A Qualitative Difference: Integrating Qualitative Data into Humanitarian Response Operations

Carleen F. Maitland

Pennsylvania State University, cmaitland@ist.psu.edu

Jean-Laurent Martin

UN Refugee Agency (UNHCR), martin@unhcr.org

Maria Gabriela Urgiles Bravo

UN Refugee Agency (UNHCR), urgilesb@unhcr.org

Alex Bertram

BeDataDriven, alex@bedatadriven.com

Recent developments in qualitative data analytics may generate helpful insights for humanitarian response. At the same time, humanitarian coordination efforts are embracing data sharing platforms to ease data flows. Combined, these two innovations could simultaneously offer operational insights across multiple humanitarian organizations. We pursue this potential through the QualMiner project, an18-month collaboration of the UN-led response to the Venezuelan forced migration crisis in Ecuador. In our efforts to integrate qualitative data, we developed applications with implications for local operations as well as platform features and analyzed data entry processes and information product designs. Our analysis finds the established *quantitative* system serves as an installed base enacting agency and generating three effects, namely framing, artifacts, and informing. We also find collaborative innovation with non-profit users results in direct and indirect factors shaping the data sharing platform's boundaries. Finally, our analysis provides a critical, yet depolarized [1], assessment of advanced analytics in the humanitarian context. These findings have implications for platform boundary theories and critical data studies in the humanitarian domain, as well as humanitarian information management practice.

CCS Concepts: • Human-centered computing→Human computer interaction (HCI); User studies.

Additional Keywords and Phrases: Humanitarian Informatics, Qualitative Data Analytics, Refugees and Forced Migrants, ICTD, Platform boundaries, Critical data studies, Venezuelan crisis

1 INTRODUCTION

The international community's systems for responding to forced migration and refugee crises continue to evolve, embracing a growing list of technologies including digital cash and food assistance, biometrics, and more recently AI [2]–[6]. To date, these systems, particularly those supporting operations, rely almost exclusively on quantitative data. As such, they miss the opportunity to generate greater insight into the unique political, social, and cultural context of the humanitarian response qualitative data affords. Recent developments in machine learning and natural language processing (NLP) have the potential to automate qualitative data analyses, helping integrate them into those humanitarian information

management systems where manual analyses are deemed too costly. Digital platforms can aid this integration by enabling distributed data collection, centralized sharing, and aggregation, generating datasets suitable for automated analyses.

A platform-based, qualitative data system would whisk information away from its source in the field into distant, centralized databases. Yet, this typical design contradicts the supposed value to be gained as well as the UN's localization agenda, which promotes greater integration of local knowledge and organizations into response operations [7], [8]. Constraints may also arise from the humanitarian system itself, where data sharing is hampered by the crisis context, siloed operations, and technical challenges [9]–[11]. Humanitarian innovation, similar to non-profit and governmental efforts, face resource constraints.

Despite these challenges, exploratory research on qualitative data use in humanitarian systems can shed light on several factors. First, many humanitarian datasets on relief operations activities are relatively small, with records ranging in the hundreds or possibly thousands, rather than millions. Given the prevalence of these relief operations 'small data,' empirical findings on qualitative data analytics in these contexts are likely to have broad impact. Second, the resource constraints affecting technology uptake also extend to staffing [12]. Recent research investigates the labor of data science, primarily in industrial contexts [13]–[15]. As data analytics tools are taken up in more diverse sectors, insights for non-profit use are needed [16], [17]. Third, as few empirical studies of data analytics specify the nature of the data – quantitative, qualitative, or mixed ([10] is an exception) – exploratory research can provide useful insights into applying automated textual analytic techniques in humanitarian operations. Fourth, critical data studies in humanitarian contexts can benefit from insight into the granular processes of systems development, which are critical for understanding the balance of benefits and harms in data use [1], [18].

We contribute these insights through the 18-month QualMiner research project investigating qualitative data use by a UN agency, and its partners, in Ecuador. Specifically, the project focuses on the response to the Venezuelan forced migration crisis, part of a regional operation. It centers on the UN agency's information management operations, including a 24-member coordination body consisting of UN and non-governmental organizations (NGOs). The regionally coordinated operation centrally collects and analyzes data across 17 national efforts.

The project team includes an academic, UN and non-profit managers, as well as the leadership and staff of the digital platform firm. Applying a sociotechnical systems and platform theory [19], we examine the effects of the sociotechnical context, including the aforementioned humanitarian system and the platform's ecosystem, on system design and evolution.

The outcomes of our project are fourfold. First, we develop a system for collecting, analyzing, and reporting on qualitative operational data designed to complement established quantitative processes. Its complementary nature resulted in multiple specialized (by sector) dashboards. The dashboard designs highlight the merits of descriptive outputs and raw data access, especially where more complex NLP analyses turn out to be infeasible**.** Relatedly, we identify three effects of the quantitative system's agency and its implications. Second, the multi-functional team's efforts exemplify platform architecture and application coevolution through collaborative innovation, with implications for theory and practice in managing platform boundaries. We reflect on the consequences of this collaboration for humanitarian operations and development in the Global South. Third, we critically reflect on of the value of automated qualitative analytics for humanitarian operations, generating insights for critical data studies in the humanitarian domain. Fourth, our analyses of qualitative data input processes and information product formats provide design recommendations for data entry prompts, qualitative data control, and humanitarian information product design.

The paper begins with a review of the literature then turns to the background material and method. This is followed by the findings. The paper concludes with a discussion of the implications for theory as well as humanitarian practice.

2 LITERATURE REVIEW

Our research integrates constructs from sociotechnical systems and sociomateriality theories into insights drawn from platform studies, informed by empirical research on qualitative data analytics. The breadth aligns with our scope of analysis, which includes the system design for data entry, analysis, and outputs through a collaboration between the platform provider and its client humanitarian information management staff.

2.1 Sociotechnical Systems and Sociomateriality

The data management systems of humanitarian crises are ensembles of soft- and hardware components, as well as the individuals and organizations involved in their development and use [19], [20]. These dynamic, layered sociotechnical systems are complex, with multiple interdependencies creating ripple effects when change is enacted in any one layer. They often are built upon an installed or embedded base, the established technical systems and standards, which enables and constrains innovation [21]–[23]. For instance, integrating innovative qualitative analytic techniques often must contend with the embedded base of quantitative technologies and workflow processes.

From a more granular perspective, the influence of the installed base on innovation can be viewed as arising from the shared agency of the system's social and material elements. The separate field of sociomaterial theory emphasizes the constitutive entanglement of people and technology, as well as the simultaneity of actions such as systems development, diffusion, and use [24]. Agency is not attributed solely to technical components nor individual actors, but instead is dynamic and interactive [25] and enacted through instantiations of a distributed system [26]. In database systems, the agency of the installed base and workflows together can serve as a hegemonic force [27].

Data also evolve, entangled with system components, other data, and people, as well as societal structures through which they are generated [28], [29]. The contextual grounding of data is often attributed solely to qualitative forms, although quantitative data merely obfuscate their grounding [29].

2.2 Digital Platforms

As stand-alone humanitarian data systems morph into collaborative architectures, they are taking on the features of digital platforms. Digital platforms connect users while enabling system extension via boundary resources, in both technical and social forms (e.g., APIs, documentation, user communities) [30], [31]. Using APIs, application developers add value to the platform and constitute key members of its ecosystem [32]. Platform categories include transaction, innovation, or hybrid forms [33].

Platforms are theorized as double-sided, connecting buyers and sellers, or multi-sided, as with data platforms that connect large numbers of users. Platforms may be provided by a 'keystone' firm [34], [35], or through an alliance or non-profit organization [36]–[38]. A platform's extensible nature creates flexible boundaries. Governance structures manage the spanning, erecting, and reinforcing of these boundaries within the platform's ecosystem [30], [39]. 'Decision rights partitioning' is one of many tools for managing boundaries, balancing developer autonomy with ecosystem coordination [35]. This balance can affect innovation, and in turn the coevolution of the platform's architecture and complementary applications [40].

Research on 'platforms for development' articulates their positive and negative impacts [41], [42]. Positive effects include improving access to resources and information, facilitating collective action, spurring application development for commercial, non-profit and government services, and increasing use of scientific information, thereby reducing poverty [39], [41]. Holistic perspectives expose their dark side as well [41]. For instance, [43] find e-hailing platforms helpfully remove institutional voids, but at the same time expand and create new ones.

Data sharing platforms are a unique sub-category, embedded in particular information environments, with varying business models and functions [36], [37]. The platform's information environment may support data sharing, such as in government open data platforms [44], or where consortia seek trusted data sharing partners [36]. Alternatively, data sharing may be highly restricted, as in healthcare [45]. Data sharing platforms may serve as marketplaces, where data are bought and sold [36], [46], or exchanges where data are shared freely in support of open data initiatives [44], or among a closed and trusted community of users [11], [38], [47], [48]. Data platforms can be associated with the standard categories of supporting transactions [36], [46], or innovation [38], or both.

In these platforms, the data itself can serve as a boundary resource [36], [37], in both social and technical forms [31]. As a social boundary resource, data support the identification and analysis of multilateral use cases, specifically through meta-data. Conversely, as "data in transit," they serve as a technical boundary resource [36].

The limited research on humanitarian data platforms examines high-level governance issues and the roles of stakeholders in design. One study examines tensions in managing the open source DHIS2 health information platform, used to develop tailored systems for humanitarian and development programs in over 70 countries [37]. The tensions include serving those who can pay for functionality versus those who cannot, supporting the platform core versus innovation in the fringes, global versus local accountability, and how to pay for maintaining the core. In research on an emergent digital identity system for use in forced migration crises, [11] find the design process lacks sufficient NGO and refugee engagement. The authors critique the platform design's embrace of notions of social identity constructed in the Global North for a system supporting refugees in the Global South.

More granular analyses of health information management systems provide useful insight into key requirements for humanitarian data management systems. These include data's potential to directly affect operations, provide accountability for staff in their work, or enhance knowledge of the operational context [49]. Templated analytics and standardization can ease system use by structuring inputs to generate automated outputs via a dashboard [10]. Standardization limits the data input burden by culling indicators to include only those used systematically in reports. Further, research on customized dashboard designs for an Indonesian health information management system based on DHIS2 finds poor staff visualization competence is a key challenge [50].

The data sharing platform at the center of our research, ActivityInfo, has evolved from a shared data repository for coordination within a single UN agency (UNICEF), to an independent software-as-a-service tool providing a platform-style shared data repository for multi-organization humanitarian coordination. The platform, similar to DHIS2, provides an architecture for data collection and sharing among non-profit organizations in the Global South and connection to external processes through its API. A key contrast between the two is DHIS2 enables adaptation to the local context through software development on the core architecture, while ActivityInfo is designed as a shared, standard (closed) core that can be adapted through the user interface.[1] Also, ActivityInfo is offered by a for-profit firm (BeDataDriven). Following [41]'s questioning of characterizing platforms according to the 'transaction – innovation dichotomy', we characterize ActivityInfo as supporting data sharing and aggregation, as well as interface configuration and, to a lesser extent, application development accommodating the unique needs of different humanitarian crises.

Our project collaboratively innovates the platform's architecture as well as applications, meeting [41]'s call for studies of platform-based innovation in the Global South. However, as discussed below, our project is embedded in the neocolonial UN system, replicating colonial power structures between nations of the Global North and South. Recent research on the UN's 'localization agenda' calls for a disentanglement of the UN system's dichotomous notions of 'local' and

[1] The cost of maintaining tens of customized architecture code bases led ActivityInfo to offer a single standardized core.

'international' and recognition of the elite status of locals employed in the system [7], [8]. Through this lens, we analyze our project's implications for development cautiously, recognizing the elite status of our 'local' team members, making us hesitant to fully embrace the characterization of 'inclusive innovation' [51] with its ties to marginalization.

2.3 Qualitative Data Analytics and Humanitarian Critical Data Studies

Automated qualitative or textual analytics tools emerging in the fields of artificial intelligence and machine learning are targeting smaller and smaller datasets through data efficient algorithms [52], [53], holding promise for the myriad organizations relying on them. Insights into the specificities of *qualitative* or text analytics implementations are limited given the paucity of empirical studies. However, a perusal of the marketing literature, for example, suggests empirical studies are underway, given the prevalence of recently published frameworks and research agendas (see e.g., [54], [55])

In contrast to the extensive scholarship on qualitative data in academic research, the limited empirical research on industrial use of qualitative data finds its most common application in operational decision making, as compared to tactical and strategic forms. Wider use among firms is hampered by the lack of easy-to-use analytic tools [56]. Drawing insights from the rich stream of academic studies using qualitative data, a meta-analysis of 61 qualitative Big Data studies finds scholars struggle with a decrease in data richness. This decrease is attributed to analytic methods, dataset reduction requirements, and the demands of analytic methods for limited noise. The problems are partially solved through human and computer pattern recognition and crowdsourcing [57].

Our project integrating automated qualitative data analytics is informed by insights on established manual methods, in particular, [9]'s granular analysis of Situation Report production, hereafter 'SitRep,' in a UN organization. A Situation Report provides a concise synopsis of the response status in the field, written for an audience typically located at national or international headquarters (see Fig. 1a below). With field work conducted in 2008-9, the authors' analysis focuses on information labor, highlighting the following contextual factors with relevance to our research.

Humanitarian information management occurs within the UN's hierarchical, distributed, and neocolonial context [18]. Within this system, official motivations for data collection vary and include accountability, detecting and remediating gaps in assistance, measuring for improvement, and sharing response outcomes with the affected community. However, [9]'s critical analysis suggests reports and dashboards primarily inform cosmopolitans at headquarters in the Global North and South, by mining the expertise field staff. This reporting is shaped by distance and power differentials. Unsurprisingly, information management work is characterized as 'feeding the beast,' creating frustration due in part to the work generating little value for those in the field [9]. A field study of malaria research data systems witnesses similar sentiments [58]. The authors note that a 'data for good' system, once created, "builds and mobilizes diverse human, organizational, and infrastructural worlds around it, who must now be dedicated to its production, management, and care." (p.153)

Collecting data in complex crises is challenging and requires coordinating with many partners, with Kobo Toolbox easing collection of quantitative data in recent years [59]. Nevertheless, differences in perceptions of what is significant still can be influenced by the political nature of humanitarian response, where host country governments may hold different agendas from those affected by the crisis. Reports can also take on instrumental value for some audiences. For instance, NGOs value 'visibility,' or having their name associated with their activities, for use in fundraising with donors. Conversely, donors value summary information and identifying gaps in support [9].

Additionally, we draw insight from a study on an NGO health information system, which makes passing reference to an application for managing qualitative information [10]. The authors report the system provided 'templated reporting' of qualitative data to ease their entry and integration with the quantitative data. Their analysis finds the qualitative data provide context, justify, and explain quantitative data to diverse audiences and are valued for their ability to bridge distances.

Beyond studies specifically concerned with qualitative data, critical data studies in the humanitarian domain offer critiques of data and the systems devised to manage and control them. While overlap exists with the aforementioned scholarship on 'platforms for development,' humanitarian critical data studies more directly confront the entanglement of data with UN and international humanitarian systems. Rightly so, many studies focus on systems directly managing sensitive data from communities of concern, in particular systems of digital identity and biometrics [60]–[63]. Yet, the field has become polarized, with [1] calling for greater balance in perspectives as well as between scholars and practitioners. Our analysis, concerned with a system more likely to surveil aid workers than affected communities, and including aid workers and systems providers as researchers, offers a unique perspective.

2.4 Summary

Our analysis examines the role of the installed base of the quantitative system and the agency that technologies and people jointly enact, while investigating factors shaping the platform's boundary. We conduct this analysis while working to integrate qualitative data into an established platform and system designed primarily for quantitative formats. The platform provides an architecture for data sharing, interface customization for adapting to the local the crisis, and APIs enabling expert humanitarian users to write customized applications to fulfill unique data management needs. Our work informs and contributes to scholarship on platform boundaries and data platform innovation in global humanitarian systems.

We generate these insights by addressing the following questions: (1) What factors affect integration of automated analyses of qualitative data into an established quantitative-based humanitarian system? (2) What value do advanced qualitative analytic tools hold for humanitarian operations? (3) What implications will our project's innovations have for humanitarian operations, the platform, and humanitarian system design? And (4) What insights can we glean into humanitarian innovation and its effects on development in the Global South?

3 BACKGROUND AND METHOD

Our research takes place within the context of the UN response to the Venezuelan refugees and migrants crisis. Below, we provide background on the humanitarian crisis and the platform involved in the study, as well as our methods.

3.1 Background

The Venezuelan humanitarian crisis has its roots in the death due to cancer of President Hugo Chávez in 2013. As the situation in the country became increasingly dire, larger numbers of Venezuelans fled starting in late 2018. With worsening conditions due to the Covid-19 pandemic, by the fall of 2021 over 5 million of the 28 million inhabitants had left[2]. The majority settled in Colombia, Brazil, Peru, and Ecuador, with the latter serving as both a host and transit country. In 2019, the World Bank estimated 400,000 of the roughly 1.1 million refugees and migrants in Ecuador are settled in the country[3]. This multi-country migration flow led the UN, in 2018, to establish a regional coordination arrangement of 17 countries called the 'R4V platform.'[4]

[2] https://www.r4v.info/en/refugeeandmigrants

[3] Challenges and Opportunities of Venezuelan Migration in Ecuador (English). Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/662681592905703335/Challenges-and-Opportunities-of-Venezuelan-Migration-in-Ecuador

[4] https://www.migrationpolicy.org/sites/default/files/publications/mpi-iom_venezuelan-profile_english-final.pdf

Globally, humanitarian crises are managed through coordination bodies made up of UN agencies, international and local non-governmental organizations (NGOs), and, in some cases, host country government representatives. Sectors, such as health, education, and shelter, are managed by designated organizational 'sector leads.'[5] Coordination may occur on sectoral, geographic, or national bases, or across all three.

In 2019, Ecuador's the national coordination body consisted of 24 partners (11 UN agencies and 13 NGOs). Facilitating systematic data collection on their operations, the coordination body jointly defined a shared monitoring framework consisting of quantitative and qualitative indicators. In subsequent years, the framework has been updated annually with aims at streamlining. At one point, it had ballooned to 729 indicators, with 66 in qualitative format.

With shared indicators, the majority of coordination body members adopted the same digital platform. Offered by the Dutch firm, BeDataDriven, the platform's design, specifically for humanitarians, avoids the complexity of popular database systems (Access, MySQL) and accommodates field use. Our project is the platform's first systematic exploration of qualitative data, despite its users' increasing use of text.[6]

In January of 2019, the joint monitoring framework was transcribed into platform-based database forms, allowing staff to begin data entry. Most of the coordination body organizations committed to the additional qualitative data entry task. Entered data are then aggregated, exported, analyzed, and linked to automatically generated reports on various dashboards. The national response data feeds the Inter-Sectoral dashboard for both the national and regional response. Our project would integrate qualitative data into this system.

In Appendix A.1 we present examples of the qualitative data emerging from the indicators, specifically short and long entries collected in two sectors. Naturally, the volume of qualitative data varied by organization and sector, with some sectors conducting activities more amenable to textual descriptions. Entries varied within sectors as well. The aggregate of the qualitative data comprised the corpus for our analyses, maximizing its size.

3.2 Method

Our research uses a case study design and qualitative methods [64]–[66], including participatory action research [67] and semi-structured interviews. These data sources are supplemented by secondary data from documents including meeting minutes, reports, SitReps, the digital platform's user manual, and training documents.

The project ran for 18 months, from June 2019 until December 2020. During the project, all team members had other responsibilities. The team consisted of an academic, staff of the digital platform, and humanitarian information management staff working in the response. The distributed team members were located, respectively, in the U.S., Europe, and Ecuador. The Ecuadorian team members included both local and international staff. The project obtained research ethics clearance through the U.S.-based university and approval of the UN agency headquarters. Interview subjects provided verbal consent for recordings based on written consent documents shared during email-based recruitment. Interview subjects, similar to research team members, largely were also humanitarian information management staff.

The research involved participatory systems development activities in addition to typical qualitative methods of interviewing subjects. In these activities, conducted in four, sometimes concurrent, phases, team members took turns leading. The first phase, led by platform staff, investigated potential techniques for analyzing the qualitative data. The second, co-led by the academic and the humanitarian staff, analyzed the processes for the qualitative data entry by implementing partners. In the third, humanitarian staff tested traditional qualitative data analysis practices and tools (e.g.,

[5] For more information on this 'Cluster Approach' see
https://www.humanitarianresponse.info/en/coordination/clusters/what-cluster-approach
[6] https://bedatadriven.github.io/QualMiner/data-in-activityinfo-over-time.html

using Atlas.ti) for potential automation through R scripts and developed three prototype information products – an Operational Report styled as a local version of the national SitRep, a set of Sectoral Dashboards, and a Control Dashboard. The fourth, co-led by the academic and humanitarian staff, explored user perceptions concerning integration of qualitative data into various forms of information products.

The sampling frame for the data entry analysis consisted of 17 of the 24 implementing partners, excluding those organizations that did not report on qualitative indicators, likely viewing them as irrelevant for their operations. Of the 17, 9 participated. As indicated in Appendix Table A.2.1, they vary based on the amount of narrative information entered on the platform, size, diversity of services provided, and offering services in sectors with few actors (e.g., LGBTQ services). These latter three characteristics provide insight into whether the type of service influenced use of qualitative data.

For each of the 9 organizations, 1-3 staff members were interviewed, for a total of 15. Interviews were conducted in person and online, as location and time permitted. As shown in Appendix Table A.2.3, the respondents hold a variety of positions related to information management, including Field Office Manager, Situation Specialist, and Monitoring & Evaluation Officers.

Interviews consisted of common and organization-specific questions. During each interview, to ground the discussion, participants viewed the regional coordination body's website and reviewed a report we provided of their organization's narrative data, including simple descriptive statistics.

The sample for the information products analysis, gathered via snowball technique, consisted of 26 staff members across 25 organizations. As shown in Appendix Table A.2.3, the resulting sample included 5 humanitarian organizations (3 small, 2 large), leaders of 2 working groups, 3 local partners, 10 coordination body partners, and 5 local authorities. The local authorities expanded the sample beyond those directly or indirectly involved in development of the information products. The interviews involved showing actual and hypothetical designs and gathering likes/dislikes.

Data collected from interviews, which were limited to the data entry and information products analyses, were analyzed through open coding methods using Atlas.ti. Themes and findings were developed jointly among the leaders, then shared with the broader team for debate and discussion.

4 FINDINGS

4.1 Qualitative Data Mismatches

The initial analytic process encountered two 'mismatches' – one between the data and the analytic techniques and a second between the data and the platform data export process. The initial text mining process, led by the platform's data scientist, applied standard cleaning and processing procedures to the unified text corpus, followed by a wide variety of analytic techniques, including n-grams frequency and sentiment analysis. At first, the results appeared to be insightful, however upon closer inspection we found they were mostly unsurprising or unactionable. For instance, word frequency and importance analyses (tf and tf-idf)[7] using several months of data reflected changes in the nature of the crisis, such as the onset of the Covid-19 pandemic, but being historical, were not particularly actionable. Rapid Automatic Keyword Extraction (RAKE) [68] found the top three most frequent phrases were 'human mobility,' 'gender violence,' and 'obstetric emergency'. Sadly, only the latter was unexpected, and therefore of interest. Finally, a sentiment analysis-generated word cloud also produced uninteresting results. In hindsight, this was unsurprising given the formal tone and lack of position statements in the data.

[7] https://en.wikipedia.org/wiki/Tf%E2%80%93idf

We assessed the cause of the data-analytic mismatch to be the alignment of the qualitative with the quantitative indicators. The team had expected the indicators to elicit stories – compelling examples reflecting the alleviation of suffering through the assistance programs. Instead, the data largely reflected dispassionate descriptions of assistance operations. The rationale for aligning the qualitative and quantitative indicators was threefold. First, limited staff meant qualitative data analytics should be automated, which would be eased by alignment with the quantitative measures. Second, the team intended the two forms of data to complement one another in outputs. Third, it would ease integration of qualitative data into the established quantitative data workflow, including report and dashboard publishing processes. As the indicators could not be changed, a new analytic strategy was required.

The second mismatch occurred between the qualitative data and the platform's data extraction feature. The feature did not easily handle qualitative data, frustrating staff unable to extract the data they had entered, which disincentivized further entry. The humanitarian team designed a flexible "Extract-Transform-Load" script, or ETL for short. It enables downloads of both quantitative and qualitative data in a single flat file appropriate for analyses, avoiding multiple downloads or manual extraction. Subsequently, the feature was integrated into the platform by its developers and the script was shared on Github.

4.2 User Requirements for Data

The new analytic strategy involved breaking the data into sector-specific sets (e.g., health, protection, communications) and conducting manual analyses prior to embarking on automated forms. Responsibility transitioned from the platform's data scientist to the humanitarian staff to tap into domain and domain-specific jargon expertise.

While resource constraints required the final analytic workflow be automated, the team undertook exploratory manual analyses to provide greater insight into the sector data and possibly a basis for automated results verification. Using standard human-in-the-loop tools (e.g., Atlas.ti) and inductive coding, the team explored summarizing the narrative data through quantitative charts, recognizing the irony of generating yet more quantitative data but understanding the value of their summarizing function. Through the analyses, the team gained awareness of the data's characteristics. For instance, surprisingly the data generated from free text fields was incomplete or merely copied from other reporting systems, making it less amenable to automated text mining.

Bringing the data cleaning and pre-processing closer to eventual users surfaced requirements concerning data loss and visibility. The existing quantitative system, specifically the Inter-Sectoral Dashboard, reported data based on the metric of 'number of persons served.' It thereby excludes secondary or indirect assistance activities, such as coordination efforts and interfacing with the government. Partners wanted assurances the new narrative data pre-processing and cleaning workflows would not lose or systematically exclude these data.

A second requirement, 'visibility,' also noted in [9]'s SitRep analysis, mandated retaining organization identifiers. Partners felt 'visibility' was valuable in interactions with potential donors, but also to track activities across multiple offices. Considering these requirements, the designed data cleaning process exports data through the ETL script, uses the Tidytext R package for tokenization, removal of stop words, and uses the Dplyr R package to select and filter the data.

From there, the automated analyses, including 'bag of words,' semantic parsing, word counts, and 'named entity recognition,' generate two primary outputs: Sectoral and Control Dashboards. Using R-markdown (for combined qualitative and quantitative outputs) and Shiny (for qualitative information combined with maps), the dashboards include summary charts, graphs, maps, and displays of processed data.

While valuable, the dashboards fall short of the deeper semantic analyses the team initially envisioned. Beyond the structure of the data, a key limiting factor was the lack of a standard humanitarian ontology or dictionary, in either English or Spanish. The digital platform team did note the availability of general Spanish natural language processing tools in

Python. However, at that time, the humanitarian information management team's investment in R had limited its experience with Python.

4.3 The Dashboards

The team's goal of delivering insights from qualitative data to operations staff became embodied in the Sectoral and Control dashboards. Each Sectoral Dashboard's key features include graphics and lightly processed text, visual representations of collaboration, search capabilities, and an export feature. Special attention was paid to optimizing loading time to accommodate rural users' more limited bandwidth. Visually, the dashboards present a series of graphics of individual indicators followed immediately by tables of lightly processed raw text attributed to specific organizations meeting that design requirement (see Figure 2c below for an example). Each indicator's browser and search features enable data retrieval on direct and indirect activities, geography, date, or partner. In addition to convenience, this feature contributes to transparency, making visible the qualitative data behind the quantitative summary statistics. Also, the export feature, which generates a pdf document, meets users expectations developed by SitReps, which are typically provided in pdf format.

As compared to a Sectoral Dashboard's individual staff users, the Control Dashboard's users are the numerous coordination teams. The Control Dashboards use database statistics to enhance quality control of both quantitative and qualitative data, displaying empty fields and narrative data length. The dashboard allows coordination leaders to know which organizations are (not) entering data, in turn, providing another source of visibility for those organizations successfully entering data. The information also provides insight into operations, to the extent they are reflected in the data. The need for this information arose organically during the project.

Following some preliminary testing, the humanitarian's Control Dashboard code was shared via Github and with the platform developers, who began integrating its features into the core platform. Despite increases in narrative data use, the platform had yet to offer a means of browsing or analyzing this type of data. The feature enabled platform-based data quality assessments, negating the need to export data to third-party software or R or Python scripts.

Finally, unfortunately, the critical issue of a formal joint data validation process was beyond the scope of the timeframe of our project. Validation was presumed to occur within individual partners at the data entry stage. However, discussions surfaced the need for a separate the multi-organization process to resolve potentially differing standards for what is deemed 'sensitive.' A potential model exists in the coordination body's two-step process for validating SitReps. The process first uses partners and sector leads, and then in the second stage relies on more senior coordination body members. That process primarily involves softening language, such as related to operational challenges or interactions with the government.

The platform lacked features promoting data quality or validation, for either quantitative or qualitative data, with expectations that qualitative data validation would be more onerous, due to potentially contested interpretations and concerns about misrepresentation. The team discussed the possibility of limiting distribution of qualitative data-based outputs to the coordination bodies until trustworthy validation workflows could be established. Ultimately, the decisions and workflow process specification were left to the Information Management Working Group.

4.4 Qualitative Data Entry and Management

The quality and quantity of the aggregated qualitative data relies on organizational data management processes. To understand partner organizations' capturing and managing processes and perceptions of qualitative data entry, principally vis-à-vis established quantitative data processes, the humanitarian staff member of the project team interviewed 15 staff members from 9 partner organizations.

Interestingly, the 9 organizations displayed a great deal of diversity in their data management approaches. For data capture, 8 distinct tools were mentioned (with number of mentions in brackets), including programmatic matrices (14), unit-based and/or general narrative reports (14), and internal systems (7). For systematic data management tasks, aggregating and disaggregating indicators (9), compiling, summarizing and narrative report revision (8), and content filtering (6), were the most mentioned among 11 separate tasks. However, those for data verification (2) and data confidentiality (1) were less so.

Perceptions of the burdensomeness of qualitative versus quantitative data entry varied, with 50% indicating qualitative data are more burdensome, 30% believe quantitative data are more so, and 20% are indifferent. One informant noted:

> *"Quantitative information is about counting every day and, at the end of the month, you have an aggregate and that is it. Narrative information is about sitting down, thinking about it… working at it, analyzing it, and you are constantly running with 50 people outside, waiting for you." – NRC*

Divergent responses were also given to questions concerning the challenges of qualitative data entry. The question generated 13 distinct responses, including: it takes staff time (10), it adds to workload and conflicts with the organizations' priorities (6), and lack of process or habit (3). Two staff members voiced concerns with the overall approach:

> *"The narrative part shouldn't be based on indicators. Indicators are very specific… We should rather focus on sectors. Which are the changes in context? What are the needs? Which are the new things one has done in child protection? I can write one paragraph that covers six or seven indicators." – UNICEF*

> *"It would be good to have testimonials of persons of concern [refugees] and show how, with their stories, they are part of the numbers we see on the dashboard." – HIAS.*

This last comment suggests a lack of clarity and divergence in expectations of the system, as well as of whose voices should be reflected. Refugee *testimonials* are controversial as they can be exploitative and need to be differentiated from other forms of amplifying voices.[8]

Greater consistency was observed in the location of data entry, whether centralized or decentralized. Centralized entry collects data outside the platform (e.g., via spreadsheets or email) and enters them centrally. In contrast, in decentralized approaches data are entered directly from multiple offices, taking advantage of the platform's web-based architecture. In our sample, 8 interviewees mentioned centralized, versus 2 for decentralized, processes. This contradicted expectations that organizations would keep data entry 'closer to the field,' the context in which it is generated. Yet, as noted above, control over what is entered, and subjecting data to centralized validation and filtering, is viewed as necessary. Also, the data sharing the platform enables raises concerns about precisely when external parties can 'see' an organization's entries. Nevertheless, centralized filtering and pre-processing can restrict data content, as noted by one staff member:

[8] For a broader discussion of refugee digital storytelling see Maitland (2020, p.5) "Access and Agency" Retrieved January 19, 2020 from UNHCR Innovation: https://www.unhcr.org/innovation/wp-content/uploads/2020/04/Access-and-agency_WEB042020.pdf and Tammas (2019) "Refugee stories could do more harm than good. The pressure of storytelling can leave refugees feeling tokenised and disempowered." Retrieved November 1, 2019 from Open Democracy: https://www.opendemocracy.net/en/refugee-stories-could-do-more-harm-good/

"Everybody has a different perception: [The] Esmeraldas [office] can think this and this is important; and maybe we don't even consider it… Maybe we are losing stuff along the way." – UNHCR

Lastly, we gathered perceptions on the platform's data entry interface prompts, due to their likely effect on the data's specificity, length, and overall value. We asked informants about their preferences for open versus precise prompts and again received mixed results. Those advocating for more precision reflected:

"The more precise the prompt, the more complete the report will be." – HIAS

"Even if you put a lot of standards [for narrative indicators], because of the fact that it is narrative, there is going to be useless information… But there would be a lot ***less*** *of it [useless information, if you put standards]." – CARE*

"The more marked out the space, the more precise the [information] that is required from us, the better. It shouldn't be open for us to write anything that comes to mind." – Dialogo Diverso

Conversely, those preferring more open prompts stated:

"It is more about a process in which people that report are a lot more conscious of what they are required to report. Prompts should be open, though it is also important to have examples." – JRS

"For me, it would be best, before going to any specific activity, to have a space to write about UNICEF's actions… that not necessarily go under a specific activity. It would be best to have flexible – open prompts under each sector to write about the general situation." – UNICEF

One informant took the middle ground, observing that the type of prompt needed for an indicator should depend on the activity that is being monitored.

"Some [indicators] don't need to be so open. Others do: those that help to show the human mobility work the organization is doing. However, sometimes [the latter] are more limited." – Misión Scalabriniana

Still others pointed out the limited effect of prompts and the range in volume that might constitute 'quality data.'

"We include the information we want there. It doesn't matter if it says: "include this [or] include that", we always include what we consider should be visible." – OIM

"There is also the fact that everybody has a writing style... A lot of people think that the more you write, the more complete. And that is not true." – HIAS

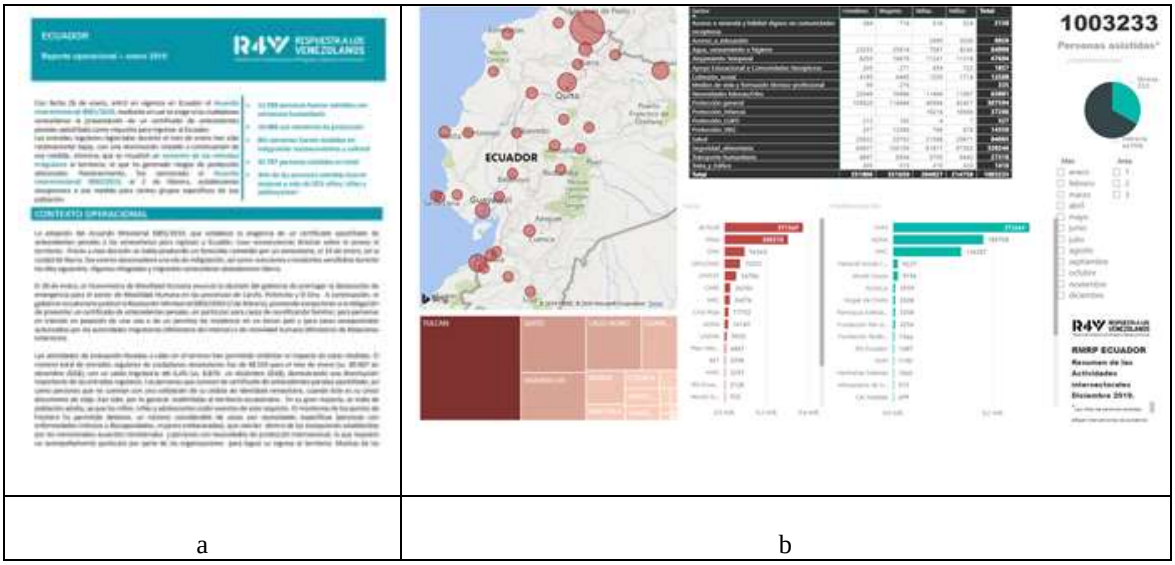
Overall, given the divergent views, tailoring the prompt to the activity, and providing open text boxes independent of specific indicators can provide a middle ground. During the 2020 monitoring framework and forms revision, open text boxes were introduced and became very popular in terms of the amount of data entered.

4.5 Information Product Designs

For the project, information products represented the final stage of communicating insights derived from qualitative data, having a broader audience than our dashboards. We investigated established, new, and theoretical product designs.

For the first comparison, from the Ecuadorian team's portfolio of 17 information products,[9] we chose two existing and two new products. The established SitRep (Figure1a) primarily uses qualitative data. Patterned on the SitRep, the new Operational Report also primarily relies on qualitative data, yet tends to be shorter (typically 2 versus 4 pages).

Conversely, the Inter-Sectoral Dashboard (Figure 1b), is largely quantitative, fed by platform data and automatically generated using Microsoft PowerBI. This dashboard displays data in various formats (e.g., maps, bar charts) based on the 'number of people assisted' metric. Its interactive elements filter information by date and geography. The new Sectoral Dashboard, depicted further below in Figure 2c, as previously described, presents summary data with raw text.



a b

Figure 1: (a) the Situation Report (SitRep); (b) the web-based Inter-Sectoral Dashboard

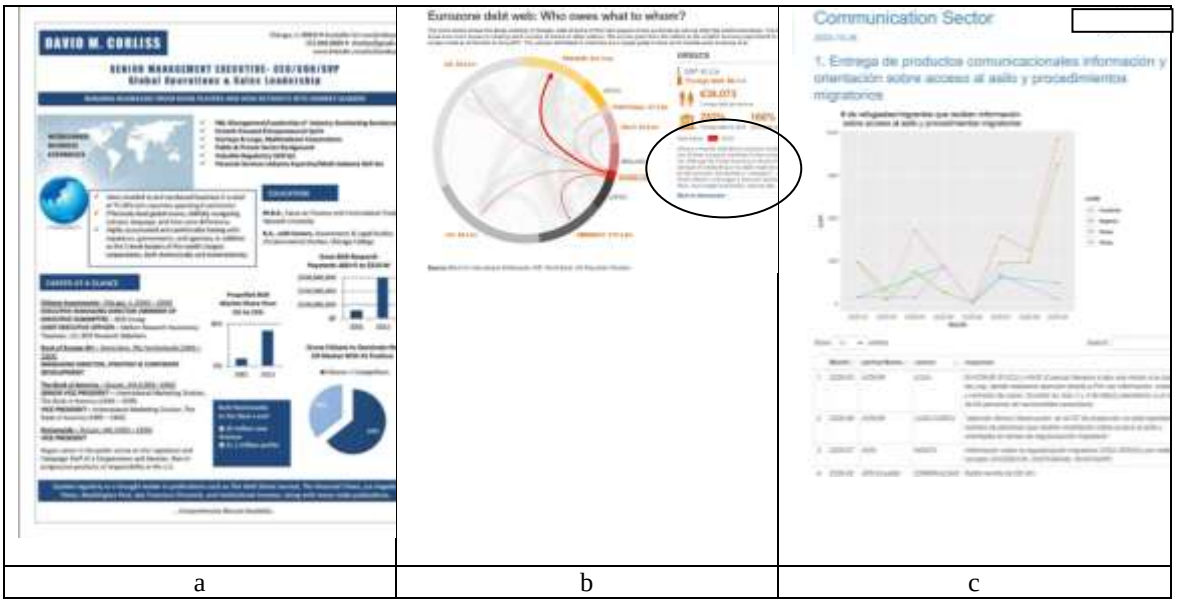
As our sample included organizations uninvolved in production, our first assessment targeted awareness of these various reports. By number of mentions, people were most aware of the SitRep (23), followed by the Operational Report (16), the Inter-Sectoral Dashboard (14), and finally the newer Sectoral Dashboard (12). The popularity of the SitRep was further confirmed by results of an unrelated UN-led Mailchimp campaign indicating the report received the most clicks. The campaign also revealed a clear increase in downloads when products are sent via email. The analysis of awareness was then followed by an assessment of general likes and dislikes for each of the four, listed in Table A.3.1 in the appendix.

Next, we conducted two comparisons, teasing out preferences for formats (1) integrating quantitative and qualitative data and (2) static versus interactive forms. Lacking a designer, we found two exemplary integrated information products online, from unrelated fields (see Fig. 2a and b, respectively, a person's CV and a European data portal). Using actual exemplary as compared with theoretical designs suggested technically feasibility. The static integrated format envisions

[9] For a detailed report of the various information products and their frequency of distribution see 'GTRM Ecuador: Catálogo de Productos de Información - Mayo 2020 [ES]' https://www.r4v.info/en/node/4689

narrative data providing context or richer meaning, complementing traditional infographics. In the interactive integrated format, applying filters refreshes narrative content on the right (circled for clarity) alongside the quantitative-data-driven circle graphic on the left. At the time of our study, few publicly available examples were found.

The final format presented for comparison was our own design of the Sectoral Dashboard, shown in Figure 2c. As previously described, it provides graphical information together with raw text and a search function. During interviews, subjects were shown a live demonstration of the hybrid and sectoral dashboards.



a b c

Figure 2: (a) Static Integrated, (b) Interactive Integrated, (c) Sectoral Dashboard

Teasing out preferences for stand-alone versus integrated qualitative data in static formats, we compared the established SitRep (1a) to the theoretical format (2a) and found the majority (18) preferred the static integrated version, a few (4) indicated both, and a minority (3) preferred the current SitRep layout. In terms of preferences for integrated content among the interactive formats (1b vs. 2b), 16 preferred the interactive integrated format, 5 preferred both, and only 4 preferred the current, primarily quantitative-driven, format of the Inter-Sectoral dashboard. Hence, in both the static and interactive modes, the informants clearly preferred the formats integrating the qualitative data.

Next, our comparison targeted preferences for static versus interactive forms among the three integrated formats (Figures 2a-c). For general features, the greatest number of positive traits was mentioned for the static format (17), while the Inter-Sectoral dashboard format was second (13), and the Sectoral Dashboard (11) last. Analysis of the likes/dislikes for each, shows the likes outweighed the dislikes, with minimal differences. As the two theoretical formats lacked actual data, comments focused on the layout. Specific like and dislikes are provided in Appendix Table A.3.2. Generally, the positive comments were unsurprising, in contrast with insightful negative ones. Informants voiced concerns about the amount of work it would take to generate these products on an ongoing basis. Also, they expressed worry that as these formats grow in complexity, sharing information might become increasingly difficult, especially with refugees, migrants, and their host communities. These concerns referenced both the use of jargon as well as the technical access and skills needed to read and interpret complex graphs.

4.6 Portal Boundaries

Our project also generates insights into the factors shaping portal boundaries, including analytics, database management, skills, and standardization. Integrated analytic features, like the platform's mapping tool, negate the need to export data to third party software, easing use. The platform staff explored whether qualitative data processing and analytics might also be conducted directly on the platform, considering intensive or light approaches. Effective on-platform analytics would need to contend with the complexity of the database structure, with multiple organizations, fields, data types, indicators, and geographies. One approach considered tweaking the database structure using inductively defined codes as links, thereby fostering data reduction in qualitative analyses, and eventually enhancing qualitative data visualization. The key would be identifying the codes for the linking field. After much reflection, the platform staff recognized that building tools merely for the data cleaning and preprocessing tasks was beyond the scope of their envisioned functionalities.

Opting for a light approach, the platform aimed to ease data export for use in standard qualitative data analysis tools (e.g., Atlas.ti, Nvivo), as it already provides for quantitative data with Microsoft's PowerBI and Tableau. Also, a nimble search feature, such as that of the Sectoral Dashboard, could serve as a preliminary analysis tool, allowing users to explore their data and make decisions about export and analysis with third party software.

Multi-organizational database management features also define the platform's boundaries. Transitioning to a multi-organizational shared data repository, the platform needed a new control feature. The platform team, inspired in part by the Ecuador Control Dashboard, developed its own web application within the Shiny framework. The feature displays data for a selected indicator, including a completion ratio and a general distribution of values. Outstanding questions remain about which qualitative data characteristics (most recent entry, largest piece of text, etc.) are most desirable.

The third factor shaping the platform's boundaries is user and developer skills. The platform's diverse user base, especially the less skilled, creates demands for ease of use. These are met through specialized materials, such as a guide for textual data. Less skilled users are more likely to want simple data entry processes and the ability to conduct analyses on the platform. In contrast, higher skilled users are more likely to demand flexible extraction scripts or APIs to facilitate use of specialized third party software. They are also likely to need mechanisms for sharing scripts, such as those for automating data cleaning, analytic processes, and visualizations.

The scripts, and underlying information management skills, also shape the platform's boundaries by relieving pressure for new features. Staff who write and publicly post scripts, for example through GitHub, can provide functionality to the community of users. This is particularly helpful given the idiosyncratic nature of humanitarian response, alleviating the need for platform-based 'global solutions.' The platform can engage in script management or simply leave it to the user community. For high-value scripts platform staff translate from one programming language to another (e.g., R to Python).

Lastly, the platform's boundaries are influenced by regional automation and standardization in data structures. To facilitate automation, regional staff were working on templates for visualizations at the sector, national and regional levels. Standardization of visualizations is aided by associated standardization of national and regional indicators, which had not yet occurred but was planned. Automation and standardization are tied into UNHCR's embrace and use of Microsoft's PowerBI for its internet-based dashboards. Together, automation and standardization, as well as use of PowerBI, help establish a clear boundary for the platform. The platform design can steer clear of replicating those functionalities while easing the flow of data between the two. At the same time, automation and standardization homogenize platform user needs, as the diverse requirements of multiple country teams evolve into greater uniformity.

5 DISCUSSION

Our exploratory project entailed analyses of the full process for qualitative data integration, from the database form design to analytic processes, and analyses of the final information products. The project's outcomes present several positive developments for qualitative data use. Users of humanitarian information products embrace the combination of quantitative and qualitative data and appreciate the transparency qualitative data lends to quantitative outputs. Users had clear requirements for qualitative data defined to overcome shortcomings in quantitative processes. Dashboards providing high level and granular views of qualitative data created value for coordination and insight into sector-specific operations. Also, the various software elements created through humanitarian and platform staff collaboration generate mutual benefit for furthering qualitative data use beyond the scope of our project.

Yet, our analysis also uncovered two broader and ongoing challenges. First, aligning the qualitative and quantitative indicators to ease automated processing (driven by resource constraints), may have influenced the qualitative data's structure and content, to the detriment of automated analysis. Our solution, embodied in the Sectoral Dashboards, is a compromise, leaving the full value of qualitative data for humanitarian operations an open question. Second, while platforms enable decentralized data entry, allowing rich qualitative data to be entered closer to its context, the need for control and validation favor centralized entry. In general, processes of aggregation and related centralization risk losing valuable contextual insights and silencing local experts.

As discussed below, our findings inform humanitarian informatics theory and practice in the areas of agency, innovation and platform boundaries, the value of advanced analytics, and for broader digitalization of humanitarian operations.

5.1 Quantitative Analytics' Agency

Throughout the project, *quantitative* data, together with associated analytic tools, workflows and human skills, exerted agency. This was partly by design, as we sought to create complementary data uses. However, the quantitative system's status as the installed base of data management [21], generated a deep entanglement with workflows, technologies, and standard forms of reporting, giving that system an outsized and systematic influence. The dynamic and interactive nature of the agency [25] positioned the system as a hegemonic force [27]. While specific to this instantiation of the platform, the design of the qualitative system, and this unique humanitarian context [26], the agency of embedded and emerging data systems is observed across the humanitarian community [5].

For our project, the quantitative system's agency has three visible effects: **framing, artifacts, and informing**. Framing reflects the quantitative system's influence on the research team's and other's expectations for the qualitative system. The quantitative system, its automation and efficiencies, became the yardstick against which many of the qualitative data practices are measured. Automated analyses, summarizing charts, and the presumption of shared interpretation, such as are feasible with quantitative data, explicitly and implicitly serve as the gold standard.

As previously stated, this agency was amplified by our complementary system design, as compared to a purely greenfield qualitative system design. Regardless, a greenfield system embedded in the humanitarian context would nearly always face the resource constraints favoring the efficiencies and ease of aggregation and centralization quantitative systems afford. In this context, the traditionally textual and only partially automated SitRep stands apart.

This framing influenced our perceived purpose of the system, and in turn, the actual structure of the qualitative data and associated data analytic tools and processes – the system's artifacts. The quantitative system's agency, exercised through the monitoring framework, influenced the structure and content of the qualitative data. In turn, this lowered the value derived from established automated qualitative data analytic tools. Subsequently, the system's affordances were re-

conceptualized to offer value within a more limited scope (e.g., offering high level insight, visibility, and transparency, rather than deep insight potentially gained from semantic analyses).

Finally, the quantitative system's agency influences the potential for informing staff and the public. In response to our proposed information product formats, which integrate text and infographics, subjects voiced concerns that infographics might be difficult for some audiences to comprehend. The text-based SitRep is still the most widely accessed and familiar report, although its primacy is being challenged by quantitative dashboards. Our proposed integration of qualitative and quantitative outputs was seen by the research team to enhance accessibility, in comparison with dashboards. The pushback and concerns they voice, echo others' fears that digital technologies, whether in the form of infographics or AI, exacerbate disconnects between aid agencies and affected communities [5].

5.2 Platform Boundaries and Collaborative Innovation

We contribute to scholarship on platform boundaries by identifying direct and indirect factors influencing their evolution. Through our project, the platform's boundaries expanded, integrating new features, including an updated data export tool, a data control or management feature, and textual data search. Several other features that may have expanded the boundaries further, were considered and dismissed. Both the adoption and dismissal occurred within the collaborative process, where applications developed by humanitarian staff influenced the platform and vice versa.

This collaborative innovation process had direct implications for boundary spanning and reinforcing, with involvement of the platform's ecosystem [30], [35], [39]. For this hybrid data sharing portal, issues of decision rights were rarely a factor in managing boundaries. Humanitarian staff developed the tools (dashboards) they required, customized to their context. These innovations helped shape those of the platform. Features transferred from the humanitarians' applications to the platform's architecture benefitted the former by relieving them of the need to maintain and update the software. They, as well as the keystone firm [30], took pride in knowing their efforts would benefit humanitarian operations in other countries.

This process is likely unique to the context of a hybrid data sharing platform, where application development by humanitarian staff is not for commercial gain. Instead, the applications, developed and shared via Github, are designed to achieve internal information management goals. Our project's context is further defined as one with a for-profit firm providing services to non-profit humanitarian organizations, its exclusive target market. The context supported collaborative innovation and the coevolution of the complementary technologies together with the platform's core architecture [32], [35], [40].

Boundary expansion through collaborative innovation may also be influenced by the information sharing context, which shapes the functionality of data sharing platforms [36], [44]. As a platform with a closed user base and trusted data sharing, the humanitarian user community and platform firm work together to enhance the efficiency of humanitarian operations. Further, the operational data, unlike its case management, refugee registration or digital identity counterparts, are readily shareable because they do not include sensitive personally identifiable information (PII). This ease of data sharing, wherein system developers can test, display, and gather feedback on results without fear of sensitive information disclosure aids in the collaborative process.[10]

We also identify indirect factors shaping the platform's boundaries, namely skills and standardization. Our sociotechnical systems framing afforded a comprehensive perspective of the platform ecosystem, including the organizations, staff, and broader technical environment. For all platforms, the skills of module developers have

[10] Importantly, it also aids the involvement of academics in research on such systems.

implications for the platform. In particular, [69] identify skills as a component affecting the 'distributed tuning' of boundary resources, a dynamic process of interdependent evolution among different forms of boundary resources. Their analysis focuses on Apple's service system, a for-profit context. Our analysis extends their findings, demonstrating skills are a force shaping boundary resources, and in turn, platform boundaries, in a non-profit and collaborative context. We highlight that a *deficit* of skills can create pressure for boundary expansion, adding easy-to-use features difficult for users to implement through third party software, despite their availability. Conversely, *availability* of skills applied to developing and publicly distributing applications or modules, can allow a platform to maintain its boundaries, as user needs are met through applications.

Finally, standardization also shapes the boundaries of data sharing platforms. Internal features of standardization, such as templated data formats and analytics [10], can ease processing within the platform. Standardization can extend externally as well, easing the flow of data into and out of the platform, with implications for boundaries. We echo [44]'s finding of the importance of standards, in their case in the context of datasets for open government data platforms. We extend their findings by highlighting their effects on boundaries. Their analysis includes both format and quality standards, while ours in concerned only with the former. Future research on humanitarian data portals should consider the effects of data quality standards for platform features and boundaries.

5.3 Questioning the Inspiration for and Value of Advanced Analytics

The successful development of the Sectoral and Control Dashboards, with their greater reliance on descriptive statistics, calls into question the value of advanced analytics. Initial expectations for semantic analyses were primed, in part, by analytic results of large, humanitarian social media data sets, although outside the operational realm [70], [71]. Our operations data, with its significantly different structure and content, created challenges for applying traditional automated text analyses. However, our compromise of providing more descriptive statistics and access to the raw data met an unstated need in organizational users. They valued direct access to the qualitative data, not typically available with highly processed results.

Our experience highlights the importance of basing expectations for advanced analytics on *similarly structured data*, rather than industry context alone. For instance, humanitarians might look to systems for analyzing doctor-entered clinical data, which has similarities including data entry by human experts, derived from a work environment, jargon ladened, and unstructured [72]. Medical text analytics are improving rapidly [73].

The value attributed to advanced analytics must also consider its effects on labor. Our research finds 50% of our subjects view qualitative data entry as more burdensome than quantitative forms. Respondents also expressed concerns about the amount of ongoing work the new information products required. These findings echo those of previous research where 'data for good' systems create nearly insatiable demands for their management and care [9], [58]. These sentiments suggest the need for systematic task and workload analyses, akin to environmental impact assessments, prior to the introduction of new analytic and data management processes [74]. They also demonstrate issues of concern to labor in data science [14], [15] also extend to less technical realms of analytics in non-profit contexts.

5.4 Qualitative Data and Platforms in Humanitarian Systems

Qualitative data, with its link to context-based knowledge, can potentially increase the value of knowledge embedded in field operations, typically geographically distant from the power centers of the Global North. Currently, humanitarian data systems, with their emphases on quantitative data and dashboards, reflect power structures associated with data 'extraction,' aggregation, and flows from the Global South to the Global North [9], [11], [18]. Such structures fly in the

face of humanitarian reform, the localization agenda, that emphasizes local engagement, local NGOs, and devolving power and authority to the field and local organizations [7]. Our project's shortcomings, resulting from tying qualitative to quantitative data management structures, may not only be due to the hegemony of quantitative systems but also the structures of humanitarianism.

The implications of our data sharing platform project for Global South innovation are unclear. It bears similarities to DHIS2, which [38] characterizes as 'not particularly generative' (p.8) in that applications developed to solve local problems are rarely transformed into generic apps, for wider use or sale. Knowledge of DHIS2 does generate consulting opportunities and expands availability of database technologies. In this respect, the closed nature and centralized architecture of our data sharing platform make it less likely to extend database access and new livelihood opportunities.

Similar to DHIS2, our project resulted in bidirectional (North/South) flows of knowledge and expertise. Our team, consisting of an academic and firm located in the Global North, and humanitarian staff, both local and from the Global North, but located in the Global South, suggests North/South distinctions in humanitarian contexts can be messy. The beneficiaries of the innovations and co-created knowledge are also complex. Innovations to the platform architecture economically benefit the Northern firm, but practically benefit the user base in the Global South. Local staff benefitted economically through enhanced employment prospects, with one subsequently sharing her expertise by taking a position in the Ecuadorian government, and another transitioning from a temporary contract to a longer-term, internal status with the UN. While these developments fall short of Global South innovation envisioned in the 'platform for development' literature, they do constitute important benefits of mutual development of expertise and systems. Future analyses of Global South innovation in the humanitarian context might head the call of [7] and more systematically unpack the international – local dichotomy as well as unpack innovation to understand the relative contribution of co-created expertise.

5.5 Recommendations

Finally, we briefly turn to concrete recommendations for implementing qualitative data systems, in non-profit or humanitarian systems, involving a data sharing platform. In terms of the systems development process, we recommend keeping the qualitative data analytics design process close to its data sources and potential users. This enables accessing the subject matter expertise for generating insights from qualitative data and identifying user requirements for data. In turn, 'data user requirements' should influence cleaning and preprocessing workflows. Systems designs should remain flexible concerning the scope of the data corpus, balancing the analyses' needs for volumes and ability to handle a broader range of sector-based jargon, with implications for developing universal versus narrow insights.

Our research generated positive outcomes for humanitarian operations by providing coordination bodies and sector leads with insight into the operations as well as vetted suggestions for potential report formats. From these results we offer the following design recommendations. First, multi-organizational data systems benefit from control information in the form of a dashboard, providing the potential to monitor and shape the data entry process, particularly for qualitative data. Also, database forms should tailor prompt openness and entry fields to individual measures. Further, systems should allow access to raw data, allowing for questioning of results and subsequent manual analyses. Finally, information product formats combining qualitative and quantitative formats can present a middle ground between accessible text and insight gleaned from infographics. To further accessibility, text-based reporting systems can be designed to flag jargon and difficult language assisting humanitarian workers to reach a broad audience, especially affected communities.

Future designs of humanitarian qualitative systems might attempt a looser coupling with established quantitative systems in order to reap qualitative data's benefits more fully. However, the extent of coupling must critically view sustainability and the costs of maintaining separate or even adjacent systems.

6 CONCLUSION

Rapid advances in machine learning for textual data are generating new applications, including in the humanitarian domain. Our exploratory project investigating the potential for these technologies to generate insight for humanitarian operations, has implications for theory and practice. We contribute to theory on platform boundaries, data platforms and critical data studies for humanitarian relief and development. Working with a global platform provider to the humanitarian community, the outcomes and findings of this project have international impact for humanitarian operations. The platform innovations developed herein are now (as of March 2022) available to humanitarian organizations operating in 34 countries.

Explorations of machine learning and artificial intelligence in humanitarian operations continue. These technologies are being applied to detect xenophobic attitudes towards forced migrants,[11] to analyze outcomes of refugees' phone-based requests for assistance, and to provide support in the form of chatbots [5]. These applications undoubtedly will generate both positive and negative effects, for humanitarian workers as well as affected communities. Informed by projects such as ours, such systems can aim to ameliorate negative and amplify positive outcomes, to improve operations and facilitate broad and accessible information sharing.

ACKNOWLEDGMENTS

This research was funded through UNHCR's Innovation Fund for 2019 and 2020 (https://www.unhcr.org/innovation/innovation-fund/). The authors would like to recognize the contributions of Mariá José Alvear of HIAS, Maarten-Jan Kallen and Metin Yazici (formerly) of BeDataDriven, and Capucine Maus de Rolley and Mirely Tobar, formerly with UNHCR.

REFERENCES

[1] K. Weitzberg, M. Cheesman, A. Martin, and E. Schoemaker, "Between surveillance and recognition: Rethinking digital identity in aid," *Big Data Soc.*, vol. 8, no. 1, p. 1-7, 2021.

[2] R. Talhouk *et al*., "Food Aid Technology: The Experience of a Syrian Refugee Community in Coping with Food Insecurity," *Proc. ACM Human-Computer Interact.*, vol. 4, no. CSCW2, 2020, doi: 10.1145/3415205.

[3] A. Irani, K. Nelavelli, K. Hare, P. Bondal, and N. Kumar, "Refuge tech: An assets-based approach to refugee resettlement," in *Extended abstracts of the 2018 chi conference on human factors in computing systems*, 2018, pp. 1–6.

[4] R. Read, B. Taithe, and R. Mac Ginty, "Data hubris? Humanitarian information systems and the mirage of technology," *Third World Q.*, vol. 37, no. 8, pp. 1314–1331, 2016.

[5] M. Madianou, "Nonhuman humanitarianism: when 'AI for good' can be harmful," *Inf. Commun. Soc.*, vol. 24, no. 6, pp. 850–868, 2021, doi: 10.1080/1369118X.2021.1909100.

[6] G. Coppi, R. Moreno Jimenez, and S. Kyriazi, "Explicability of humanitarian AI: a matter of principles," *J. Int. Humanit. Action*, vol. 6, no. 1, pp. 1–22, 2021.

[7] K. Roepstorff, "A call for critical reflection on the localisation agenda in humanitarian action," *Third World Q.*, vol. 41, no. 2, pp. 284–301, 2020.

[8] K. Pincock, A. Betts, and E. Easton-Calabria, "The rhetoric and reality of localisation: refugee-led organisations in humanitarian governance," *J. Dev. Stud.*, vol. 57, no. 5, pp. 719–734, 2021.

[11] https://www.unhcr.org/innovation/wp-content/uploads/2018/01/White20Paper20Social20Media203_0.pdf

[9] M. Finn and E. Oreglia, “A fundamentally confused document: Situation reports and the work of producing humanitarian information,” in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 1349–1362.

[10] D. Huser, A. Bon, and A. Anifalaje, “Uncovering generative mechanisms of information use for project monitoring in humanitarian health management information systems,” *Electron. J. Inf. Syst. Dev. Ctries.*, no. April, pp. 1–18, 2021, doi: 10.1002/isd2.12184.

[11] S. Madon and E. Schoemaker, “Digital identity as a platform for improving refugee management,” *Inf. Syst. J.*, no. October 2019, pp. 1–25, 2021, doi: 10.1111/isj.12353.

[12] C. S. Wang, S. L. Lin, T. H. Chou, and B. Y. Li, “An integrated data analytics process to optimize data governance of non-profit organization,” *Comput. Human Behav.*, vol. 101, no. September 2018, pp. 495–505, 2019, doi: 10.1016/j.chb.2018.10.015.

[13] M. Muller *et al.*, “Human-centered study of data science work practices,” *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1–8, 2019, doi: 10.1145/3290607.3299018.

[14] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, “‘Everyone wants to do the model work, not the data work’: Data Cascades in High-Stakes AI,” pp. 1–15, 2021, doi: 10.1145/3411764.3445518.

[15] A. X. Zhang, M. Muller, and D. Wang, “How do Data Science Workers Collaborate? Roles, Workflows, and Tools,” *Proc. ACM Human-Computer Interact.*, vol. 4, no. CSCW1, pp. 1–23, 2020, doi: 10.1145/3392826.

[16] C. L. Aasheim, S. Williams, P. Rutner, and A. Gardiner, “Data analytics vs. data science: A study of similarities and differences in undergraduate programs based on course descriptions,” *J. Inf. Syst. Educ.*, vol. 26, no. 2, pp. 103–115, 2015.

[17] Z. Radovilsky, V. Hegde, A. Acharya, and U. Uma, “Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis,” *J. Supply Chain Oper. Manag.*, vol. 16, no. 1, p. 82, 2018.

[18] M. Madianou, “Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises,” *Soc. Media + Soc.*, vol. 5, no. 3, p. 1-13, Jul. 2019, doi: 10.1177/2056305119863146.

[19] W. J. Orlikowski and C. S. Iacono, “Research commentary: Desperately seeking the ‘IT’ in IT research—A call to theorizing the IT artifact,” *Inf. Syst. Res.*, vol. 12, no. 2, pp. 121–134, 2001.

[20] R. Kling and W. Scacchi, “The web of computing: Computer technology as social organization,” *Adv. Comput.*, vol. 21, pp. 1–90, 1982.

[21] G. C. Bowker and S. L. Star, *Sorting things out: Classification and its consequences*. MIT press, 2000.

[22] M. Aanestad, M. Grisot, O. Hanseth, and P. Vassilakopoulou, “Information infrastructures and the challenge of the installed base,” *Inf. infrastructures within Eur. Heal. care*, pp. 25–33, 2017.

[23] H. Shen, C. Faklaris, H. Jin, L. Dabbish, and J. I. Hong, “‘I Can’t Even Buy Apples If I Don’t Use Mobile Pay?’ When Mobile Payments Become Infrastructural in China,” *Proc. ACM Human-Computer Interact.*, vol. 4, no. CSCW2, pp. 1–26, 2020.

[24] W. J. Orlikowski, “Sociomaterial practices: Exploring technology at work,” *Organ. Stud.*, vol. 28, no. 9, pp. 1435–1448, 2007.

[25] D. Cecez-Kecmanovic, R. D. Galliers, O. Henfridsson, S. Newell, and R. Vidgen, “The sociomateriality of information systems,” *MIS Q.*, vol. 38, no. 3, pp. 809–830, 2014.

[26] W. J. Orlikowski, “The sociomateriality of organisational life: considering technology in management research,” *Cambridge J. Econ.*, vol. 34, no. 1, pp. 125–141, 2010.

[27] C. Bopp, L. M. Benjamin, and A. Voida, “The coerciveness of the primary key: Infrastructure problems in human services work,” *Proc. ACM Human-Computer Interact.*, vol. 3, no. CSCW, pp. 1–26, 2019.

[28] R. Crooks and M. Currie, “Numbers will not save us: Agonistic data practices,” *Inf. Soc.*, pp. 1–19, 2021.

[29] P. Dourish and E. Gómez Cruz, “Datafication and data fiction: Narrating data and narrating with data,” *Big Data Soc.*, vol. 5, no. 2, pp. 1–10, 2018, doi: 10.1177/2053951718784083.

[30] M. De Reuver, C. Sørensen, and R. C. Basole, “The digital platform : a research agenda,” pp. 124–135, 2018, doi: 10.1057/s41265-016-0033-3.

[31] V. Dal Bianco, V. Myllärniemi, M. Komssi, and M. Raatikainen, “The role of platform boundary resources in software ecosystems: a case study,” in *2014 IEEE/IFIP Conference on Software Architecture*, 2014, pp. 11–20.

[32] C. Leong, S. L. Pan, D. E. Leidner, and J.-S. Huang, “Platform leadership: Managing boundaries for the network growth of digital platforms,” *J. Assoc. Inf. Syst.*, vol. 20, no. 10, p. 1, 2019.

[33] M. A. Cusumano, D. B. Yoffie, and A. Gawer, “The future of platforms,” *MIT Sloan Manag. Rev.*, vol. 61, no. 3, pp. 46–54, 2020.

[34] M. De Reuver, C. Sørensen, and R. C. Basole, “The digital platform: A research agenda,” *J. Inf. Technol.*, vol. 33, no. 2, pp. 124–135, 2018, doi: 10.1057/s41265-016-0033-3.

[35] A. Tiwana, B. Konsynski, and A. A. Bush, “Research commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics,” *Inf. Syst. Res.*, vol. 21, no. 4, pp. 675–687, 2010.

[36] B. Otto and M. Jarke, “Designing a multi-sided data platform: findings from the International Data Spaces case,” *Electron. Mark.*, vol. 29, no. 4, pp. 561–580, 2019, doi: 10.1007/s12525-019-00362-x.

[37] M. Schreieck, M. Wiesche, and H. Krcmar, “Governing nonprofit platform ecosystems–an information platform for refugees,” *Inf. Technol. Dev.*, vol. 23, no. 3, pp. 618–643, 2017.

[38] B. Nicholson, P. Nielsen, J. Saebo, and S. Sahay, “Exploring tensions of global public good platforms for development: The case of DHIS2,” *IFIP Adv. Inf. Commun. Technol.*, vol. 551, pp. 207–217, 2019, doi: 10.1007/978-3-030-18400-1_17.

[39] S. K. Jha, A. Pinsonneault, and L. Dub, “The evolution of an ICT platform-enabled ecosystem for poverty alleviation: The case of eKutir,” MIS Quarterly, vol. 40, no. 2, pp. 431-445/June 2016.

[40] P. Constantinides, O. Henfridsson, G. G. Parker, P. Constantinides, O. Henfridsson, and G. Parker, “Introduction — Platforms and Infrastructures in the Digital Age,” *Inf. Syst. Res.*, vol. 29, no. July 2019, pp. 381–400, 2018.

[41] C. Bonina, K. Koskinen, B. Eaton, and A. Gawer, “Digital platforms for development: Foundations and research agenda,” *Inf. Syst. J.*, no. January, pp. 1–34, 2021, doi: 10.1111/isj.12326.

[42] B. Nicholson, P. Nielsen, and J. Saebo, “Digital platforms for development,” *Information Systems Journal*, vol. 31, no. 6. Wiley Online Library, pp. 863–868, 2021.

[43] R. Heeks *et al.*, “Digital platforms and institutional voids in developing countries: The case of ride-hailing markets,” *World Dev.*, vol. 145, p. 105528, 2021.

[44] C. Bonina and B. Eaton, “Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo,” *Gov. Inf. Q.*, vol. 37, no. 3, p. 101479, 2020, doi: 10.1016/j.giq.2020.101479.

[45] S. S. Dhruva *et al.*, “Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform,” *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–9, 2020.

[46] A. E. Abbas, W. Agahari, M. van de Ven, A. Zuiderwijk, and M. de Reuver, “Business Data Sharing through Data Marketplaces: A Systematic Literature Review,” *J. Theor. Appl. Electron. Commer. Res.*, vol. 16, no. 7, pp. 3321–3339, 2021.

[47] D. E. Hellmann, “The Generative Effect of Expertise and Articulation Work in Organic ICT Based Collaborative Communities of Practice,” Doctoral Dissertation, The Pennsylvania State University, 2018.

[48] D. Hellmann, C. Maitland, and A. Tapia, “Collaborative analytics and brokering in digital humanitarian response,” in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 1284–1294.

[49] M. M. Vila-Pozo and S. Sahay, “Institutional shaping of affordances: Implications on information use in global humanitarian organizations,” in *International Conference on Social Implications of Computers in Developing Countries*, 2019, pp. 496–507.

[50] A. Chrysantina and J. I. Sæbø, “Assessing user-designed dashboards: A case for developing data visualization competency,” in *International Conference on Social Implications of Computers in Developing Countries*, 2019, pp. 448–459.

[51] R. Heeks, C. Foster, and Y. Nugroho, “New models of inclusive innovation for development,” *Innovation and Development*, vol. 4, no. 2. Taylor & Francis, pp. 175–185, 2014.

[52] A. Adadi, “A survey on data-efficient algorithms in big data era,” *J. Big Data*, vol. 8, no. 1, pp. 1–54, 2021.

[53] G.-J. Qi and J. Luo, “Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods,” *IEEE*

Trans. Pattern Anal. Mach. Intell., 2020.

[54] L. Hagen *et al.*, "How can machine learning aid behavioral marketing research?," *Mark. Lett.*, vol. 31, no. 4, pp. 361–370, 2020.

[55] J. Berger, A. Humphreys, S. Ludwig, W. W. Moe, O. Netzer, and D. A. Schweidel, "Uniting the tribes: Using text for marketing insight," *J. Mark.*, vol. 84, no. 1, pp. 1–25, 2020.

[56] C. Stepniak, "The role of qualitative data in the management of enterprises," *Proc. 28th Int. Conf. Inf. Syst. Dev. Inf. Syst. Beyond 2020, ISD 2019*, 2019.

[57] S. Suvivuo, "Qualitative Big Data's Challenges and Solutions: An Organizing Review," *Proc. 54th Hawaii Int. Conf. Syst. Sci.*, vol. 0, pp. 980–989, 2021, doi: 10.24251/hicss.2021.119.

[58] P. Sanches and B. Brown, "Data Bites Man: The Production of Malaria by Technology," *Proc. ACM Human-Computer Interact.*, vol. 2, no. CSCW, pp. 1–19, 2018.

[59] Y. Xu and C. Maitland, "Participatory data collection and management in low-resource contexts: a field trial with urban refugees," in *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, 2019, pp. 1-12 .

[60] M. Cheesman, "Self-sovereignty for refugees? The contested horizons of digital identity," *Geopolitics*, vol. 27, no. 1, pp. 134–159, 2022.

[61] G. Iazzolino, "Infrastructure of compassionate repression: making sense of biometrics in Kakuma refugee camp," *Inf. Technol. Dev.*, vol. 27, no. 1, pp. 111–128, 2021.

[62] A. Martin and L. Taylor, "Exclusion and inclusion in identification: Regulation, displacement and data justice," *Inf. Technol. Dev.*, vol. 27, no. 1, pp. 50–66, 2021.

[63] M. Madianou, "The Biometric Assemblage: Surveillance, Experimentation, Profit, and the Measuring of Refugee Bodies," *Telev. New Media*, vol. 20, no. 6, pp. 581–599, 2019, doi: 10.1177/1527476419857682.

[64] V. Braun and V. Clarke, *Successful qualitative research: A practical guide for beginners*. Sage, 2013.

[65] J. W. Creswell and D. L. Miller, "Determining validity in qualitative inquiry," *Theory Pract.*, vol. 39, no. 3, pp. 124–130, 2000.

[66] R. K. Yin, "Case study methods." In H. Cooper, P.M. Camic, D.L. Long, A.T. Panter, D. Rindskopf, & K.J. Sher (Eds), APA handbook of research methods in psychology, vol. 2. Research Designs: Quantitative, qualitative, neuropsychological, and biological. pp. 141-155, 2012.

[67] J. M. Chevalier and D. J. Buckles, *Participatory action research: Theory and methods for engaged inquiry*. Routledge, 2019.

[68] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents," *Text Min. Appl. theory*, vol. 1, pp. 1–20, 2010.

[69] B. Eaton, S. Elaluf-Calderwood, C. Sorensen, and Y. Yoo, "Distributed tuning of boundary resources: the case of Apple's iOS service system," *MIS Q. Manag. Inf. Syst.*, vol. 39, no. 1, pp. 217–243, 2015.

[70] G. Müller-Stewens, T. Dinh, B. Hartmann, M. J. Eppler, and F. Bünzli, "Humanitarian Aid 2.0: Social Media Analytics and Stakeholder Engagement at the International Committee of the Red Cross," in *The Professionalization of Humanitarian Organizations*, Springer, 2019, pp. 65–81.

[71] M. Miranker and A. Giordano, "Text mining and semantic triples: Spatial analyses of text in applied humanitarian forensic research," *Digit. Geogr. Soc.*, vol. 1, 100005, pp. 1-10, 2020.

[72] I. Spasic and G. Nenadic, "Clinical text data in machine learning: systematic review," *JMIR Med. informatics*, vol. 8, no. 3, p. e17984, 2020. doi: 10.2196/17984.

[73] F. Shen *et al.*, "Family history extraction from synthetic clinical narratives using natural language processing: overview and evaluation of a challenge data set and solutions for the 2019 National NLP Clinical Challenges (n2c2)/Open Health Natural Language Processing (OHN," *JMIR Med. Informatics*, vol. 9, no. 1, p. e24008, 2021. doi: 10.2196/24008 .

[74] J. Foster, J. McLeod, J. Nolin, and E. Greifeneder, "Data work in context: Value, risks, and governance," *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 12, pp. 1414–1427, 2018.

A APPENDICES

A.1 Qualitative Text Examples

Table A.1:Examples of Qualitative Text from the Livelihoods and Education Sectors

Sector	Typical short	Typical long
Livelihoods	"9 women and 2 men attended for the first time employment training; 3 women and 1 man and in follow-up 2 men and 3 women participated in employability workshops"	"[organization name] has adapted tools to work with migrant, refugee and local populations on entrepreneurship and employability issues. So far, 88 people have been trained in employability issues (37 foreigners and 51 locals) and 76 people in business ideation (entrepreneurship). The methodology includes an initial diagnosis of the profiles in order to know and guide the beneficiaries in the best way. It works with beneficiaries that are already established in the country."
Education	"Guidance on education issues has been provided."	"Because the public school system is closed by the end of the educational year there is a significant percentage of children out of school. In that context, the families have been provided with guidance so that they can go to the Education District with the documents of the minors and can advance the allocation of the level of education that corresponds to the next school year. As the same time, in cases that do not have any documents, they have been informed so that they can start the location testing process. On May 14, an informational workshop on access to education for people in human mobility was held with the Education District. In this space, the families presented the documents of their children and the representative of the District gave them the guidance on whether they

		were valid and the route to follow so that they can continue with the schooling process. Access to child development services is limited so from [government ministry] it has been indicated that the identification of cases will be carried out directly when more quotas are opened.”

A.2 Sample Characteristics

Table A.2.1: Input Analysis Sample Characteristics

Partner	Partners w/more narrative info	Biggest partners (by # services provided)	Partners w/more sector coverage	Partners participating in sectors with few actors
ACNUR	X	X	X	X
HIAS	X	X	X	X
NRC	X	X	X	
OIM	X	X	X	X
UNICEF	X	X	X	
CARE		X	X	
JRS Ecuador			X	
Diálogo Diverso				X
Misión Scalabriniana				X

Table A.2.2: Input Analysis Interviewees’ Organizations and Positions

Organization / # staff interviewed	Interviewees’ Positions (FO = Field Office)
HIAS (3)	Quito South FO Manager, Quito Programme Assistant, Venezuela Situation Specialist
ACNUR (2)	Lago Agrio FO Programme Associate, Esmeraldas FO Programme Associate
UNICEF (1)	National Emergency Officer
Diálogo Diverso (1)	Information and Reference Center Coordinator
IOM (3)	Interagency Coordination Assistant, M&E Assistant, Information Management Officer
CARE (2)	Humanitarian Assistant, M&E Officer
JRS (1)	Monitoring and Follow Up Officer
Misión Scalabrinia (1)	National Advocacy Coordinator
NRC (1)	M&E Officer

Table A.2.3: Information Product Analysis Sample

Categories	Organizations	Position
Smaller Partners (3)	Alas de Colibrí	- Project General Coordinator
	Misión Scalabriniana	- Executive Director
	CISP	- Coordination Assistant
Larger Partners (2)	Jesuite Refugee Service	- Project Follow-Up Technicians
	PAHO/WHO	- Coordination Assistant
Specific working groups and subroups (2)	GBV and Human Trafficking Protection Subgroup	- Sector lead
	Education Working Group	- Sector lead
Local partners (3)	AVSI	- Legal Representative (Quito)
	Ayuda en Acción	- Human Mobility Supervisor (Ibarra)
	CDH	- Executive Secretary (Guayaquil)
Local GTRM interviews (10)	UNFPA	- Southern Border Social and Reproductive, GBV Response Coordinator (Huaquillas)
	HIAS	- Programme Analyst (Quito) - Head of Field Office (Tulcán)
	COOPI	- Project Coordinator (Machala)
	Fundación Tarabita	- Coordinator (Lago Agrio)
	Fundación Lunita Lunera	- Psicosocial Worker (Esmeraldas)
	FEPP	- Community Worker (Esmeraldas)
	FUDELA	- Community Worker (Esmeraldas)
	Plan International	- Project Analyst and Focal Point (Huaquillas)
	GIZ	- Local Advisor (Ibarra)
	PMA	- Field Specialists (Lago Agrio)
Local authorities (5)	Public Health Ministry	- 4th District Supervisor (Tulcán - San Pedro de Huaca)
	Municipality of Tulcán	- International Cooperation Advisor (Tulcán)
	Human Rights Protection Cantonal Council (Milagro)	- Executive Secretary (Milagro)
	Ministry of Foreign Affairs and Human Mobility	- 8th Zone Coordinator (Guayaquil)
	Public Defender	- Provincial Public Defender (Ibarra)

A.3 Information Products Likes/Dislikes

Table A.3.1: Established Information Product Likes and Dislikes

	Likes (# of mentions)	Dislikes (# of mentions)
SitRep	overall design and content (6), complete/appropriate summary of the overall effort (4), meets my information needs (4), clear and understandable (3), narrative is fine (3),	hard to read with so little time (7), lacking information (e.g., age, gender)(4), not agile or dynamic (4), difficult to share with external audiences (3),

	identifies gaps and priorities (3), trustworthy information (2), timely information (2)	lacks specificity on one time or ongoing assistance (2)
Operational Report	executive presentation makes it easier to read (4), it provides a good summary of what is being done (3), it identifies gaps and priorities (3), factual and precise (2), it provides useful information (2), and its overall design and content (2)	doesn't respond to outsiders' information needs (5), too much text (4), it provides more than I need (I only read my part) (3), there is insufficient information on the gaps (2), and there is a lack of explaining the origin of the data (2).
Intersectoral Dashboard	Cross tabs and multiple selections (8) Pertinent information (5) Maps allow geographical view (4) Interactive (4) Overall design and content (4) Good summary/synthesis of partners' activities (3) Easy to understand (2) Factual and precise (2) Allows to evaluate progress (2)	Requires basic knowledge/time to understand (5) Overwhelming/too much information (3) My organization is not represented (3) Disaggregation is missing (2) My organization is not represented because we do not provide direct assistance (2)
Sectoral Dashboard	Useful information (5) Complete sector information (a 'zoom' in of the intersectoral dashboard) (4) Overall design and content (3)	Disaggregation is missing gender and age My organization is not represented No difference between direct and indirect assistance

Table A.3.2: Selected Likes and Dislikes Comparing Integrated Static and Integrated Interactive Formats

	Likes (# of mentions)	Dislikes (# of mentions)
Static hybrid	Good balance of quantitative/qualitative information (6) Graphs and numbers make it more understandable (5) Easier to read and go through data (3) Eye catching (3) Useful information (2) Allows comparisons between sectors (2)	Looks like it would take a lot of work to produce (2) Difficult to share with refugees/migrants and host communities (1)
Interactive hybrid	Descriptive elements with stats (6) Eye catching (4) Intuitive and useful (3) Dynamic and pertinent information with few clicks (3)	Duplicative (Qualitative data is available in other reports)(1)
Sectoral Dashboard	Covers expectations for own/other organizations' reports – practical and provides visibility (5) Valuable detailed data (4) Very useful information (3) Maps, graphs and charts save time (3) Interesting to see activities of all partners (2) Motivate partners to report better (1) Easy to share link (1)	Too much text, long sentences (1) Needs standards (1) Needs validation (1) Complicated to adapt when an organization leaves (1) Looks like a lot of work to produce (1)