Aligning Data with the Goals of an Organization and Its Workers: Designing Data Labeling for Social Service Case Notes

Apoorva Gondimalla
School of Information
University of Texas at Austin
Austin, Texas, USA
apoorva.gondimalla@utexas.edu

Varshinee Sreekanth
School of Information
University of Texas at Austin
Austin, Texas, USA
varshinee@utexas.edu

Govind Joshi
School of Information
University of Texas at Austin
Austin, Texas, USA
govindjoshi@utexas.edu

Whitney Nelson
School of Information
University of Texas at Austin
Austin, Texas, USA
whitney.nelson@utexas.edu

Eunsol Choi
Department of Computer Science
University of Texas at Austin
Austin, Texas, USA
eunsol@utexas.edu

Stephen C. Slota
School of Information
University of Texas at Austin
Austin, Texas, USA
steveslota@gmail.com

Sherri R. Greenberg
LBJ School of Public Policy
University of Texas at Austin
Austin, Texas, USA
srgreenberg@mail.utexas.edu

Kenneth R. Fleischmann
School of Information
University of Texas at Austin
Austin, Texas, USA
kfleisch@ischool.utexas.edu

Min Kyung Lee
School of Information
University of Texas at Austin
Austin, Texas, USA
minkyung.lee@austin.utexas.edu

ABSTRACT

The challenges of data collection in nonprofits for performance and funding reports are well-established in HCI research. Few studies, however, delve into improving the data collection process. Our study proposes ideas to improve data collection by exploring challenges that social workers experience when labeling their case notes. Through collaboration with an organization that provides intensive case management to those experiencing homelessness in the U.S., we conducted interviews with caseworkers and held design sessions where caseworkers, managers, and program analysts examined storyboarded ideas to improve data labeling. Our findings suggest several design ideas on how data labeling practices can be improved: Aligning labeling with caseworker goals, enabling shared control on data label design for a comprehensive portrayal of caseworker contributions, improving the synthesis of qualitative and quantitative data, and making labeling user-friendly. We contribute design implications for data labeling to better support multiple stakeholder goals in social service contexts.

CCS CONCEPTS

• **Human-centered computing** → **Qualitative research**.

KEYWORDS

Social Work, Nonprofits, Case Notes, Data Collection Practices, Data Labeling, Design Ideas

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642014

ACM Reference Format:
Apoorva Gondimalla, Varshinee Sreekanth, Govind Joshi, Whitney Nelson, Eunsol Choi, Stephen C. Slota, Sherri R. Greenberg, Kenneth R. Fleischmann, and Min Kyung Lee. 2024. Aligning Data with the Goals of an Organization and Its Workers: Designing Data Labeling for Social Service Case Notes. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3642014

1 INTRODUCTION

Nonprofit social work is increasingly adopting a data-driven approach for performance evaluation, funding reports, policies, and program management [4, 38, 57]. Data-driven approaches require social workers to collect data for objectives beyond meeting the needs of a particular client, such as performance assessments of social workers, programs, and organizations. Unlike automated or systematized data collection processes in other industries, social work data collection relies on manual recording in the field because of the nuanced and subjective nature of the data [5]. Social workers are best suited for this task due to their close relationship with clients and understanding of the services provided. Prior work highlighted several challenges in data collection by social workers for performance and funding reports, such as misalignment with caseworkers' goals [4, 23], lack of motivation [6, 22], and system usability [5, 59], leading to inadequate and inaccurate data.

While prior human-computer interaction (HCI) research discusses challenges in data collection by social workers, very few studies explored how to improve this process [5, 8]. As a step in this direction, our study used a multi-stakeholder approach to explore the design of data labeling for social service case notes in homeless case management. Data labeling is a data collection process caseworkers use to assign predefined labels to each client interaction. The labels are then aggregated for funding and performance reports. Data labeling is critical to communicating performance and funding needs because it easily quantifies, complex and large qualitative case note data. However, while data labeling is a task central to a caseworker's daily duties, it diverges from their core role of providing "care" to the client. This research examines the perspectives of multiple stakeholders engaged with data labels through the lens of fifteen design ideas. These ideas were inspired by caseworker interview insights and an understanding of the caseworker's existing data labeling system. We reviewed perspectives and concerns through speed dating methodology [17], where caseworkers were shown a sequence of storyboards for potential data labeling solutions to elicit their reactions. Our findings encompass the assessment of design ideas across three dimensions: 1) Intrinsic and extrinsic motivation, by aligning data labeling with caseworker objectives and showing its impact on client service provision (4.1), 2) Caseworker acceptance of data collected, by navigating multiple data collection goals including the comprehensive representation of casework (4.2), 3) Usability, through clarity on the data labels and an intuitive labeling interface (4.3). Drawing from our findings, we discuss design implications for the data labeling system to align with the objectives of the organization and the workers.

Our work contributes to the HCI literature that investigates computing practices in nonprofit organizations. Our study supports prior research on challenges in nonprofit data collection, focusing on the domain of homeless care. We propose a set of 15 design ideas to improve data labeling by aligning with the goals and values of both the organization and its workers. We provide design implications that include aligning data labeling with gaining insights for the case and program management decisions, creating a shared control of data collected, enabling the synthesis of qualitative and quantitative data for diverse stakeholders, and improving system usability.

2 RELATED WORK

2.1 Data Collection in Nonprofit Social Work

2.1.1 Prominence of Data in Nonprofits. Data collection is highly prominent in nonprofit organizations. A wide array of studies have noted that nonprofits are under pressure from various stakeholders, including funders, government agencies, and the general public, to showcase data-driven evidence on the performance of their programs [4, 12, 13, 35, 38, 57]. The need to quantify performance has led to an emphasis on the collection of "performance data", which illustrates the effectiveness and efficiency of the organization's work. According to Verschuere, "Effectiveness can be defined as the ratio between the objectives an organization sets and the outcomes that are the result of its efforts, while efficiency can be defined as the ratio between organizational inputs and outputs" [58]. Prior work into the data collection practices of various nonprofits has shown that a focus on performance measurement and evaluation leads to substantial improvements in the outcomes of an organization [35].

Data-driven decisions and evidence collection are seen as necessary solutions for increasing budget requests and appeasing external stakeholders. The amount of data collection needed and the required accuracy, therefore, has also increased. Nonprofits generally collect "financial, client satisfaction, output, and outcome data", [4] where output data refers to services provided by the organization, and outcome data reflects the impact of the work. Output and outcome data collected can include program expenditures, the number of clients served, demographic information, and narrative or anecdotal data.

The increased focus on data-backed performance metrics has led to nonprofits collecting more data for other use cases. On top of evaluating performance, accountability, assessment, and planning, Reamer [44] finds that case documentation is adopted in social work to serve functions including "service delivery, the continuity and coordination of services, and social work supervision." He concludes that the role of documentation has evolved and "social workers have begun to appreciate the relevance of documentation for risk-management purposes, particularly as a tool to protect clients and to protect practitioners in the event of an ethics complaint or lawsuit."

Despite the cited benefits of data collection on organizational outcomes and casework service delivery, through a scoping review, Kuorikoski [33] finds that "documentation has a low status in adult social work and recording practices are inadequate" which, along with lack of time, knowledge, tools, and even deliberate resistance, often results in incomplete or even incorrect data. While the advent of Electronic Information Systems (EIS) has resulted in drastic improvements in the ease of creation, storage, retrieval, and management of data, their rise can result in a shift in focus from service provision to data collection [21, 62] and an over allocation of time and resources to data collection that does not serve an organizational purpose. These systems are optimized for managerial purposes and usually are not structured to the practical needs of social workers who may find the purpose of certain data collection practices to be unclear [40]. Nonprofit data collection systems are perceived to align with the goals of management and external stakeholders and not with the goals of caseworkers.

2.1.2 Performance Evaluation in Social Work Organizations. Performance evaluation is a key area of data use and collection at nonprofits. According to Carman and Fredericks [13], nonprofits view evaluation practices in three ways: external promotional tools, or strategic management tools, resource drains and distractions. These evaluation processes and data analysis are driven by funders and external stakeholders [13, 37]. Performance evaluation data is used to inform strategic planning, improve grant applications, and for marketing to the community [12]. However, several studies find that though organizations may dedicate time and resources to data collection for evaluation, that may not always translate to effective data use.[37, 57].

Prior work has shown that efforts to improve data collection and methods to showcase performance reports are not enough to constitute evaluation and organizational growth, and that data collection does not equate to data use. Nonprofit organizations need a culture of evaluation [39] to encourage not only the effective use of data but also the accurate collection of data.

2.1.3 Survey on How Performance Data is Collected. While intended to improve client outcomes and service delivery, data collection is perceived as an additional burden to caseworkers. The practice of data collection is heavily reliant on caseworkers who interact directly with clients [1, 34, 47, 55]. Caseworkers are the subject matter experts on organizational services and client needs. During or soon after meeting with a client, caseworkers summarize the interaction, input key pieces of information relevant to their service, and other notable details identified by the caseworker or required for reporting [4, 33, 55, 56].

As more data collection and record-keeping become electronic, caseworkers are pushed to collect more data for performance measurement and move towards standardization [18, 24, 56]. Stricter standards of documentation can lead caseworkers "to engineer workarounds and shortcuts" [24, 60].

As part of data collection, caseworkers are also expected to perform data labeling. Data labeling is a qualitative coding process that represents free text case notes as quantitative performance measures. Caseworkers perform data translation work by labeling the raw data collected during client interactions. The labels aid in the translation of social work into outputs and outcomes for performance measurement [47]. Based on specific requirements of funders and other external stakeholders for performance measurement, caseworkers must label more data [26]. This labeling process becomes an additional part of a caseworker's day-to-day responsibilities on top of providing services to their clients [4].

2.1.4 HCI Challenges Discovered for Data Collection in Social Work.

HCI scholars have identified several challenges with data collection at nonprofits, including misalignment with data collection and caseworker goals. Other challenges include the burden of data collection as an additional task and privacy issues related to the sensitive nature of the data collected [2, 4, 6, 38, 51, 60]. These challenges bring to the forefront the tension among caseworkers, nonprofit management, and external stakeholders. A caseworker's main priority is helping their client get the services they need, while management's main priority is communicating impact to external funders, government bodies, and the public. These conflicting priorities create a misalignment in the purpose of data labeling and lead to problems in the data collection, such as inconsistencies and inaccuracies.

Caseworkers who are assigned data collection for performance and funding reports in addition to their service to the client perceive it as a burden. Nonprofit organizations rarely have the resources to hire a data analyst or invest in sophisticated data software. Data collection, therefore, becomes an additional responsibility of caseworkers and other nonprofit staff [20, 60]. Caseworkers sometimes come up with their own methods of data collection, which can vary within an organization, are not interoperable, and require manual upkeep [60].

Overall these challenges lead to what Bopp et al. [7] have identified as a "cycle of data disempowerment" at nonprofit organizations. Bopp et al. [7] argues that the data disempowerment cycle is created out of the desire to make data-driven decisions. Caseworkers become disempowered by the necessity of data collection where what is collected and how it is evaluated is not in their control. Nonprofit organizations, to secure funding and adhere to policy and legal requirements, must collect data. Previous research identifies the challenges of goal misalignment but does not explore potential solutions. Given the need for performance data collection in the existing nonprofit and stakeholder framework, we build upon prior work to improve this data collection by exploring ways to align the data with the values and goals of social workers. By centering caseworker goals and responsibilities, and not just managerial goals, within the data collection process, our work aims to improve data collection for all relevant stakeholders and potentially reduce the impact of the cycle of data disempowerment.

While the challenges arising from this environment are well documented among HCI researchers, there are very few proposed solutions to improving data collection such as labeling for social work. Salvador et al. proposed three potential data collection frameworks for aligning data collection with nonprofit performance and service improvement: the CIT model of civic engagement, the REAP Metrix, and the Dual Capacity-Building Framework for Family-School Partnerships [46]. The three proposed data collection frameworks aim to better align community or client values with nonprofit goals. Our paper expands on this research by placing caseworker and managerial values in alignment through a human-centered co-design approach [46].

2.2 HCI Research in Social Service Work

HCI research with nonprofits further highlights the challenges with data collection. HCI research related to child services and homelessness social work specifically showcases how data collection and use can directly impact client interactions. Many child social service agencies across the USA have implemented data-driven models and solutions to assist social workers in determining risk when making decisions. Several HCI studies have evaluated the effectiveness of these tools and highlighted the miscalibration between the tool and the workflows of caseworkers [11, 15, 23, 47]. A participatory design approach with caseworkers better identifies the "value metrics" that data-driven solutions can improve [23]. Value metrics may focus more on care and human-oriented data instead of quantifying success in ways preferable to funders. For example, funders may evaluate nonprofits by the number of clients served or the number of visits per client, while value metrics consider softer, human-focused ways of evaluating success, such as client confidence or emotional well-being. Most data collection implemented at nonprofits to evaluate performance does not consider the "temporality of risk" that caseworkers must navigate and the procedural nature of their role [47]. Prior work suggests including value metrics in the data collection process may better align caseworker goals with data use and improve caseworker data collection [23].

Homelessness is another area where HCI research has identified alignment with caseworkers as crucial to the successful implementation of data-driven solutions [32, 54]. Caseworkers regularly perform data translation work that is crucial to building trust with their clients and community [54]. Data translation is the work of "translating information from public institutions... ...for their communities" and translating the data gathered from communities into information for public institutions. The effort and time involved in data translation are typically unrecognized by nonprofit management and external stakeholders. Using participatory design methods, such as "comic boarding", caseworkers can be empowered to provide valuable feedback on a data system's design and implementation [32].

2.3 HCI Solutions in Other Domains for Improving Data Collection

Although there is HCI research on improving data collection in different domains, such as crowd work, the values and context largely vary in this space. For example, crowd work is profit-driven, with the labeling itself as the primary responsibility or role, not an additional task as it is for caseworkers [42]. The HCI solutions proposed therefore primarily focus on creating and improving annotation tools [41], creating more efficient learning algorithms [52], or building domain expertise into the model itself [3, 14, 31, 48]; and do not address the unique environment of nonprofit organizations. Solutions such as monetary incentives and gamification for motivation address goals to improve the number of labels recorded rather than improving care.

While motivation-based data labeling is a prominent concept in HCI research, there is limited research on solutions to improve data labeling by social workers for purposes other than their direct client service [53]. HCI research provides motivation-based data labeling techniques within four main categories: games with a purpose [30], gamification display via milestones, rankings, etc. [16], worker compensation [28], and individual performance comparison [25]. However, the most impactful solutions for caseworker data labeling align motivation with improving community and care values [19]. This is distinctly different from the motivation-based labeling techniques from current HCI research[19].

Our study expands the research on motivation for data labeling by identifying needs through semi-structured interviews and using a co-design method to investigate ways to improve data labeling for caseworkers by addressing factors influencing their motivation. Participatory co-design methods can shift the concept of value for a data solution to align with the motivation and values of the caseworkers and managerial stakeholders [23]. We build upon the work of Robinson, which shows that social worker staff welcomes data collection solutions that help them enhance their clients' situations and improve service delivery [45]. By directly addressing caseworker motivation, our research aims to improve the data labeling process for caseworkers and address management's data collection goals.

3 METHODS

This section begins by introducing the study context, providing details about the organization and its primary data collection methods (case notes and data labels). Next, we describe the participants and the study design involving interviews and design feedback sessions.

3.1 Study Context

3.1.1 Organization. We worked with a government-led nonprofit organization that has been serving people experiencing homelessness to achieve long-term stability through an Integrated Case Management (ICM) program over the past two decades in a mid-sized city in the U.S. They provide services aimed at long-term living and housing stability, such as creating housing search plans, counseling services, identifying appropriate programs or treatments for medical health needs, etc. Our prior research interviews [49–51], as part of the bigger engagement with the organization, revealed an ongoing expansion of responsibilities beyond what is strictly necessary for ICM. These responsibilities include participation in emergency response, coordination of transportation for cold weather shelters and protective lodges, and serving as guides to the overall system of services available to people on the homelessness continuum. They also provide short-term, on-demand aid and advice to other homeless individuals as walk-in services. These walk-in services include activities such as holding mail, obtaining bus tickets, assistance with obtaining documents, etc. Walk-in clients are served based on the immediate needs of the client and have a limited meeting time of about 30 minutes, each caseworker deals with as many as 50-60 clients per day. Whereas, ICM clients have dedicated caseworkers with a case management plan.

The organization is composed of about 30 clinically trained caseworkers dealing with ICM and walk-in clients. The caseworkers are overseen by two managers, who supervise caseworkers, oversee program management, and are also responsible for bringing in funding and performance reports. Lastly, two program analysts have been appointed for a holistic investigation of the current program and to build strategies that enhance case management practices and their data systems. Although the organization is a government service, its functions are essentially nonprofit social work. Hence, they need to periodically justify the impact of their work and resource requirements to sustain their services and secure appropriate funding [4, 12, 13, 35, 38, 57]. Prior work shows that being data-driven is essential for nonprofits (2.1). This organization shows a real-world example of the impact of data-driven processes and the challenges in recording and processing data in their effort to promote data-driven decision-making. Additionally, they were open to collaborating with researchers and sharing case data. The caseworkers at the organization primarily collect data in two forms: they write detailed free-text notes that describe the caseworker's interaction with that client and assign labels from a predefined list of data labels to represent the outputs and outcomes of the interaction for funding and performance reports. Caseworkers write the case notes during or after a client interaction and then assign relevant data labels.

3.1.2 Case Notes. Case notes are free-text records of caseworker and client-related interactions, written during or after such interactions. They assist caseworkers in tracking clients' case histories, and informing their decisions on the next steps for client service. Caseworkers consult past case notes before or during client meetings to access pertinent information such as contact details, service requests, and pending applications for housing programs.

The dataset shared by the organization contains case notes that span from October 2016 to September 2022. There are a total of 63,485 case notes across 1,691 clients written by 30 caseworkers. These notes vary in length, ranging from a few words to over 130 sentences. The content includes contemporary functions of documentation such as assessment and planning, service delivery, and continuity and coordination of services [44]. They chronicle events related to clients, encompassing not only direct interactions between clients and caseworkers but also other interactions such as email exchanges with service providers, records of received mail, phone calls, etc. Both Intensive Case Management (ICM) and walk-in clients have case notes, though the nature of these notes may slightly differ. Walk-in clients' case notes often focus on immediate actions and services rendered, while ICM clients' notes may cover longer-term plans and more personal details. We conducted a thematic analysis [10] of case notes to understand the data collected and their implications for case management practices. Case notes content can be represented in six major themes: 1) action items such as updating an application, renewal, or housing options; 2) client status updates such as job, application fill up, or caseworker task updates such as received mail; 3) requests for pass/cards/services; 4) scheduling and tracking meetings or appointments; 5) emotional/general conversation snippets; 6) potential next steps.

3.1.3 Data labels. Caseworkers also assign data labels from a predefined list of labels to describe the outputs or outcomes [36] of a particular meeting with a client. Each case note or client interaction can have multiple labels associated with it. The IT department returns the totals of each label at the end of the month which is uploaded to the Homeless Management Information System (HMIS), and used for performance and funding reports. HMIS is a local information technology system used to collect client-level data and data on the provision of housing and services to homeless individuals in each Continuum of Care (CoC). Caseworkers do data labeling after writing case notes, usually post-client interactions. Data labeling is crucially important to the organization's funding and success as it directly informs performance and funding reports. The data labels are the foundation for these reports for external stakeholders and represent qualitative data from case notes as measurable quantities for evaluation. The importance and challenges involved with the data labeling process were emphasized by the organization's management and by caseworkers throughout our findings.

There are two primary types of labels, "Contact Type" and "Interventions". "Contact Type" labels refer to the mode of client interaction, such as 'direct contact' with the client in or out of the office, or 'collateral contact' for interactions with other organizations for client-related tasks. The "Interventions" labels refer to outputs and outcomes, such as providing a referral for treatment or service, successful housing, completing client assessments, and providing bus or food passes. We analyzed data from October 2016 to September 2022. There are a total of 113 labels, 28 (24.8%) are "Contact Type" labels, whereas 85 (75.2%) are "Intervention" labels. In the given period, caseworkers assigned 108,704 labels across 63,485 notes for a total of 1691 clients, bringing the average number of labels assigned per note to 1.688. Of the 108,704 labels collected, 100,784 (92.7%) are "Contact Type" labels, and 7920 (7.3%) are "Interventions" labels. The distribution for contact type labels is heavily skewed. The top 4 labels, "Direct Contact" (38.55%), "Collateral Contact" (24.57%), "Client Contact out of office" (12.58%), and "Client contact in office" (11.36%) make up just over 87% of all contact type labels. Moreover, our interview findings showed that several duplicate labels are often chosen interchangeably, and apart from "Collateral Contact," all of the remaining labels in the top 4 are used to refer to the same type of contact. In the current system, the "Interventions" data is arranged in a hierarchy. There are five total top-level interventions such as housing, income, medical, and 26 specific program collaborations, and their 17 outcomes such as accepted, denied, and declined. This data is similarly skewed; the "Income" category accounts for 76% of all interventions labeled, while there is significantly low data on other interventions performed.

3.2 Participants

We recruited five caseworkers who collected the data, two managers who used the data for reports, and two program analysts who assessed the data systems. These stakeholders pursuing different objectives offered unique viewpoints. We crafted the recruiting emails, which the organization collaborator distributed among their employees for the voluntary sign-up. The participants were chosen from volunteers. Our participants varied in terms of their tenure in the organization and education. Table 1 shows the aggregated demographics of the participants for different stakeholder groups to prevent the identification of individual workers. All five caseworkers and one manager were interviewed. Then, four caseworkers and the remaining stakeholders joined for the design feedback session. Participants took part in the design feedback session based on availability.

3.3 Study Design

The study was performed using a combination of in-depth interviews and design feedback sessions involving multiple stakeholders to understand the organization's current challenges, opportunities, and approach to data labeling.

3.3.1 Interviews. To understand the challenges and needs of the data labeling system at the organization, we conducted 30-minute semi-structured interviews with caseworkers. Our prior collaboration with the organization established familiarity and trust in our research team for the participants. The interviews focused on topics such as the purpose, utility, and quality of existing data labels, and the process of labeling. We explored the role of data labels in day-to-day case management, potential motivating factors for labeling, and caseworkers' reflections on their labeling practices. We analyzed the notes and transcripts from Zoom following [43]'s qualitative data analysis method. The emerging themes were grouped to note the benefits and limitations that participants perceived in the current data labeling design and ideas that they shared on how data labeling may be improved.

3.3.2 Design Feedback Session: Speed Dating. Our research team brainstormed 15 design opportunities based on specific needs and scenarios described by the caseworkers in the interviews to improve the data labeling system. Some ideas were rooted in participant insights, while some were inspired by literature in motivation [19] and our prior analysis of case notes and existing data labels (sections 3.1.2 and 3.1.3). The findings section presents details on interview insights and specific literature that led to each design idea. Team members reviewed the collected needs, generated ideas, and discussed them to improve and finalize. We utilized the initial phase of speed dating for needs validation by presenting these design ideas to the users through a series of storyboards. This approach allowed us to synchronize the design opportunities we found with the needs users perceived [17]. Consequently, we gained an understanding of where the observed and perceived needs of caseworkers for data labeling aligned, providing deeper insights into the interviewees' perspectives (Table 2), needs, and concerns. Placing participants in

Table 1: Participant Demographics and Interview Attendance

Group	Role	Count	Gender		Race		Age Range		ID	Interview	Speed Dating
							25-34	1	C1	Y	Y
	Clinical professionals		Female	4			35-44	2	C2	Y	Y
Caseworkers	who manage individual	5			White	4	45-54	1	C3	Y	Y
	cases of homelessness		Non Binary	1			65-74	1	C4	Y	Y
									C5	Y	N
Program Analysts	Investigates current data systems						35-44	1	P1	N	Y
	at the organization, and	2	Female	2	White	2					
	build or integrate technology tools						Undisclosed	1	P2	N	Y
	Supervises the case management		Female	1			35-44	1	M1	Y	Y
Managers	program, responsible to bring	2			White	2					
	in funding and resources		Male	1			55-64	1	M2	N	Y

familiar scenarios with new interventions representing potential future scenarios [63] facilitated a deeper exploration of true needs and investigation of the challenges and feasibility of the proposed ideas.

We comprehensively assessed various stakeholder viewpoints to consider the organization's needs. Each session had 1-2 participants, depending on availability. Each session took about 90 minutes. During the study session, we presented the design ideas through individual and group activities. First, participants individually reviewed the ideas and recorded their initial impressions on a notes sheet which allowed them to familiarize themselves with the ideas and form opinions. Subsequently, we presented each design idea to the participant(s) with a brief description and initiated a discussion to assess the designs and understand perspectives. We followed up on their initial impressions, discussed alignment with their daily needs, investigated challenges or concerns, and explored potential improvements.

In line with Zimmerman and Forlizzi [63], digital storyboards were used to represent the design concepts allowing rapid visualization of the possible futures. Each storyboard consists of four panels, where stick-figure characters walk the reader through the possible future by showcasing the context, need, application, and result or impact of the idea (figure 1). Depending on the idea, the storyboards were positioned from different perspectives, such as seasoned or new caseworkers or the manager. The storyboards were reviewed by the research team and piloted by an external member to ensure clarity and consistency in the representation. The final storyboards can be found in the supplementary materials.

We conducted and recorded the study on Zoom, with the participant's consent. Each session was facilitated by 2-3 researchers, including a moderator, a note-taker, and an observer. In total, we held five sessions, with paired sessions for program analysts (P1, P2) and two caseworkers (C1, C2), and individual sessions for the remaining participants (C3, C4, M1, M2). Caseworkers were shown all the ideas, while certain ideas were skipped for other stakeholders based on relevance to their responsibilities. The notes and transcripts from Zoom were analyzed following Patton [43]'s qualitative data analysis method. The first four authors initially analyzed participants' feedback on each idea with a focus on what resonated with them and what they found more or less useful and created thematic groups across the ideas. They then discussed the findings with the entire research team through a weekly meeting and derived final insights.

3.4 Researcher Stance

Our research team included people with diverse backgrounds in human-computer interaction, artificial intelligence, and communications. We sought to bring about positive change by exploring design interventions. The exploratory nature of our study was communicated throughout the collaboration with the organization. Several of the researchers have conducted research with the organization previously. This relationship and familiarity helped us gain access to the research site.

4 FINDINGS

In this section, we describe our research findings in three major themes: aligning data collection with case management goals, comprehensive representation of caseworkers' work in the data labels, and usability of data labels and data labeling. Each section first describes the caseworkers' issues with data labeling, then introduces the proposed ideas, and details the perspectives of the caseworkers, managers, and program analysts.

4.1 Aligning Data Collection with Case Management Goals

Caseworkers consider data labeling extra work because they do not perceive its connection to improving client services. Caseworkers understand that data labels serve performance and funding purposes, but consistently expressed that they do not perceive the value of labeling. For example, caseworkers explained, "It is extra work"(C1), "It is not useful or part of my case management"(C4), and "not sure how exactly it is useful, but we were told to do this"(C2). Consequently, it is perceived as a data-focused activity rather than a client-centered activity, unlike case note writing that directly informs caseworkers' decisions for the clients. Caseworkers, who are driven by a sense of care (C2) for their clients, lack the motivation to engage in tasks that do not directly contribute to client service and rarely participate in labeling. As noted by a manager, caseworkers expressed, "Data input isn't hard, it is just not fun. You know, so it is hard to motivate yourself to do that so you can justify it in your mind like, oh, that wasn't a very major interaction, so I'm not going

Table 2: Idea Usefulness by Participant. This table presents the list of design ideas generated by the research team, along with participant preferences color-coded into four categories and 'not shown'.

Idea	Design Idea	C1	C2	C3	C4	P1	P2	M1	M2
1	Filtering case notes based on data labels to facilitate targeted search	Useful	Useful	Useful	Useful	Useful	Conditionally Useful	Useful	Useful
2	Client analytics dashboard that leverages data labels to showcase trend	Useful	Useful	Useful	Conditionally Useful	Useful	Useful	Useful	Useful
3	Dashboard with information on the periodic impact of data labels	Conditionally Useful	Useful	Skeptical	Not useful	Useful	Conditionally Useful	Useful	Useful
4	Redesigning the data labels with increased granularity	Useful	Useful	Useful	Useful	Useful	Skeptical	Useful	Conditionally Useful
5	Standardizing case notes to access contextual information underlying the labels	Conditionally Useful	Skeptical	Useful	Skeptical	Skeptical	Conditionally Useful	Useful	Useful
6	Streamline the addition and subtraction of data labels	Useful	Useful	Useful	Useful	Conditionally Useful	Useful	Useful	Useful
7	Instant access to the data label definitions and examples of labeling	Useful	Useful	Useful	Useful	Useful	Useful	Useful	Useful
8	Periodic training sessions on data labeling	Useful	Useful	Useful	Useful	Useful	Useful	Useful	Useful
9	AI tool to identify redundant data labels	Useful	Useful	Useful	Useful	Useful	Conditionally Useful	Useful	Useful
10	Search feature to find specific labels	Useful	Useful	Useful	Useful	Not shown	Not shown	Not shown	Not shown
11	Visual feedback to navigate through the data labels	Conditionally Useful	Conditionally Useful	Useful	Useful	Not shown	Not shown	Not shown	Not shown
12	AI tool that analyzes current case note content and suggests data labels	Conditionally Useful	Conditionally Useful	Useful	Useful	Useful	Skeptical	Useful	Conditionally Useful
13	Presenting client's most frequent and past interaction data labels	Useful	Useful	Useful	Useful	Not shown	Not shown	Not shown	Not shown
14	Reminders on data labeling objectives	Not useful	Not useful	Not useful	Useful	Not shown	Not shown	Not shown	Not shown
15	Displaying labels that have been rarely or never recorded	Useful	Useful	Conditionally Useful	Useful	Conditionally Useful	Useful	Useful	Useful

Not useful – Caseworker stated explicitly the idea was not useful for data labeling and their casework
Skeptical – Caseworkers did not explicitly state the idea was not useful but were skeptical of its use
Conditionally Useful – Caseworker stated the idea could be useful for certain use cases or client contexts only
Useful – Caseworkers stated explicitly the idea was useful for data labeling and their casework
Not shown - These ideas were not shown to specific participants due to relevance

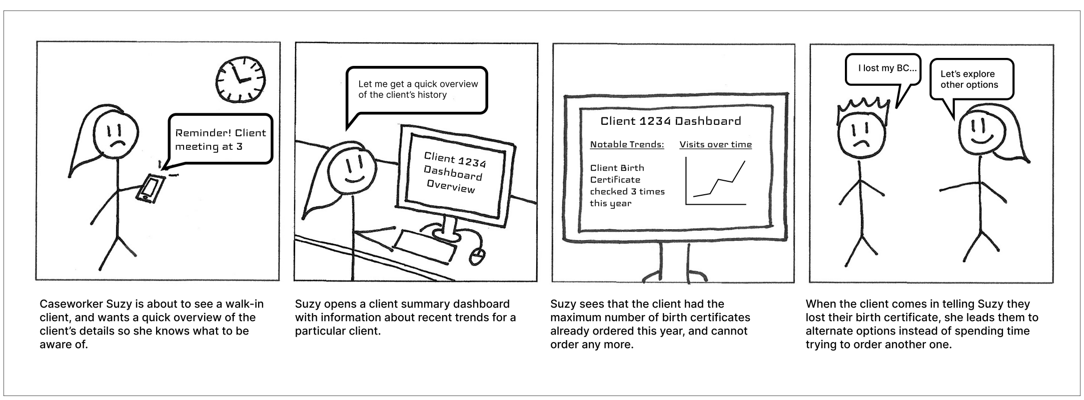

Figure 1: Example storyboard: client analytics dashboard that leverages data labels to showcase trends (Idea 2). This idea seeks to align data labeling with caseworkers' goal of understanding client case context to provide more effective assistance.

to, you know, spend time logging into the system and logging that interaction because it is going to take more time for me to log this information than the actual conversation" (M2).

Hence, we observed the perceived disconnect between labeling data and providing care to significantly diminish caseworker motivation, affecting data labeling. Drawing from intrinsic and extrinsic motivation, we generated design ideas. We aimed to enhance extrinsic motivation [25, 28] by directly bringing the value of data labeling to caseworkers' daily tasks(4.1.1) and improve intrinsic motivation[19] by increasing awareness about the value of data labeling(4.1.2) as explained in the following subsections.

4.1.1 *Find Ways to Make Labeling Useful to Caseworkers.*

Our investigation aimed to explore if labels can be utilized by caseworkers in their daily activities to address client needs and, in turn, motivate labeling. Primarily, we ideated ways to enhance identifying specific information from case notes that characterize the clients' case, such as previous housing options explored. This information could then inform the caseworkers' subsequent actions, such as recommending potential housing options or following up on past ones.

Our focus on improving information retrieval from case notes is derived from our case note analysis (see section 3.1.2) and caseworker interviews that revealed inefficiencies in the current process. The inefficiencies can be attributed to various factors, including the nuances introduced by free-text case notes, such as significant variations in terminology and abbreviations to convey similar information. For example, terms like "Detox program" and "A****** house program" refer to the same service, while "BC," "Birth Certificate," and "Identity card" are used interchangeably. Furthermore, the absence of specific keywords in documenting certain services results in the loss of valuable information during the search. To cope with these challenges, caseworkers resort to manual and iterative keyword-based searches through free-text notes, which is time-consuming (C1, C4). Manual searching is also prone to errors, leading to instances where caseworkers fail to adequately retrieve previous options explored, hindering their ability to re-evaluate housing or treatment choices effectively.

Our team developed two ideas to align labeling with the caseworkers' goal of efficient retrieval of past information on the client's case. These involve providing mechanisms that enable caseworkers to analyze case notes more effectively using labels.

First, we proposed **filtering case notes based on data labels to facilitate targeted search (Idea 1)** to enable caseworkers to effectively find relevant information and make informed decisions. For instance, When deciding which housing options to explore next, caseworkers can filter case notes by choosing all data labels associated with housing applications to identify case notes with detailed insights on outcomes of past housing options. This aligns data labeling directly with the caseworkers' goal of better serving their clients.

Second, we proposed a **client analytics dashboard that leverages data labels to showcase trends (Idea 2)**, such as the number of requests and prior outcomes of attempts to receive various services such as housing and mental health treatment. The purpose of this dashboard is to enable caseworkers to identify unique client trends, which will aid in devising personalized strategies to help clients achieve their goals. For instance, a significant number of clients aim to register for mental health programs as they play a crucial role in enhancing overall living stability and the ability to cope with the challenges of housing search. However, the clients' history, such as criminal background, intensity of substance abuse, or violent behaviors, may hinder their acceptance into these programs. By analyzing trends, such as repeated denials for specific programs, and frequent access to substance abuse treatments, caseworkers can gain valuable insights to reassess their strategies and tailor their efforts to better address individual client needs and circumstances.

Participants in the speed-dating session strongly preferred both of these ideas to address the current inefficiencies in retrieving client case information and trends. They were also perceived to benefit managers by facilitating easy and accurate identification of client case characteristics. A caseworker noted that "If these data labels function better, then there would be more of a push, I think, to record things while the client is there" (C3). As caseworkers discussed additional use cases, they highlighted automated information retrieval for efficient management of client records, such as ensuring meeting specific ID requirements for housing applications and other services. One caseworker highlighted this need, saying, "There's a limit to how many birth certificates we can order for a client within a year if we could, and as of now, like, I'll control F search for information. But if we could just like, see how many times we have gotten this birth certificate for this client. You can really see the pattern" (C2). Caseworkers also emphasized how visual representation through a dashboard could quickly highlight trends for walk-in services where there is less time to go through past interactions with the client. For example, identifying clients who repeatedly request the same vital documents due to misplacement allows caseworkers to devise tailored solutions, such as having multiple copies ready to save time and enhance the effectiveness of their assistance (C1, C2, C3). Both program analysts and managers underscored the dashboard's potential for swiftly analyzing client behaviors and delivering valuable insights, particularly for new caseworkers who often face information gaps on the client in the initial stages.

Program analysts further highlighted the need to implement these solutions to benefit all stakeholders. They stated the dashboard features should facilitate the needs of both caseworkers and managers to avoid any double work, providing an example of IDs ordered attribute to be aggregated not just on the client level to know the trend but also on an organization level to know the funding needs. "I also want to make sure that a manager is able to say, 'How many birth certificates are we getting in total this year because that's going to change our funding need to ask for money there. I want to make sure that it works for everyone, and they're not doing double work'" (P2). Managers echoed this sentiment and emphasized the value of an advanced and adaptable search system that enables caseworkers and managers to identify relevant information based on their needs (M2).

4.1.2 *Increase Awareness of the Value of Data Labeling:*

From an organizational perspective, a constant demonstration of outcomes and service provision (outputs) is essential for securing resources to run efficient services for the client. Aggregated data labels demonstrate the specific outcomes and funding needs. However, caseworkers indicate a lack of this knowledge and emphasize the need to understand the purpose and rationale behind their efforts in labeling. As noted by caseworkers, "What we need to do and why we we're doing", the answers to these questions. Knowing that it translates into something bigger can help motivate recording the data labels"(C1) and “Realizing that our data labeling has a direct influence on our capabilities, the two are linked”(C5).

To enable an understanding of how data labeling can translate to better services to the clients, we proposed a **dashboard with information on the periodic impact of data labels (Idea 3)**. The dashboard portrays the connection of data labeling to the value provided by presenting a periodic aggregate of labels and the corresponding effects on the organization's funding and recognition by other entities. This can be achieved by presenting number of clients referred to the organization over time. For instance, consider displaying the total count of state IDs the organization has successfully obtained over time, and the subsequent increase in the number of clients referred to the organization for State IDs. This demonstrates that caseworkers can showcase their proficiency in efficiently processing State IDs by consistently recording IDs issued, a task often challenging due to the need for collaboration with multiple departments. The recognition can prompt other organizations to refer more clients, expanding access to suitable services for a larger group of people, and in turn, enhances program efficiency as caseworkers can focus on services in which they excel.

All participants agreed with the need to demonstrate the value of data labeling. However, caseworkers had mixed opinions on using dashboards as a form of communication. While one expressed curiosity to learn about the outcomes through dashboards (C2), others either preferred dashboards for predicting trends and allocating resources effectively (C3) or were not sure if they would actively refer to a dashboard to verify impact (C4). However, they emphasized the need for transparency (C1, C4) and sought this information through staff training on data labels (Idea 8), “I know I personally value transparency. I want to know what you know. Why my work, like, why, these things are important, and you know, and of course, I have a general idea of why. When we're actually able to tie real outcomes, too. Okay, this. This led to x amount of funding, or we got, you know, a,b,c,d,e, from this, and it also informs us honestly, like where we need to focus more and less as well. It is another way to figure out gaps in services” (C4). The same caseworker highlighted periodic training, “This is an opportunity for folks to see like, Oh, well, this is the impact we're having, like when we're doing this. And this is how it is directly tied. I think, the more tied to this data people are, or the more invested in it, the more likely they're going to utilize it” (C4). Program analysts, however, highly believed the dashboard was the most efficient in conveying the impact to caseworkers.

Managers proposed that a dashboard could also be quite useful for them to gain insights into the client case progress and, in turn, the caseworkers' current workload. For instance, a manager explained how knowing the trends and current client case status can help allocate caseloads - “You know how many clients are currently unhoused, because we know those who are unhoused typically are much more high need, particularly with the conditions that exist in our community around housing and affordable housing. There's a lot of effort that has to be put in by their caseworkers to identify housing. You're gonna get a lot of doors shut your face like no, no options available, no options available. And then, once you have someone that can help, the load reduces. So kind of, you know, gauge-like, where's someone's level of effort? - needs to be evaluated of how that can be adjusted to help create more equilibrium" (M2).

4.2 Comprehensive Representation of Caseworkers' Work in the Data Labels

The caseworkers are dissatisfied with the current data labels as they consider them inadequate in representing their work. Since the data label aggregates are used by the City and other collaborators for the organization's performance evaluation, it is crucial to accurately capture all casework for appropriate recognition. This recognition leads collaborators to refer more clients, improving access to services for individuals in need. Moreover, it allows caseworkers to concentrate on services they excel at, ultimately enhancing the program's efficiency.

Current data labels encompass measures required by funding agencies, such as counts of IDs processed to accommodate application fees and in-office client visits to facilitate office utilities. Data labels also include measures for standard performance reports by the City, such as counts of clients successfully housed, clients contacted, and mental health treatments obtained. However, caseworkers and managers stated that they fail to efficiently capture other crucial aspects of casework that impact clients' housing stability and well-being. Data labels only portray end results such as housed or income acquired, overlooking case complexities, leading to a perception of low organizational performance (M1). For instance, securing stability for clients with behavioral issues demands extra effort to ensure task completion and maintain progress.

Similarly, the scale of coordination required with multiple entities is not considered, “I might deal with four different agencies in 30 min and send referrals and continue care stuff, all of that work get lost as far as getting captured"(C4). Moreover, data labels do not capture other casework outputs such as obtaining identity documents, these are crucial for job applications or accessing public resources for stable living. "I asked why we only captured the two, birth certificates, and State IDs. I know we don't pay for social security cards. But we're spending a ton of time ordering those" (C1). By using these drawbacks identified in interviews regarding the current labels, we created design ideas. These ideas focused on capturing complexities(4.2.1) and sharing the changing relevance of labels(4.2.2) with management, as explained in the following subsections. We aimed to address the perceived dissatisfaction of caseworkers with labels representing their work, to improve data labeling.

4.2.1 Enable Capturing of Casework Complexities. Caseworkers are concerned that data labels reduce their work to numbers that do not represent the complexities of client cases they manage (C1, C4). For instance, they explain how capturing just the total number of housed or unhoused clients fails to consider the varying amount of work completed for them. Clients with criminal histories often face limited housing options, while clients with severe substance abuse issues require treatments to qualify for housing applications. "One case took two years of work and 40 housing applications while the other took 3 months of work and two applications - if you didn't have the steps caught accurately, then it will just be a number, this person is housed, and that person not" (C4).

We proposed **redesigning the data labels with increased granularity (Idea 4)** to provide a more accurate representation of the casework by capturing the caseworkers' efforts that indicate the case complexities. For instance, adding a data label to record the counts of people or entities contacted for a service can capture the scale of coordination needed. "If they talk to three different agencies in one interaction with the client, then that should count as 3 collateral contact instances, not just one. A recent extremely medical fragile client I had, involved me all week communicating with his emergency room doctor, a nurse at the emergency room, his partner, his insurance company, and potential nursing homes" (C4). Additionally, data labels that record the number of attempts made for housing before success, and mental health treatments sought before acceptance, can also account for the complexities of the casework.

Overall all the participants agreed with this idea. Caseworkers strongly resonated with the need to capture their casework efforts. They also considered granular labels to enhance the utility of filtering through case notes (Idea 1) and client analytic dashboards (Idea 2) with more precision. For instance, with an additional filter to distinguish between successful and unsuccessful housing applications, caseworkers can quickly identify areas of improvement from past applications. Managers found granular labels valuable for conveying the organization's challenges to funders, the City, and collaborators, by enabling them to build a narrative backed by quantitative evidence on the extent of denials and efforts made to succeed in client goals (M1, M2). However, there were concerns about misrepresenting certain efforts captured without evaluating the underlying reasons behind the numbers (P2, M2). For instance, directly interpreting the total number of denials or total time spent on housing applications as measures of higher complexity, "an application can also be denied for logistic errors like missing the submission of a required document" (P2).

In addition to granular labels to demonstrate complexities, managers and program analysts also asserted the need to reform certain labels into broader categories. This is to keep the labels concise and simplify the labeling process for caseworkers. "How granular, you know, do we go? Are we losing any efficiencies by having ten options... Where's the sweet spot?" (M2). For instance, they suggested combining different food coupons provided into a single label, as only the aggregate amount of food coupon requests is sufficient to allocate the budget for food support (P2). Similarly, all non-funded IDs could be combined into a single category called "IDs" instead of a list (M2).

Besides investigating the quantification of caseworker efforts to capture case complexities, we also explored enhancing access to qualitative case note information for managers to identify case complexities. We proposed **standardizing case notes to access contextual information underlying the labels (Idea 5)**, making detailed case information easily available to others. With better access to qualitative case note data, managers can incorporate the reasons behind the aggregated data label numbers to represent case complexities.

All participants desired case note content to be accessible for comprehensive communication of casework for performance and funding reports. However, caseworkers raised concerns that over-restricting case note writing, generally performed during a client interaction, could distract their engagement with the client making it more data-centered. They emphasized the importance of maximizing time and attention to clients during interactions and proposed implementing overarching guidelines with flexibility. Additionally, a program analyst considered training as a more effective method for standardization over guidelines (P1), while the other asserted the usefulness would depend on designed guidelines (P2). Caseworkers stressed that standardized case note guidelines would help new caseworkers grasp what information to include to effectively inform their next steps in serving the clients. Caseworkers from diverse backgrounds have varied note-writing approaches, resulting in differences in the level of detail and types of information in their notes. For instance, one caseworker initially focused heavily on noting specific services provided to the client but later recognized the importance of recording their assessment of the client's state in their notes, such as the client's attitude and mental state. "I leaned in really hard on the very specific services I had provided, and really just including that, and I realized I was actually not seeing folks completely as people for a period of time. Not that I didn't care about them, and I wasn't trying to help them, but I wasn't looking beyond processing a referral to see that you know they didn't really seem to be doing too well" (C3).

Furthermore, managers believed that having access to detailed casework information would help identify program inefficiencies by providing contexts, such as for delays in treatments or extended periods of homelessness. "It would be so much more efficient for those who are coming in and trying to get a snap of what's going on with this call" (M1).

4.2.2 Provide Ways for Shared Creation of Data Labels by Management and Workers. Periodic assessment of data labels is essential to ensure they comprehensively capture changes in casework. Caseworkers' work may evolve over time due to new client needs, such as voter registration, or organizational changes, such as new collaborations on new treatment programs or housing services. Without specific labels representing these tasks, the work done would go unnoticed during performance evaluation. For example, caseworkers highlighted the need for additional data labels for obtaining social security cards, driver's licenses, and voter registrations. These are frequently handled tasks demonstrating their workload and expertise. Caseworkers noted, "These data labels will show that we're working on them often, and how skilled we are in it" (C5), "Oh, my God, I'm waiting on like a 1,000 social security cards to come in right now. Both walk-in clients and my own case managed clients. And yet there is not a social security card drop-down" (C3). Only some caseworkers approached management on adding a social security card label, while other labels were never brought to management's attention, indicating a lack of communication on updates to data labels.

To ensure the data labels are capturing current casework, we proposed to **streamline the addition and subtraction of data labels (Idea 6)** through a digital channel. This allows caseworkers to propose new labels, which managers can review to update the labels list. Managers can also gather quick feedback on data label changes from caseworkers to aid their decisions.

All participants unanimously supported streamlining data label creation. As one caseworker noted, "When a new need comes up and caseworkers are putting so much effort into it, being able to add immediately would be awesome" (C2). Program analysts encouraged the idea but were uncertain if caseworkers would actively suggest labels, indicating the need for testing. Managers underscored streamlining to also help preserve knowledge on past data labels, such as their definitions, intended purpose, and reasons for removal. This information is considered essential to compare aggregated values and assess differences over time (M2).

4.3 Usability of Data Labels and Data Labeling

Caseworkers highlighted two main usability concerns with data labeling. Firstly, they found the process of identifying relevant labels from a long list time-consuming, hampering client interaction time. Secondly, label meanings were considered ambiguous resulting in uncertainty when choosing them for client interactions, leading to inconsistency and abandonment of labeling (C3, C5). For example, the "birth certificate" label was associated with multiple instances such as when it was ordered, denied, and successfully received. This caused multiple entries for a single order, leading to inaccurate funding calculations for their application fees. We aimed to address the usability concerns hindering caseworker's data labeling and its accuracy. Drawing from the major pain points identified from interviews on labeling efficiency(4.3.1) and label clarity(4.3.2), we generated design ideas as explained in the following subsections.

4.3.1 Tools to Efficiently Identify Relevant Labels: Manual omissions in labeling were found to be commonplace. These omissions are attributed to caseworkers lacking sufficient time to assess each label for relevance (C4, M1) and remember everything that should be recorded. One caseworker stated, "There are instances you feel like you have missed out on data labels that you want to record, but it is either too time-consuming to go through all of them, or you forget in that instant" (C1). When client interactions get lengthy or highly active, caseworkers often struggle to remember and select all the relevant data labels corresponding to the performed activities. To efficiently identify relevant labels, we proposed four ideas.

First, an **AI tool that analyzes current case note content and suggests appropriate labels (Idea 12)** to guide caseworkers on labeling. Caseworkers commended the utility of AI recommendations. However, they stressed the importance of autonomy in choosing the final labels. They also opposed interrupting features such as continuous reminders or pop-ups that could impede their interaction with the client. Managers and program analysts encouraged this idea but cautioned against caseworkers' over-reliance on AI recommendations which are prone to inaccuracies (M1, P2).

Second, displaying **clients' most frequent and last interaction data labels (Idea 13)**. This aimed to facilitate caseworkers to quickly assign relevant labels from past interactions. Caseworkers acknowledged the utility and believed it could reveal important client case characteristics. For instance, a frequent "no show" label for a client indicates that the client frequently misses appointments, prompting the caseworker to provide additional reminders. Additionally, labels from the previous interactions could help review recently explored options and identify potential next steps, "Hey? Let's, you know, go in this direction, or it looks like this has been tried. But how do you feel about going this other way with" (C4).

Third, **displaying labels that have been rarely or never assigned(Idea 15)**. Caseworkers intuitively prioritize assigning some labels over others. Showing rarely assigned labels could encourage reviewing all the labels for relevance. Overall, this idea elicited mixed responses among caseworkers. Some appreciated that it increases awareness of underutilized labels (C1, C4), while there were concerns that exposing the low-assigned labels per caseworker to everyone could induce performance pressure for specific individuals. Managers recognized the idea's benefit in addressing caseworkers' lack of awareness of specific existing labels. "I think that solves the issue of staff members not being aware that certain items may exist, and I think over time staff develop a blind spot" (M2).

Fourth, providing **reminders on data labeling objectives (Idea 14)**, with an option to edit assigned labels to promote accuracy and completeness. Caseworkers had a neutral stance on this idea in influencing their labeling but accepted it as long as it didn't impede their work. They preferred having a clear understanding of data labels (4.3.2) and their utility (4.1.2).

In addition to efficiently identifying relevant labels, we evaluated ideas addressing specific navigation issues in their interface. Data labeling significantly reduced when caseworkers transitioned to a new system. This decline was attributed to the new interface that groups data labels into drop-downs requiring multiple clicks to browse through the labels, in contrast to the old system's easier browsing through a single-page labels list. We presented two ideas to facilitate easier navigation, a **search feature to find specific labels (Idea 10)**, reducing the need to browse through lists. And **visual feedback to navigate through the data labels (Idea 11)**, such as presenting assigned and unassigned labels separately, to enable reviewing and ensure completeness.

4.3.2 Clarify Label Meanings: Caseworkers expressed the need for clearer data label definitions to judge relevance. There are also numerous redundant data labels with overlapping meanings in the existing list, a consequence of a lack of cross-checking between new and old labels (C1, C4). For example, four labels in the current list represent a similar client-caseworker contact type during an interaction. The "client contact" label, which represents any contact with the client subsumes "direct contact," which is in-person contact with the client. Further, the "direct contact" includes "direct contact in office" and "direct contact out of office." Caseworkers stressed the need to remove redundant labels to improve usability. "I would like the system to be more streamlined, with redundant and useless options removed. I don't want to go through a list of many labels trying to figure out which one is appropriate" (C4). Hence, we proposed three ideas to improve clarity in data labels.

First, an **AI tool to identify redundant data labels (Idea 9)** to enhance the process of identifying and refining the labels. The AI tool will regularly analyze all case notes and corresponding labels assigned to identify potential redundancies. It does so by analyzing patterns, such as different labels assigned for similar case note content.

The second is to provide one-click access to data label definitions with appropriate examples within the interface where the labels are assigned. This **instant access to the data label definitions and examples of labeling (Idea 7)** could foster a common understanding among the caseworkers.

The third idea is based on the caseworkers' suggestion to provide **periodic training sessions on data labeling (Idea 8)** to caseworkers. These training sessions serve as opportunities to clarify ambiguous labels and ensure that caseworkers have a better understanding of labeling.

In general, all participants strongly preferred these ideas to enhance caseworkers' understanding and accuracy in data labeling. A manager noted that removing redundancy is one of their current goals (M2). Discussion around using AI for this purpose was, however, limited, likely due to unfamiliarity with AI's functionality to identify redundant labels. However, one caseworker with a technical background expressed enthusiasm citing the example of an AI tool that suggests combining music albums based on their content similarity (C3). Caseworkers considered definitions as valuable tools for newer caseworkers who may often feel uncertain about appropriate labels. As one caseworker stated, "I'm killing it on labeling. You know the right data labels. And then I just don't know. I hit a slump, where I'm like say 'direct contact', and that's all I get. And then I start feeling confused about what I should be labeling" (C3). Managers considered access to definitions to empower caseworkers by reducing reliance on management for clarification (M2).

5 DISCUSSION

Overall our study serves to identify solutions to enhance data labeling for social service case notes for performance and funding reports. We investigated the perspectives of caseworkers, managers, and program analysts using a set of fifteen design ideas inspired by caseworkers' interviews and an understanding of their current system. Our qualitative approach enables a rich understanding of the challenges and motivations of data collection within nonprofit organizations, particularly in the context of casework. Our design implications have the potential to apply to other nonprofits. Our findings suggest ways to address the challenges presented in prior literature and the implications for designing effective data labeling. We discuss these implications in the following four themes.

5.1 Designing Tools that Utilize Labeled Data to Address Caseworker's Information Needs

Our study underscores the value of crafting assistive tools that align data collected for performance and funding needs with the informational needs of caseworkers during client interactions. We found that caseworkers commonly lack motivation for labeling data, which is consistent with the prior literature [4, 6, 22, 33, 54], and as such, they often fail to label data [33, 40]. Our findings suggest that this alignment could bolster caseworkers' motivation for data labeling. As captured in the words of a caseworker, "If these data labels work better, it would encourage us to record information while interacting with clients" (C3). Prior research states social workers welcome tools that help assess clients' situations, as it enhances their professionalism [9, 45]. Given the caseworkers' waning motivation to label data they do not utilize, and intending to foster intrinsic motivation [19], we identify two potential opportunities for data labels to facilitate improved assessment of client situations.

- Enable extraction of past information relevant to specific client outputs. For instance, employing data labels to filter case notes (Idea 1).
- Generate quantitative insights about clients' behaviors or trends in case outputs to complement qualitative information. For instance, utilizing data analytics dashboards (Idea 2).

Throughout our design discussions, caseworkers consistently communicated their aspiration for ideas that capitalize on collected data labels to support their informational needs. Apart from the data analytics dashboards (Idea 1) and case note filtering using data labels (Idea 2) ideas proposed by our team, caseworkers proactively suggested harnessing ideas like granular data labels (Idea 4) to amplify (Idea 1) and (Idea 2). They also expressed appreciation for displaying most frequent labels for a client (Idea 13) for enabling insights into notable client behaviors.

Aligning data labeling with the caseworkers' information needs is anticipated to yield multiple advantages. Firstly, it empowers social workers to make informed and effective decisions by assisting assessment of clients' situations. Secondly, bringing label utility to caseworkers can promote their active engagement in creating and updating labels. This can lead to shared control on shaping the labels by collaborative decision-making between caseworkers and managers instead of only managers being in charge. Moreover, since the utility of assistive tools hinges on accurate labels, it cultivates accuracy in caseworkers' labeling.

5.2 Enabling Shared Control on Data Label Design to Accommodate Diverse Goals

We recognize the necessity of developing a system that facilitates shared control in the creation of effective data labels. The prior literature shows that caseworkers are dissatisfied with data collected to portray only end results [8, 20]. They emphasize the degree to which the nuances of casework fail to be captured, as the focus tends to be on numbers [2, 4, 6, 54, 60]. Further, based on our findings and the literature [54], we find that caseworkers do not always communicate required changes to data labels to the managers. We assert that reconstructing data labels for adaptability across diverse stakeholder objectives could serve as a fundamental strategy for enhancing the effectiveness and practicality of data labeling. Our proposal envisions a scenario of mutual benefit in which caseworkers engage in data labeling to inform client assessments while also labeling data required specifically for managers or funders. We present a dual approach driven by motivation. Firstly, by leveraging intrinsic motivation through raising awareness about the impact of external data labels (Idea 3) designed specifically for funders or performance reports. Secondly, by tapping into extrinsic motivation by enhancing the usefulness of labeled data for caseworkers (Idea 1, Idea 2). However, to facilitate diverse utilities of data labeling across multiple stakeholders, a collaborative formulation of the composition of the data labels plays a pivotal role (C1, P2, M2).

We propose that our investigation into a streamlined process for adding or removing labels (Idea 6) could be extended into a broader communication channel for iterative label decisions. Such a system can enable the collaborative formulation of quantitative combinations that can assist stakeholders in their diverse objectives. We identify the following implications for the system in designing effective data labels.

- There is a need for enabling a continuous improvement process, in line with Kim et al. [29]. The evolving tasks over time, due to new client needs such as voter registration, or organizational changes such as new collaborations on treatment programs or housing services, require iterative conversations on ensuring labels are up to date with representing the diverse goals. Additionally, NLP tools can contribute to continuous refinement efforts by identifying redundant labels [2] (Idea 9).
- There is a need for collaborative exploration of defining granular yet usable data labels. Discussion on assessing the right granularity of the labels (Idea 4) for comprehensive representation of caseworkers' work surfaced challenges. While caseworkers and managers discussed further granularity of the labels to fully capture case complexities, such as caseworker activities and outputs, such as attempts, requests, denials, and potential time taken, all participants highlighted the concern of creating too many labels. Too many labels can impact the efficiency of recording and the difficulty in managing their exclusivity. It is essential to be open to many label suggestions, which should be iteratively discussed to reach a consensus. Natural Language Processing (NLP) tools offer the potential to aid in the identification of suitable labels for case notes through methods like topic modeling. However, the implications must be supported by collaborative discussions, accounting for the inherent constraints of case note content [47].
- Contrasting opinions are to be expected throughout the process, on the appropriateness of certain labels. Discussion should be driven by how specific data labels would be used for final outcome and evaluated. For example, it is critical to consider the impact of making previously hidden measures more visible. One concern was that creating data labels aligned with the organization's key metrics could create pressure on caseworkers who do not have many "labels" on their case notes, as it may indicate low performance. Additionally, this pressure could lead to over-assignment of data labels by workers. It's imperative to create a shared understanding of how to and how not to interpret the data labels before they are introduced to the data labeling system.

In addition to better representation of work, effective data labels improve assistive tools such as data analytics dashboards to be useful for both caseworkers (gauge outcomes for client situations) and managers (gauge efforts for caseloads) goals.

5.3 Enabling Synthesis of Both Qualitative and Quantitative Information to Enhance Data Labeling

Discussions on the opportunities of utilizing labeled data for caseworkers' information needs also uncovered a novel perspective on the role of quantitative performance data in providing internal insights for caseworkers. Previous research has predominantly highlighted the advantages of quantitative data in terms of aggregation for seamless sharing [7], temporal scaling for comparisons over time [4, 35]. Further, many studies talk about supplementing quantitative performance data with qualitative data for story-telling [4, 8, 20, 27, 35]. However, there is a lack of exploration of the need for quantitative information for caseworkers. Our findings indicate that integrating quantitative data, particularly through analytics, can supplement caseworkers' use of existing qualitative information. By processing past data labels, data analytics can furnish comprehensive insights into case attributes and client behaviors. For instance, if there are multiple instances of housing rejections linked to "no-shows," this could signal a client's lack of compliance, possibly indicating the necessity for behavioral intervention plans. However, we note participants had concerns that extracted quantitative data can lead to misinterpretation of caseworker's performance by management. One example is measuring performance based on the number of client visits required to resolve the case. More client visits could be due to case complexities, such as the client's criminal background restricting housing options. Higher client visits do not necessarily equal poor caseworker performance. It is critical to complement extracted quantitative data with qualitative notes to allow caseworkers and managers to interpret the quantitative measures. This could be facilitated through standardized formats or advanced search, which we describe below.

Our discussions also continually surfaced the lack of access to the context-rich qualitative case notes and a desire to obtain them to complement the objectives of managers and program analysts. Existing research has emphasized the value of complementing quantitative data with qualitative insights to craft narratives that inform financial needs and performance assessment [20, 27]. However, this approach fell short of providing usefulness to diverse stakeholders due to the continued inaccessibility of case notes [5]. In the context of the studied organization, the initial attempt towards mapping quantitative and qualitative information is the presence of data labels for each case note that corresponds to single client interaction. However, we found the integration of qualitative information is still hampered by challenges in accessing unstructured case notes. We provide the following opportunities to improve the accessibility of case note content:

- Accessible data collection formats: Case notes should be structured to be comprehensible by various stakeholders. Caseworkers showed a positive inclination towards standardization despite the limitations it imposed on their current practice of creating free-text case notes. They noted that they understand the significance of enhancing access to qualitative content for contextual details. However, too much prescription for how case notes are written was cautioned [24, 61] by all participants. Participants were concerned that rigid standardization would shift the focus of client meetings to collecting data instead of serving client needs. Case note standardization requiring the input of a set of fields, whether relevant to the client or not, emphasizes data collection and not client needs. A potential approach could also be to apply advanced NLP techniques, which enable caseworkers to freely write case notes while ensuring the correct data labels are identified.

- Advanced search tools: Feedback from participants suggested utilizing advanced search capabilities within case notes to access specific client behaviors and challenges. Employing advanced Natural Language Processing (NLP) tools such as question-answering could facilitate the efficient extraction of information from unstructured case notes. This could encompass searching for reasons behind recurring denials of housing applications for a client or the various steps involved in procuring ID documents. This might highlight the reasons, such as misinformation from a client on their personal details, requiring coordination with other departments to gather data. Nevertheless, owing to the requirement for high accuracy and the absence of a standardized evaluation, training and assessing these models might pose challenges.

Looking ahead, it is imperative to explore efficient solutions that effectively integrate qualitative insights with quantitative measures to meet the diverse needs of various stakeholders.

5.4 Improving the Usability of the Data Labeling Process

Our research expands upon the existing understanding of the crucial role of usability in caseworkers' data labeling [5], largely influenced by constraints in time, funding, and expertise [59]. Our findings emphasize enhancing usability, particularly in terms of navigation (Idea 10, Idea 11) and clarity on labels(Idea 7). In addition to a usable interface, our research surfaces considerations of label visibility, user autonomy, and efficiency as crucial in designing data labeling systems.

- The data labeling system should ensure the fair visibility of all labels to caseworkers. Individuals can naturally tend to favor certain labels over others which can distort the organization's profile, potentially resulting in discrimination of care services for certain individuals and the inadequate representation of certain caseworkers' work. The concern about ignored labels was evident from the discussions of ideas that addressed the presentation of all labels, such as displaying rarely and never assigned labels (Idea 15) and selected and excluded labels (Idea 11). This concern was also raised along with the potential over-reliance on AI-recommended labels (Idea 12). Therefore, it is crucial to ensure that labels with varying degrees of usage are prominently displayed in the interface.
- Data labeling systems should uphold user autonomy, particularly as data labeling often overlaps with direct client interactions. Caseworkers prioritize autonomy during client interactions in deciding when to engage in data activities and when to focus on the client. While managers suggested possible interventions on the interfaces, such as reminders or mandatory actions to nudge caseworkers on data labeling, caseworkers firmly resisted such features viewing them as distractions and disruptions during client interactions.
- Data labeling systems should explore solutions facilitating fast identification of pertinent data labels such as through AI recommendations (Idea 12) and display of most frequent client labels (Idea 13). This not only enhances efficiency but also reduces the risk of manual errors to drive accuracy.

6 LIMITATIONS

Our findings provide insights from workers within a specific organization utilizing a particular data labeling system. This focus could result in a bias in our findings toward the specific characteristics of their system and organization. Future research should explore diverse organizational settings to assess the generalizability of our findings. Although prior literature [4, 20, 38, 51, 60] highlights shared challenges among social workers prioritizing client service and care, further investigation is necessary to validate the applicability of our brainstormed ideas to other non-profit sectors. Thus, this study is limited in demonstrating the representativeness of brainstormed ideas for non-profit organizations.

Furthermore, we did not assess whether the proposed ideas for improving data labeling would be effective if put into practice, and therefore, the study limits its scope in investigating potential concerns. Therefore, the study is limited in demonstrating the practicality of the brainstormed ideas for non-profit organizations. For example, the ideal usage of design suggestions, such as making data labels more detailed or standardizing case notes depends on how they are implemented. Future research could involve field testing these prototypes to determine their feasibility and alignment with the needs of both social workers and organizations. The broader applicability of our design implications should be confirmed through validation on a larger scale and across diverse social service domains.

7 CONCLUSION

As data-driven methods gain importance, it becomes even more crucial to tailor data collection for social service work. We collaborated with a nonprofit that serves people experiencing homelessness and elicited the perspectives of multiple stakeholders such as caseworkers, program analysts, and managers to explore ways to improve data labeling of case notes. Our findings suggest potential design implications to better support social workers' and organizations' needs. This includes aligning data labeling with gaining insights for the case and program management decisions, creating a shared control of data collected, enabling the synthesis of qualitative and quantitative data for diverse stakeholders, and improving system usability. We hope that our work can serve to inform future HCI research on advancing data collection for performance and funding assessment within the domain of social work.

ACKNOWLEDGMENTS

This research was partially supported by the following: the National Science Foundation CIVIC-PG 2228661, DGE-2125858 grants; Good Systems, a UT Austin Grand Challenge for developing responsible AI technologies; and UT Austin's School of Information. We are grateful to our participants and the organization for their trust in our work and thoughtful insights during the sessions. We are also thankful to the anonymous reviewers for providing invaluable feedback.

REFERENCES

[1] Paula Allen-Meares and Bruce A. Lane. 1990. Social Work Practice: Integrating Qualitative and Quantitative Data Collection Techniques. *Social work (New York)* 35, 5 (1990), 452–458.

[2] Asbjørn Ammitzbøll Flügge, Thomas Hildebrandt, and Naja Holten Møller. 2021. Street-level algorithms and AI in bureaucratic decision-making: A caseworker perspective. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23.

[3] David Andrzejewski, Xiaojin Zhu, and Mark Craven. 2009. Incorporating domain knowledge into topic modeling via Dirichlet forest priors. In *Proceedings of the 26th annual international conference on machine learning*. 25–32.

[4] Lehn M Benjamin, Amy Voida, and Chris Bopp. 2018. Policy fields, data systems, and the performance of nonprofit human service organizations. *Human Service Organizations: Management, Leadership & Governance* 42, 2 (2018), 185–204.

[5] Ashley Boone, Carl Disalvo, and Christopher A Le Dantec. 2023. Data Practice for a Politics of Care: Food Assistance as a Site of Careful Data Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

[6] Chris Bopp, Ellie Harmon, and Amy Voida. 2017. Disempowered by Data: Nonprofits, Social Enterprises, and the Consequences of Data-Driven Work. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3608–3619. https://doi.org/10.1145/3025453.3025694

[7] Chris Bopp, Ellie Harmon, and Amy Voida. 2017. Disempowered by data: Nonprofits, social enterprises, and the consequences of data-driven work. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3608–3619.

[8] Chris Bopp and Amy Voida. 2023. " Showing the Context": A Need for Oligopticonic Information Systems in Homelessness Measurement. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–24.

[9] Lieve Bradt, Rudi Roose, Maria Bouverne-De Bie, and Maarten De Schryver. 2011. Data recording and social work: From the relational to the social. *British Journal of Social Work* 41, 7 (2011), 1372–1382.

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[11] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[12] Joanne G. Carman. 2007. Evaluation Practice Among Community-Based Organizations: Research Into the Reality. *American Journal of Evaluation* 28, 1 (2007), 60–75. https://doi.org/10.1177/1098214006296245 arXiv:https://doi.org/10.1177/1098214006296245

[13] Joanne G. Carman and Kimberly A. Fredericks. 2010. Evaluation Capacity and Nonprofit Organizations: Is the Glass Half-Empty or Half-Full? *American Journal of Evaluation* 31, 1 (2010), 84–104. https://doi.org/10.1177/1098214009352361 arXiv:https://doi.org/10.1177/1098214009352361

[14] Ming-Wei Chang, Lev Ratinov, and Dan Roth. 2007. Guiding semi-supervision with constraint-driven learning. In *Proceedings of the 45th annual meeting of the association of computational linguistics*. 280–287.

[15] Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency*. PMLR, 134–148.

[16] Florian Daniel, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah, and Mohammad Allahbakhsh. 2018. Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Computing Surveys (CSUR)* 51, 1 (2018), 1–40.

[17] Scott Davidoff, Min Kyung Lee, Anind K. Dey, and John Zimmerman. 2007. Rapidly Exploring Application Design Through Speed Dating. In *UbiComp 2007: Ubiquitous Computing*, John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 429–446.

[18] Jochen Devlieghere and Rudi Roose. 2019. Documenting practices in human service organisations through information systems: When the quest for visibility ends in darkness. *Social Inclusion* 7, 1 (2019), 207–217.

[19] Robert Dur and Robin Zoutenbier. 2015. Intrinsic motivations of public sector employees: Evidence for Germany. *German Economic Review* 16, 3 (2015), 343–366.

[20] Sheena Erete, Emily Ryou, Geoff Smith, Khristina Marie Fassett, and Sarah Duda. 2016. Storytelling with data: Examining the use of data by non-profit organizations. In *Proceedings of the 19th ACM conference on Computer-Supported cooperative work & social computing*. 1273–1283.

[21] Philip Gillingham. 2011. Computer-based Information Systems and Human Service Organisations: Emerging Problems and Future Possibilities. *Australian Social Work* 64, 3 (2011), 299–312. https://doi.org/10.1080/0312407X.2010.524705 arXiv:https://doi.org/10.1080/0312407X.2010.524705

[22] Philip Gillingham. 2013. The development of electronic information systems for the future: Practitioners,'embodied structures' and 'technologies-in-practice'. *British Journal of Social Work* 43, 3 (2013), 430–445.

[23] Naja Holten Møller, Irina Shklovski, and Thomas T Hildebrandt. 2020. Shifting concepts of value: Designing algorithmic decision-support systems for public services. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–12.

[24] Victoria Hoyle, Elizabeth Shepherd, Andrew Flinn, and Elizabeth Lomas. 2019. Child social-care recording and the information rights of care-experienced people: A recordkeeping perspective. *The British Journal of Social Work* 49, 7 (2019), 1856–1874.

[25] Shih-Wen Huang and Wai-Tat Fu. 2013. Motivating crowds using social facilitation and social transparency. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. 149–152.

[26] Saila Huuskonen. 2014. Recording and use of information in a client information system in child protection work. (2014).

[27] Naveena Karusala, Jennifer Wilson, Phebe Vayanos, and Eric Rice. 2019. Street-level realities of data practices in homeless services provision. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[28] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. worker motivation in crowdsourcing–a study on mechanical turk. (2011).

[29] Mirae Kim, Cleopatra Charles, and Sarah L. Pettijohn. 2019. Challenges in the Use of Performance Data in Management: Results of a National Survey of Human Service Nonprofit Organizations. *Public Performance & Management Review* 42, 5 (2019), 1085–1111. https://doi.org/10.1080/15309576.2018.1523107 arXiv:https://doi.org/10.1080/15309576.2018.1523107

[30] Masatomo Kobayashi, Shoma Arita, Toshinari Itoko, Shin Saito, and Hironobu Takagi. 2015. Motivating multi-generational crowd workers in social-purpose work. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1813–1824.

[31] Rajasekar Krishnamurthy, Yunyao Li, Sriram Raghavan, Frederick Reiss, Shivakumar Vaithyanathan, and Huaiyu Zhu. 2009. SystemT: a system for declarative information extraction. *ACM SIGMOD Record* 37, 4 (2009), 7–13.

[32] Tzu-Sheng Kuo, Hong Shen, Jisoo Geum, Nev Jones, Jason I Hong, Haiyi Zhu, and Kenneth Holstein. 2023. Understanding Frontline Workers' and Unhoused Individuals' Perspectives on AI Used in Homeless Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[33] Tuija Kuorikoski. 2022. Client documents in social work with adults as research data: scoping review of opportunities and challenges. *Nordic Social Work Research* 0, 0 (2022), 1–13. https://doi.org/10.1080/2156857X.2022.2130406 arXiv:https://doi.org/10.1080/2156857X.2022.2130406

[34] Ana M. Leon and Julie Pepe. 2010. Utilizing a Required Documentation Course to Improve the Recording Skills of Undergraduate Social Work Students. *Journal of Social Service Research* 36, 4 (2010), 362–376. https://doi.org/10.1080/01488376.2010.494086 arXiv:https://doi.org/10.1080/01488376.2010.494086

[35] Kelly LeRoux and Nathaniel S. Wright. 2010. Does performance measurement improve strategic decision making? Findings from a national survey of nonprofit social service agencies. *Nonprofit and Voluntary Sector Quarterly* 39, 4 (2010), 571–587. https://doi.org/10.1177/0899764009359942

[36] Kristen Lynch-Cerullo and Kate Cooney. 2011. Moving from outputs to outcomes: A review of the evolution of performance measurement in the human service nonprofit sector. *Administration in Social Work* 35, 4 (2011), 364–388.

[37] Heather MacIndoe and Emily Barman. 2013. How Organizational Stakeholders Shape Performance Measurement in Nonprofits: Exploring a Multidimensional Measure. *Nonprofit and Voluntary Sector Quarterly* 42, 4 (2013), 716–738. https://doi.org/10.1177/0899764012444351 arXiv:https://doi.org/10.1177/0899764012444351

[38] Duncan J Mayer and Robert L Fischer. 2022. Exploring Data Use in Nonprofit Organizations. *Evaluation and Program Planning* (2022), 102197.

[39] George E. Mitchell and David Berlan. 2018. Evaluation in Nonprofit Organizations: An Empirical Analysis. *Public Performance & Management Review* 41, 2 (2018), 415–437. https://doi.org/10.1080/15309576.2017.1400985 arXiv:https://doi.org/10.1080/15309576.2017.1400985

[40] Nancy O'Brien MSW, Thomas McClellan MSW, and Diane Alfs MSW. 1992. Data Collection:. *Administration in Social Work* 16, 2 (1992), 89–99. https://doi.org/10.1300/J147v16n02_07 arXiv:https://doi.org/10.1300/J147v16n02_07 PMID: 10121745.

[41] Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. doccano: Text annotation tool for human. *Software available from https://github. com/doccano/doccano* (2018), 34.

[42] Soya Park, April Yi Wang, Ban Kawas, Q Vera Liao, David Piorkowski, and Marina Danilevsky. 2021. Facilitating knowledge sharing from domain experts to data scientists for building NLP models. In *26th International Conference on Intelligent User Interfaces* (College Station TX USA). ACM, New York, NY, USA.

[43] Michael Quinn Patton. 2014. *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

[44] Frederic G Reamer. 2005. Documentation in Social Work: Evolving Ethical and Risk-Management Standards. *Social work (New York)* 50, 4 (2005), 325–334.

[45] Gwen Robinson. 2003. Technicality and indeterminacy in probation practice: A case study. *British Journal of Social Work* 33, 5 (2003), 593–610.

[46] Karen Salvador, Erika J Knapp, and Whitney Mayo. 2023. Moving beyond data collection: Fostering community engagement in community music schools. *International Journal of Music Education* (2023), 02557614231163556.

[47] Devansh Saxena, Erina Seh-Young Moon, Aryan Chaurasia, Yixin Guan, and Shion Guha. 2023. Rethinking" Risk" in Algorithmic Systems Through A Computational

Narrative Analysis of Casenotes in Child-Welfare. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[48] Patrice Y Simard, Saleema Amershi, David M Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, et al. 2017. Machine teaching: A new paradigm for building machine learning systems. *arXiv preprint arXiv:1707.06742* (2017).

[49] Stephen C Slota, Kenneth R Fleischmann, Sherri Greenberg, Michelle Surka, Keyanna S Evans, James Snow, Sarah Rodriguez, and Tara Zimmerman. 2021. Just infrastructure? Field research on a standardized assessment tool for a continuum of care for people experiencing homelessness. *Proceedings of the Association for Information Science and Technology* 58, 1 (2021), 327–336.

[50] Stephen C Slota, Kenneth R Fleischmann, and Sherri R Greenberg. 2022. A brief typology of time: Temporal structuring and dissonance in service provision for people experiencing homelessness. In *International Conference on Information*. Springer, 211–224.

[51] Stephen C Slota, Kenneth R Fleischmann, Min Kyung Lee, Sherri R Greenberg, Ishan Nigam, Tara Zimmerman, Sarah Rodriguez, and James Snow. 2023. A feeling for the data: How government and nonprofit stakeholders negotiate value conflicts in data science approaches to ending homelessness. *Journal of the Association for Information Science and Technology* 74, 6 (2023), 727–741.

[52] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. *Advances in neural information processing systems* 30 (2017).

[53] Daniel L Strachan, Karin Källander, Augustinus HA Ten Asbroek, Betty Kirkwood, Sylvia R Meek, Lorna Benton, Lesong Conteh, James Tibenderana, and Zelee Hill. 2012. Interventions to improve motivation and retention of community health workers delivering integrated community case management (iCCM): stakeholder perceptions and priorities. *The American journal of tropical medicine and hygiene* 87, 5 Suppl (2012), 111.

[54] Cella M Sum, Anh-Ton Tran, Jessica Lin, Rachel Kuo, Cynthia L Bennett, Christina Harrington, and Sarah E Fox. 2023. Translation as (Re) mediation: How Ethnic Community-Based Organizations Negotiate Legitimacy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[55] Noel Timms. 2018. *Recording in social work*. Routledge.

[56] Pelle Tracey, Patricia Garcia, and Ricardo Punzalan. 2023. Recordkeeping, logistics, and translation: a study of homeless services systems as infrastructure. *Archival Science* (2023), 1–27.

[57] Shahid Umar and Shahidul Hassan. 2019. Encouraging the Collection of Performance Data in Nonprofit Organizations: The Importance of Organizational Support for Learning. *Public Performance & Management Review* 42, 5 (2019), 1062–1084. https://doi.org/10.1080/15309576.2018.1481118 arXiv:https://doi.org/10.1080/15309576.2018.1481118

[58] Bram Verschuere and Ben Suykens. 2020. *Effectiveness and Efficiency of Nonprofit Organizations*. https://doi.org/10.1007/978-3-319-99675-2_116-1

[59] Amy Voida, Ellie Harmon, and Ban Al-Ani. 2011. Homebrew databases: Complexities of everyday information management in nonprofit organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 915–924.

[60] Amy Voida, Ellie Harmon, and Ban Al-Ani. 2011. Homebrew Databases: Complexities of Everyday Information Management in Nonprofit Organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 915–924. https://doi.org/10.1145/1978942.1979078

[61] Amy Voida, Ellie Harmon, Willa Weller, Aubrey Thornsbury, Ariana Casale, Samuel Vance, Forrest Adams, Zach Hoffman, Alex Schmidt, Kevin Grimley, et al. 2017. Competing currencies: Designing for politics in units of measurement. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 847–860.

[62] Katri Ylönen. 2023. The use of Electronic Information Systems in social work. A scoping review of the empirical articles published between 2000 and 2019. *European Journal of Social Work* 26, 3 (2023), 575–588. https://doi.org/10.1080/13691457.2022.2064433 arXiv:https://doi.org/10.1080/13691457.2022.2064433

[63] John Zimmerman and Jodi Forlizzi. 2017. Speed dating: providing a menu of possible futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (2017), 30–50.

A STORY BOARDS

Figure 2: Idea 1 - Filtering case notes based on data labels to facilitate targeted search. This idea seeks to align data labeling with caseworkers’ information needs while assisting their clients

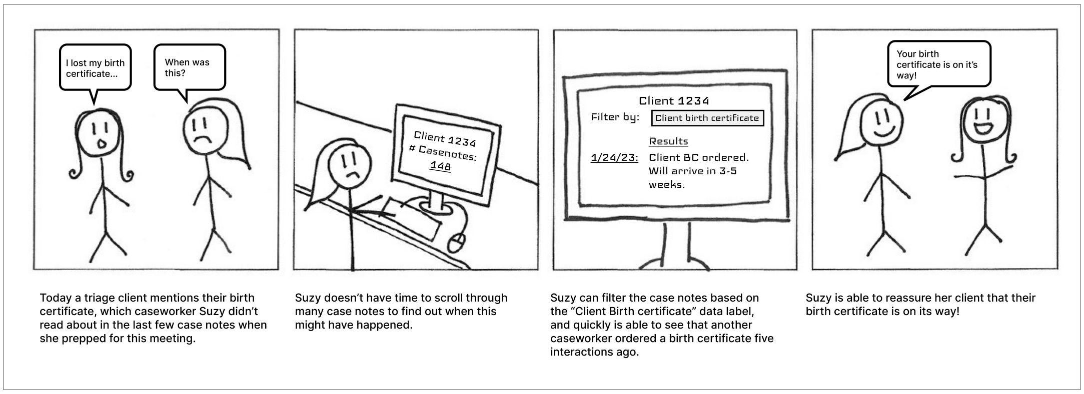

Figure 3: Idea 2 - Client analytics dashboard that leverages data labels to showcase trends. This idea seeks to align data labeling with caseworkers’ goal of understanding client case context to provide more effective assistance

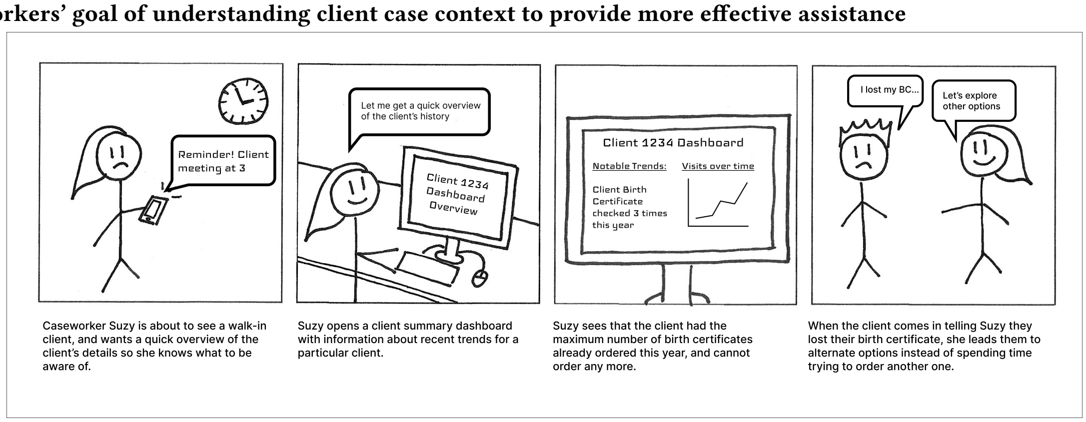

Figure 4: Idea 3 - Dashboard with information on the periodic impact of data labels. This idea seeks to connect data labeling with caseworkers’ goal of providing client service.

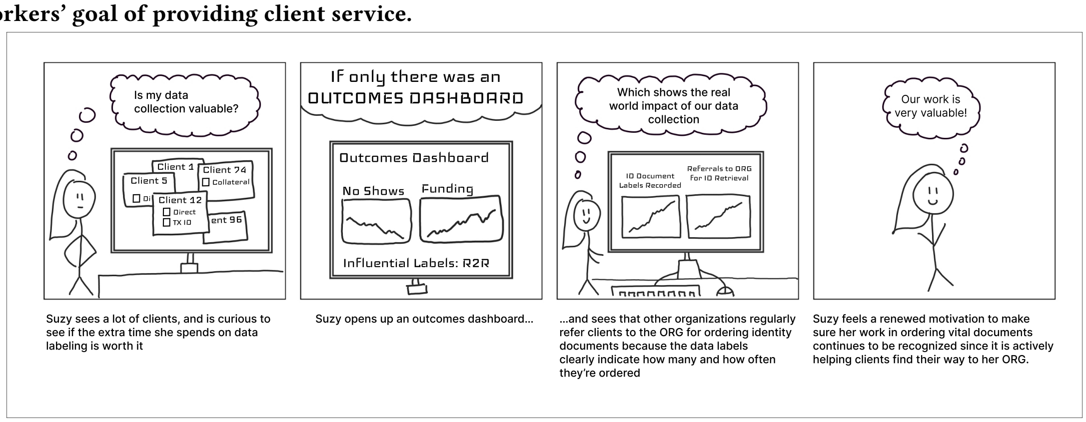

Figure 5: Idea 4 - Redesigning the data labels with increased granularity. This idea seeks to align data labeling to represent caseworkers' values of service provision.

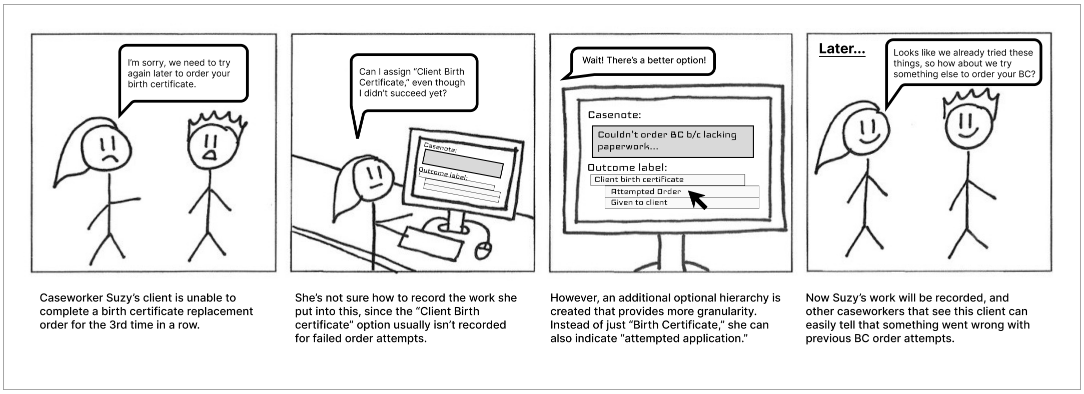

Figure 6: Idea 5 - Standardizing case notes to access contextual information underlying the labels. This idea seeks to align data labeling to represent caseworkers' values of service provision.

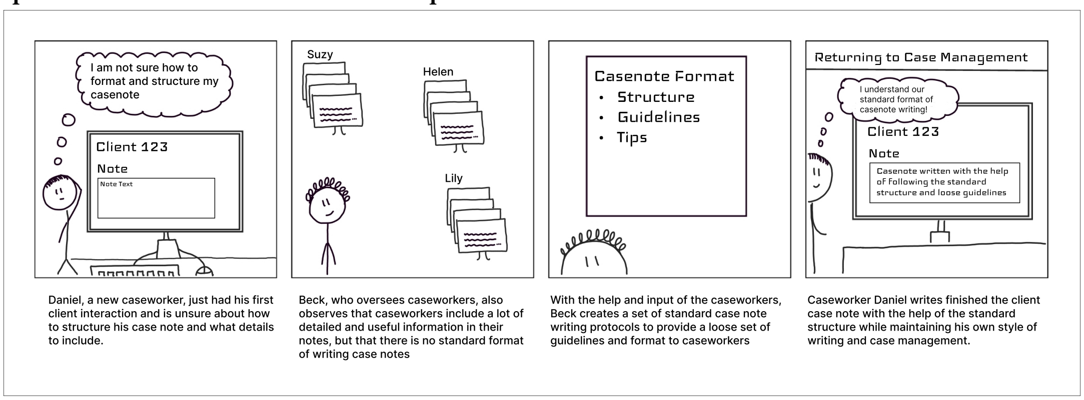

Figure 7: Idea 6 - Streamline the addition and subtraction of data labels. This idea seeks to align data labels with caseworkers' values by enabling collaboration.

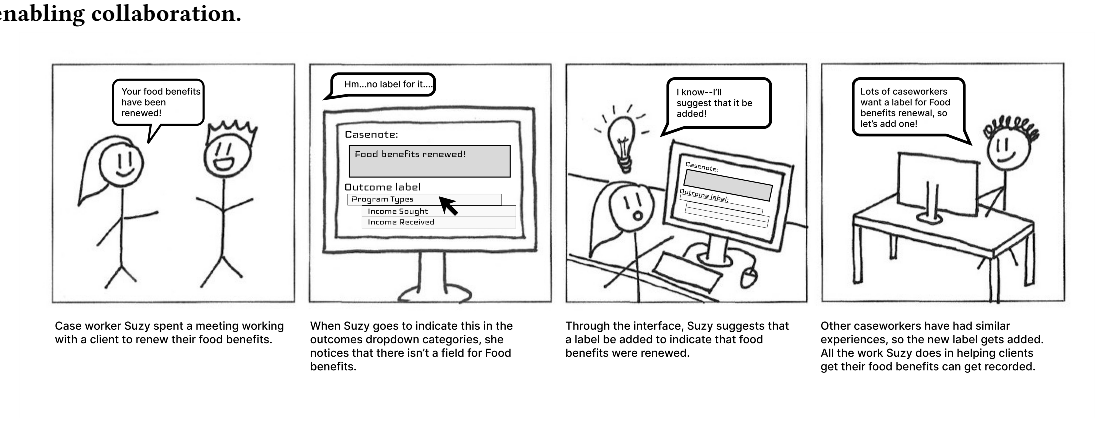

Figure 8: Idea 7 - Instant access to the data label definitions and examples of labeling. This idea seeks to align data labeling with caseworkers' usability needs.

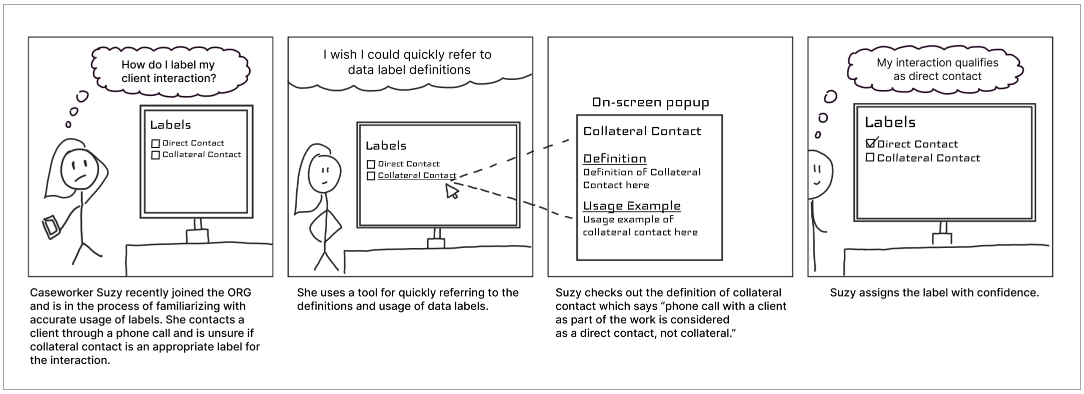

Figure 9: Idea 8 - Periodic training sessions on data labeling. This idea seeks to align data labeling with caseworkers' usability needs.

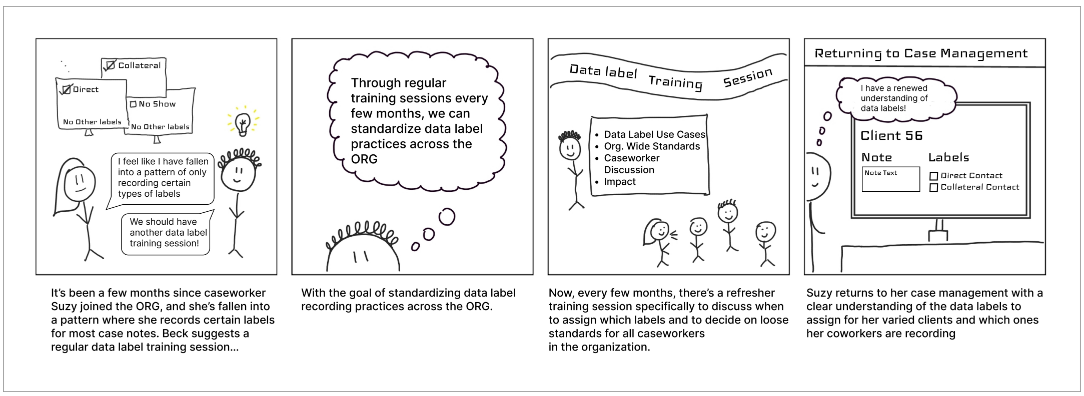

Figure 10: Idea 9 - AI tool to identify redundant data labels. This idea seeks to align data labeling with caseworkers' usability needs.

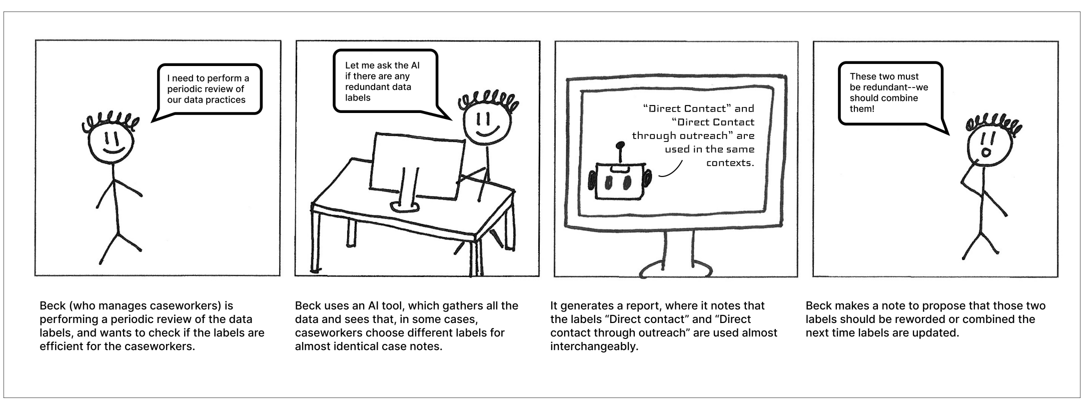

Figure 11: Idea 10 - Search feature to find specific labels. This idea seeks to align data labeling with caseworkers' usability needs.

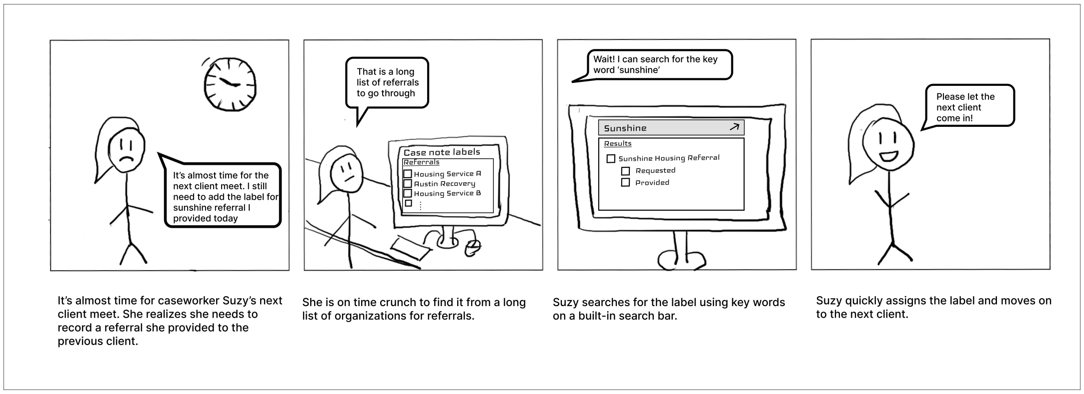

Figure 12: Idea 11 - Visual feedback to navigate through the data labels. This idea seeks to align data labeling with caseworkers' usability needs.

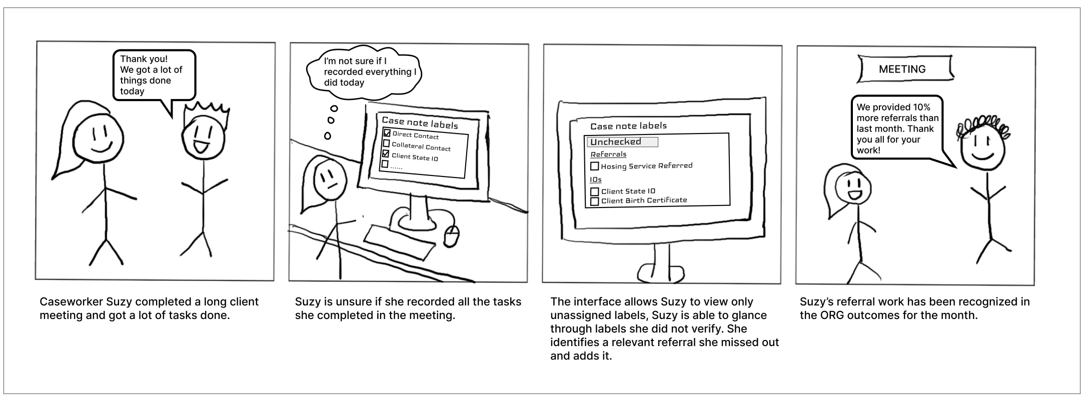

Figure 13: Idea 12 - AI tool that analyzes current case note content and suggests data labels. This idea seeks to align data labeling with caseworkers' usability needs.

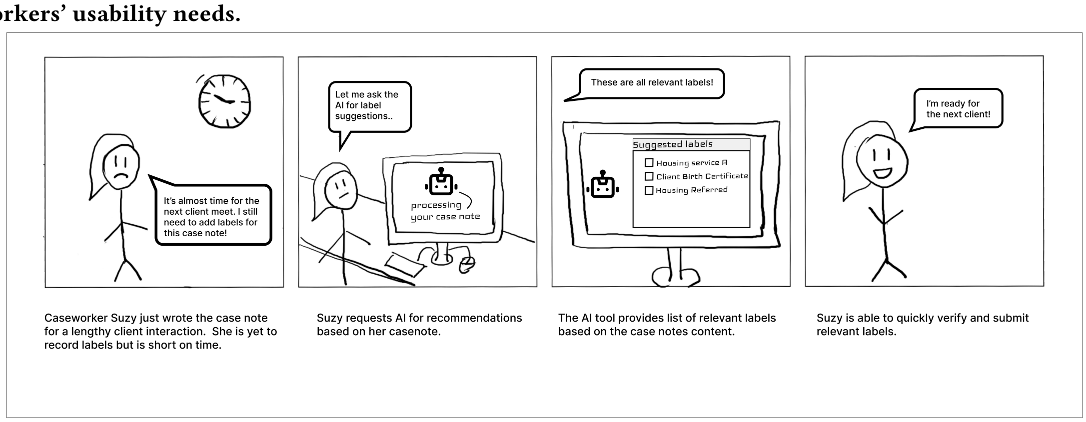

Figure 14: Idea 13 - Presenting client's most frequent and last interaction data labels. This idea seeks to align data labeling with caseworkers' usability needs.

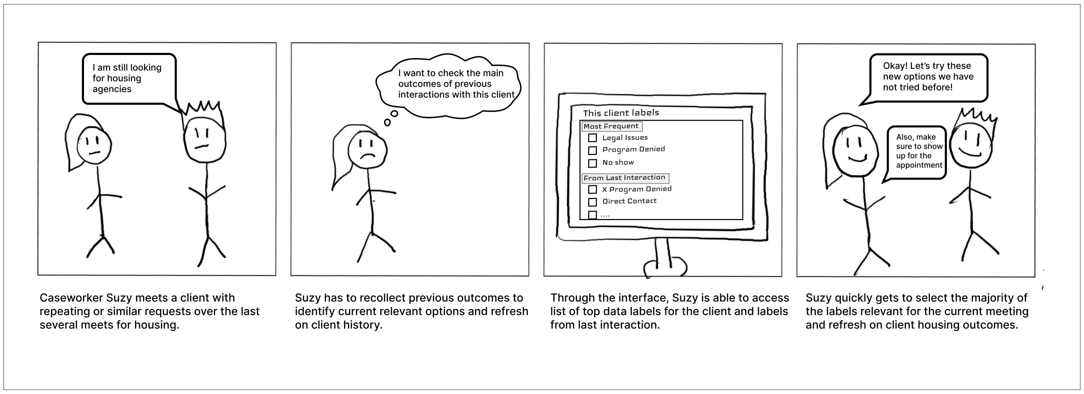

Figure 15: Idea 14 - Reminders on data labeling objectives. This idea seeks to align data labeling with caseworkers' usability needs.

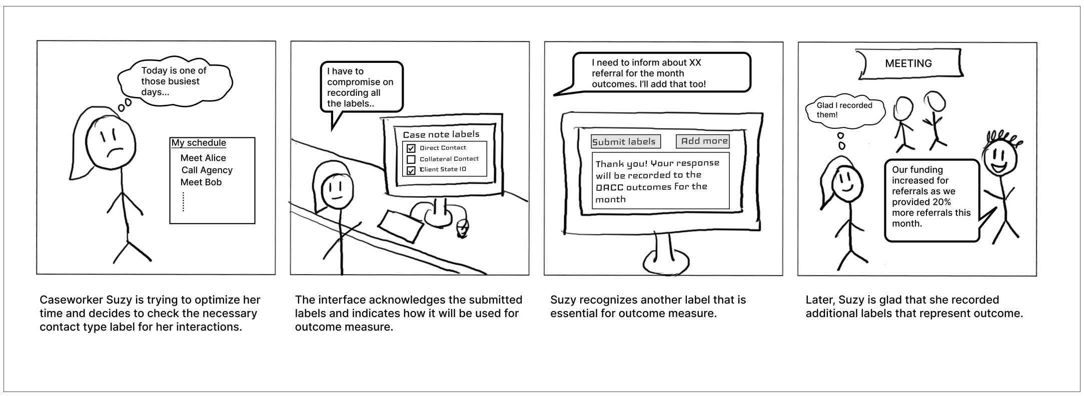

Figure 16: Idea 15 - Displaying labels that have been rarely or never assigned. This idea seeks to align data labeling with caseworkers' usability needs.