to CS Open Rankings, (university_name> is ranked <ranking_number> in <ranking_type> among US and Canadian Universities.
>
> The listings consist of tenure-track faculty members who can advise CS PhD students. If there are any corrections for your current faculty or alumni, please feel free to edit the CS professors' data. We're also happy to share the raw data if you contribute.

I Additional Results

Post-hoc Comparisons per Visit	Average Session Length in Seconds ($\|r\|$)	Average Interactions ($\|r\|$)
Static vs. Organic	0.18	0.17
Asking vs. Organic	0.09	0.05
Asking vs. Static	0.05	0.08
Dynamic vs. Organic	0.19	0.25
Dynamic vs. Static	0.46	0.45
Dynamic vs. Asking	0.34	0.32

Table 8. Mann-Whitney U standardized effect sizes (rank-biserial correlation r) for post-hoc comparisons of average engagement session length in seconds and average interactions per visit across sources ranging from small (0.05) to large (0.46). We remove directionality reporting r as $|r|$.

I.1 Qualitative Insights from Drafty's Visitors

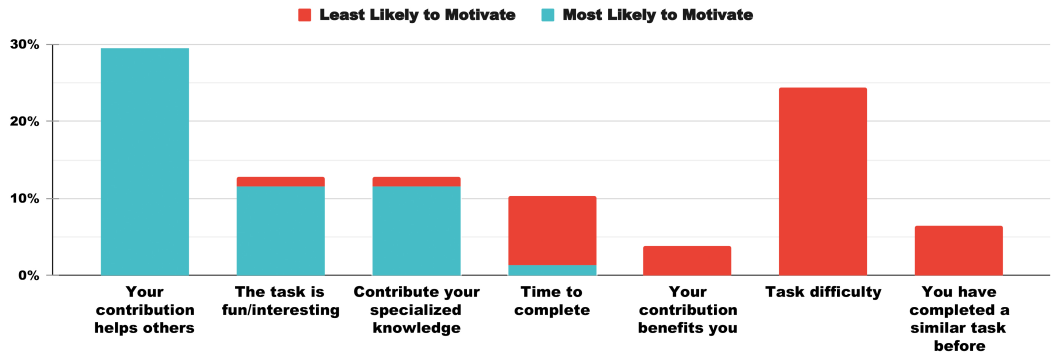


Fig. 9. Survey participants were asked to select what motivated them most and least from seven predefined options: "What motivates you to contribute to public data systems: like Drafty". The most common answer at 29% is that they wanted to help others. This beneficial intrinsic motivator mirrors prior crowdsourcing research [6, 90]. Another takeaway is that reducing the difficulty to contribute (complete a task) would reduce negative motivations for participants to edit and contribute data to Drafty.

In late 2022 and early 2023, we conducted two surveys with Drafty's anonymous visitors to gain qualitative feedback on their motivations for visiting and editing its data. Visitors could freely access our Qualtrics survey from a button in the spreadsheet interface. Visitors who completed the survey could submit their email for a one-in-four chance to receive a $25 Amazon gift card. Our institution's IRB approved the survey procedures. In total, 53 visitors freely chose to take the survey.

Participants answered three questions: Q1 was an optional open-ended prompt asking participants what motivates them to contribute voluntarily to public data systems like ours (see Table 9); Q2 asked them to select what motivated them most and least from seven predefined options from prior research [90, 141] (Figure 9); and Q3 was a 5-point Likert scale: "How difficult was it to contribute to Drafty?" All 53 participants completed Q2 and Q3; 32 also answered Q1. Effectively, Q1 is the open-ended version of Q2, participants answered Q1 before Q2.

We inductively coded Q1 responses using themes drawn from prior work [90, 141] and frequently occurring participant comments. Table 9 shows that motivations were primarily intrinsic and altruistic; for example, P5 reported they "added a missing professor from my university" and were motivated by "helping my community."

Participants also expressed broad interest in our CS faculty dataset, aligning with findings that intrinsic interest supports sustained engagement in peer production [79, 141]. Our companion site, CS Open Rankings, likely contributes to this interest: from January 1, 2023 to May 13, 2025, it directed 112,784 visits to Drafty from 5,860 unique CS faculty profiles.

In contrast to earlier versions of Drafty (pre-2020), we prioritized reduced effort to contribute in our redesign for first-time and return visitors alike. Our results (see Figure 9) show our visitors' motivation is likely more negatively affected by the effort to contribute rather than the time to contribute, mirroring prior work on unpaid contributors [57]. For Q3, 90% of participants stated that it was Very Easy, Easy, or Neutral to contribute to Drafty, while 0 participants stated it was Very Difficult.

Create / Maintain High-Quality Data	Your Contrib. Helps a Friend	Your Contrib. Helps Others	The Task is Fun / Interesting	Contribute your Specialized Knowledge	Your Contrib. Benefits You	Task Difficulty	Time to Complete
40.6%	6.3%	56.3%	9.4%	21.9%	9.4%	6.3%	6.3%

Table 9. Over half of the participants stated they contribute to systems and public datasets like Drafty, because they want to help others. This intrinsic motivation mirrors prior research [90]. This result mirrors the responses to their ranking question in Figure 9. This table codifies participants' answers to an optional open-ended question about what motivates them to contribute voluntarily to public data systems like ours. **Contrib.* is an abbreviation for Contribution.

Received 13 May 2025; revised 13 January 2026; accepted 9 April 2026