

## Small packages for Big (qualitative) Data

---

KIM ERWIN and THEODORE POLLARI

IIT Institute of Design

*Smart devices and online research platforms are changing the landscape of qualitative data collection and analysis. While data collection mechanisms have flourished, analytic tools to work with that data have not meaningfully evolved. Changes in professional practice and advances in technology are creating new opportunities—and new pressure—to develop software tools that are focused, simple to use, fit flexibly with a variety of analytic processes, adapt to different data sets and do not lock data into proprietary formats or researchers into predefined analytic processes. We call such tools Small Packages for Big (Qualitative) Data. This paper defines the concept and introduces three such early stage tools—Voyant, Mandala Browser and Nineteen, and links qualitative research to another field experiencing similar changes and tool development, the Digital Humanities. Lastly, we present a case study to demonstrate how Small Packages can focus investigations, build early-stage familiarity with data, and inform subsequent analysis.*

### THE CHALLENGES OF BIG QUALITATIVE DATA

The qualitative research field is experiencing a confluence of factors that are collectively creating an opportunity to rethink analytic tools and their role in qualitative research. Most of today's analytic tools were developed in the 20th century, on 20th century platforms and for a context of practice that is quite different than the 21st century conditions in which many researchers practice today. Today's conditions, and the factors these authors see as relevant to software development for qualitative research, are as follows:

#### Technology-driven abundance

Field researchers are rapidly adopting smart devices, apps and internet-based tools to bring new efficiencies to their own data collection. Mobile devices, apps and wireless data transmission are also being integrated into a new class of research platforms—such as dScout, Revelation, QualVu, ethos, and others—that allow researchers to engage informants remotely through self-reporting. These tools are powerful additions to the researcher's toolkit: they open up new user groups for study, allow researchers to launch studies at a global scale and to engage study participants 24/7. However, with the power comes new data abundance.

While data abundance has always been a hazard of the profession, digital collection can quickly magnify the challenge. Now, more than ever, we have the ability to generate vast amounts of research data from more participants, in shorter time. We call this Big Qualitative Data, a reference not simply to the amount of qualitative data that can be collected and accumulated—not yet on par with the revolution in quantitative Big Data revolution—but for the complexity of that data, the speed of its accumulation and the

resulting challenge to researchers who must manage and analyze it using tools that were not designed for this scale of work.

We call out the adoption of online research platforms in particular because they are also creating a new condition for the researcher: large-scale self-reporting studies can generate potentially massive amounts of user data that is all new all at once. As a consequence, the researcher's task of building a mental model or a structured understanding of the study data can become overwhelming, making analysis more difficult, tools more important, and rigor all the more essential.

### **Tighter timeframes, limited resources**

This newfound ability to scale up qualitative data collection coincides with an increase in pressure to further reduce analysis time. While corporations are accelerating their integration of ethnographic methods into development processes (Cefkin, 2010; Rhea and Leckie, 2006; Malefyt, 2009), corporate timescales have continued to shrink (Malefyt, 2009; Thrift 2000). Compressed business cycles have been a driver of “rapid ethnography” practices, a response by the research community to better serve organizational needs and to keep qualitative researchers at the corporate table (Ladner, 2012; Cefkin, 2013; Isaacs, 2013). Another response is the adoption of mixed methods research to better ensure that research projects produce more than one type of investigative perspective and lean towards the predictive. These and other changes in qualitative research in corporate settings are driving new kinds of research problems, new users of the output (designers, strategists, planners, etc.) and new forms of relevant data. While our study approaches and data collection methods are advancing, we argue, our tools are not keeping pace.

### **New forms of data – and an opportunity**

Digital qualitative data has more predictable forms and formats. Again, we call out online research platforms as a coherent and concentrated example of this: Online platforms tend to deliver data as pre-segmented, bite-sized chunks of text, photos or short videos, rather than large bodies of textual narrative or hours of uncut video. Additionally, XML encoding and .csv files are now fairly standard. Standard formats and predictable data open the door to new computational support—the kind that quantitative analysts have always had but qualitative researchers have not.

And here we come to the pressing issue: As data sets sprawl and analysis time shrinks, analytic tools and computational support for qualitative research have not meaningfully evolved. Twenty-five years ago, mass market tools emerged that materially advanced the efficiency of researchers by bringing computer support to the research process. These tools, such as NVivo, ATLAS.ti and MAXQDA, were built as comprehensive analysis platforms, architected around a linear analytic model and a proscribed approach for organizing and analyzing qualitative data.

Today, changes in the context of practice—from advances in digital data collection tools to the rise of corporate ethnography and its demanding timeframes to growing interest in qualitative perspectives in new domains—has produced a class of practitioners who would

benefit from greater flexibility than established tools can provide. This includes the ability to tailor workflow to fit the variable nature of individual projects and team processes, as well as the means to experiment with or advance new analytic processes. This class of practitioners, and perhaps even traditional practitioners, we believe, would benefit from a more open design paradigm in software development.

## **A NEW VIEW: SMALL PACKAGES FOR BIG (QUALITATIVE) DATA**

What might an open approach to software seek to address? We propose the following: tools that fit flexibly with a variety of analytic processes; that adapt to different data sets; that are focused and simple to use; and that do not lock data into proprietary formats nor researchers into predefined analytic processes. We call this paradigm *Small Packages for Big (Qualitative) Data*. Its key attributes are as follows:

1. Modular, loosely-coupled, purpose-driven tools;
2. Visualization-driven interfaces to engage large data sets; and
3. Dynamic interaction environments for exploration and sense-making.

### **Principle 1: Modular, loosely-coupled, purpose-driven tools**

How does a Small Packages paradigm flexibly support analysis? We think of Small Packages like a nurse's toolkit: a collection of small, focused tools that require minimal training and effort to use, that have clear contexts of use and that *extend* the natural abilities of practitioners, rather than replace them. We also propose that the tools be "loosely coupled," referring to a concept in system and interface design that seeks to reduce the interdependencies of components in a system in order to increase independent functioning and create more flexible responses (Orton & Weick, 1990). Loosely-coupled tools, then, are designed to work well together, but do not require each other or necessitate a particular sequence to produce useful results.

A suite of independent, purpose-driven analytic tools that work together through common, standards-based data formats would better fit the way qualitative researchers work. Specifically, they pursue research problems using a series of independent methods and tools to investigate aspects of the problems. Tools to support this, then, should be developed to tackle individual, focused problems as well: how to visualize, organize and investigate data related by tags; how to investigate data based on language patterns; how to visualize large numbers of diary entries, etc. A suite of purpose-driven, independent tools would allow researchers to run their data through as many Small Packages as they see fit, each optimized around a particular task, and none forcing the researcher into a pre-determined process.

### **Principle 2. Visualization-driven interfaces to engage large data sets**

Visualization of qualitative data is an important attribute of Small Packages and key to its ability to aid researchers with large data sets. As datasets generated by online platforms or collected in databases grow in raw size, they can quickly surpass the ability of researchers to

apprehend in a reasonable amount of time. While software cannot increase the reading speed or working memory of an individual, it can provide tools that tap our native abilities to visually process large amounts of data with minimal conscious interpretation, so as to spot patterns more quickly (Healey, et al., 1996; Ware, 2008; Few, 2009; Tversky 2011). Quantitative expert Steven Few calls this “thinking with our eyes.” Data visualization has long been valued in the quantitative world (Bertin, 1967; Tufte, 1983; Cleveland & McGill, 1984; Slone, 2009) but its adoption in qualitative settings is much more recent (Miles & Huberman, 1994; Slone, 2009; Erwin 2011).

In qualitative analysis, the visual display of information can be an effective offset to the human tendency to “jump to hasty, partial, unfounded conclusions,” and to “overweight vivid information” when engaging large data sets (Miles & Huberman, 1994, p. 11). Objective displays of data, they argue, help researchers draw valid interpretations and take needed action (p. 91). Cognitive scientists agree: as Nickerson, et al., (2013) highlight, externalized visual representations add a measure of permanence to insights and free up cognitive resources by allowing researchers “to use working memory for inferences and mental revisions” (p. 14).

It is the “bigness” of Big Qualitative Data that opens up new opportunities to apply visualization methods developed for quantitative data to qualitative data: By using colour, size, shape, position and other visual variables, qualitative data can be visually coded. That is, units of data can be temporarily assigned visual attributes. Common variables for visual coding might include day, time, participant, activity, tagged words, etc. However, any variable collected consistently across a qualitative data set can be assigned a visual code, i.e., “store shopped at” or “media used.” Once variables are represented in a visual way—circles of a given colour, for instance—the data can be represented abstractly and compactly to fit onscreen in a unified display. Visually-coded qualitative data can then be clustered, ordered or juxtaposed to show quantity, correlations and other relationships.

### **Principle 3: Dynamic interaction environments for exploration and sense-making**

This third attribute of Small Packages advances Miles and Huberman’s central thesis that the building of the data display by the researcher is integral to the analytic progression (Qualitative Data Analysis, p. 11). It also reflects Bowen’s notion that iterative interplay between data collection and analysis is key to discovery (Bowen, 2008). Small Packages enables these actions by turning data displays into dynamic interaction environments that allow researchers to quickly configure, explore and then reconfigure their data in new ways. Stephen Few (2009) notes that technology used this way allows us to hold a “dynamic dialog between the analyst and the data.” We call this ability to quickly prototype data, “data poking.”

*Data poking* is an important concept in Small Packages, as it encourages researchers to see and touch all their data in an informal manner multiple times before engaging in deep analysis. This act of exploring and prodding the data prior to coding creates knowledge that can drive efficiencies in the formal analytic process—Given & Olson (2003), for example, advocate an upfront step aimed at the organization and preparation of data as critical to

effective analysis. Data poking is a complementary concept, focused instead on preparation of the *researcher*. Upfront explorations create familiarity with the data, raise questions about its nature, and aid the researcher in building a mental model of the dataset and its contents to carry into analysis. This kind of upfront familiarization stage is particularly important when the data has been generated via online research platforms, as the researcher may not have had direct experience of the data during its collection.

### THREE “SMALL PACKAGE” EXAMPLES

In this section, we introduce three tools that fit the Small Packages paradigm. *Voyant* and *Mandala* come from the Digital Humanities, a field that shares the qualitative researcher's interest in visually enhanced, computer-supported inquiry into large amounts of texts. The third, *Nineteen*, comes from the design research field, specifically from the two authors of this paper.

The Digital Humanities offers particularly fertile ground for qualitative researchers seeking new tools, especially tools that work with large data sets in the focused, independent manner we advocate with Small Packages. Not well known to qualitative researchers, the Digital Humanities is a field of study emerging from the introduction of computing to the humanities (a more precise definition of the Digital Humanities has not come together in a way that scholars in the field can agree on; however, few disagree with this bare bones description). As with qualitative research, the introduction of computing has changed humanities scholarship in important ways. The first is the creation of large digital collections (texts, images, video and even artefacts) that offer unprecedented quantities of sometimes hard-to-access material for investigation. For example, the Medici Archive project is an exhaustive collection of the courtly archives of early modern Europe, currently over four-million letters occupying 6,429 volumes and a mile of shelf-space in printed form (see the <http://dhcommons.org/projects> for this and other active DH projects). While this scale of material may seem unfathomable to qualitative researchers typically working at the project level, it is not out of the realm of possibility, as digital data collection makes massive compilations possible (these authors are aware of at least one organization that is in the process of collecting and compiling user activity research from countries across the world). The challenges and opportunities of digital collections are similar for both fields: digital collections offer a new ability to combine previously separate datasets into a single corpus for study; they open up new forms of inquiry that were either not possible or not useful before; and they challenge the current tools and approaches that have not been optimized for data sets of such great diversity, complexity or scale.

Of particular relevance to the qualitative researcher is a second effect of computing in the humanities: their development of new digital tools. Digital Humanists are developing data visualizations, text-mining algorithms and interfaces that support exploratory inquiry to take advantage of data now in computational form. Such efforts are often government funded, including support from the relatively new Office of Digital Humanities at the NEH. As a result, the Digital Humanities is rapidly advancing computer-enhanced text analysis techniques and visualizations of encoded data, and opening up new ways to perform research in the process.

Below we show three examples of Small Package-style tools, all working with the same data set for easier comparison. The data used is the output of Revelation, an online research platform, used to engage 25 participants in a study of household management behaviours over a two-week period. This data set contains 118 entries of their shopping diaries; each diary entry contains ten variables, such as store shopped, item shopped for, description of the experience, satisfaction level, shopping style (online, offline or a mix), in addition to time, date, participant and segment. This dataset, therefore, contains 1180 units for analysis—modest in scale but one of thirty activities users were asked to engage in over the course of the two-week study.

## Voyant

*Voyant* (<http://voyant-tools.org/>) is an online text analysis environment designed to explore and compare large texts. It employs a dashboard interface (see Figure 1) with multiple tools that can be kept open or closed, allowing the researcher to optimize their work environment. Sinclair, its creator, describes *Voyant* as “designed for humanists who wish to spend more time exploring their corpus than learning complicated, statistical and analytical software” (Sinclair & Rockwell, 2012 p. 259). *Voyant*’s default interface offers six fundamental tools, but can be customized from a larger library containing twenty text investigation tools.

On opening a file, *Voyant* creates a Cirrus cloud to represent high-frequency words in the text (researchers can apply and edit numerous Stop Word lists). From there, researchers are presented with multiple tools to help analyze or explore a single text, or to compare multiple texts. As an example of Small Packages, *Voyant* itself is a modular series of tools—some statistical, some visual, some exploratory. While *Voyant* is designed to be both an analytic and a reading environment, qualitative researchers may find it most useful as a means to get familiar with new data: Its core toolset quickly filters and clusters subsets of text, accelerating the identification of telling language and larger themes for deeper coding.

## Mandala

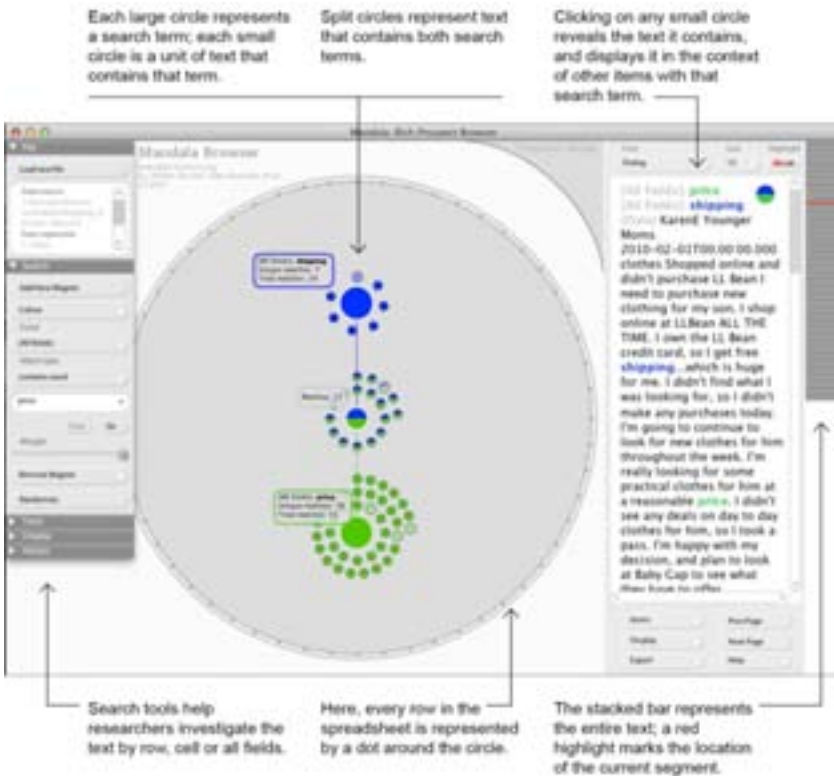
*Mandala* (<http://mandala.humviz.org/>) is a desktop tool that allows researchers to build rich visualizations of a text based on search terms, and to see how those texts interrelate when two or more search terms are applied. The interface uses a magnet metaphor to search and “attract” units of text based on search terms entered by the researcher. These are represented as a cluster of circles on the screen (see Figure 2) that then act as a direct interface to those units of text. *Mandala* is built on Ruecker’s principles of Rich Prospect Browsing (2006, 2011), which advocate that effective visual interfaces represent all elements of a collection at all times, that those representations become the means for accessing further data, and tools are provided to the researcher for manipulating those elements into meaningful representations.



**FIGURE 1.** Here we've arranged the Voyant interface to show three of its tools. Note the Cirrus cloud highlights that "price" is a high frequency word in the shopping diary (82 instances). Clicking on "price" brings up every instance in the panel on the right, allowing the researcher to quickly establish the contexts of use. Clicking on any line on the right pulls the full text into the central panel for closer reading.

Both Mandala and Voyant are tools explicitly developed to enhance what humanities scholars' term "distant reading." Distant reading, which focuses on non-linear reading across texts, stands in contrast to traditional, "close reading" (reading from beginning to end) of a small number of texts. Digital humanities scholars believe "distant reading... is a *condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems" (Moretti, 2000). The concept of distant reading is a substantive topic too large to explore in detail in this paper; however, it offers a potentially powerful point of overlap between digital humanities scholars and qualitative researchers: both professions utilize text-based methods of inquiry; both are encountering larger and larger amounts of text to assess; both suspect that larger samples and the patterns they hold may bring forth new meaning that smaller samples cannot; both would agree that focusing on a few individual texts/people only makes sense if the researcher is convinced that only a few of them matter. From a Small Packages perspective, it makes sense for both

fields to collaborate in the development of new tools that advance new analytic practices not possible before the digitization of data and the widespread use of computers for analysis.



**FIGURE 2.** *Using Mandala, we pursue the insight from Voyant that price is an important concept to understand. First, the researcher searches for “price” (all text that contains this term is represented by small circles at the bottom of the screen). Next, on a hunch, the researcher searches for “shipping” (small circles at the top). Notably, Mandala highlights that 17 entries use both price and shipping (split circles in the middle)—a co-occurrence that bears looking into. To read all text containing either term, the researcher can browse the source data using the right-hand screen.*

## Nineteen

Like Mandala, *Nineteen* (<http://data.pollari.org/>) also reflects the principles of Rich Prospect Browsing, but is a web-based tool designed to support researchers who work with spreadsheets to manage their data. *Nineteen* represents every row of data in a spreadsheet as an individual unit onscreen (see Figure 3). A small set of controls allow researchers to



## PAPERS / Session 1, Big Data/Ethnography or Big Data Ethnography

display, explore and read these units, which can be clustered or ordered based on any variable in the columns. Nineteen is useful in that it helps researchers see all their data at once, and allows them to arrange and rearrange that data to quickly discover correlations (i.e., activity patterns by time or date), outliers (such as over or underactive participants or segments) and other issues that are hard to spot in a sprawling spreadsheet. The intended use of Nineteen is as a first-stage exploration of digital data, using dynamic visualizations and direct manipulation to generate a quick mental model of the data and to accelerate and inform subsequent data analysis. Nineteen's features are explained more fully in the case study immediately following.



**FIGURE 3.** Here, *Nineteen* represents every diary entry as a square. Using the control panel on the left, the researcher has arranged diary entries by participant, from most to fewest entries, and colored them by the 4 participant segments. We quickly see that two participants in the same segment have generated a substantial number of the overall entries—a potential issue to investigate and factor into pattern detection efforts. Adding in the search term “price” highlights all entries/squares that mention price, and again allows the researcher to quickly browse that subset of entries.

## A CASE STUDY WITH NINETEEN

In this section, we explore an application of *Nineteen*, as used by a design planning team at the IIT Institute of Design on a research-driven design project for Chicago Public Media and its public radio station, WBEZ.

In 2011, WBEZ partnered with a class of graduate students, including author Ted Pollari, led by Professor Tomoko Ichikawa and in consultation with author Professor Kim Erwin. The class was asked to execute a rapid process of research, analysis, and concept generation, moving from initial recruitment of study participants to presentation of final concepts in approximately eight weeks. This was accomplished with the use of an online research platform, Revelation, to collect research data and Nineteen as a support tool for engaging that data.

The study produced a great deal of data in a short period of time. From an initial pool of 125 self-selected WBEZ listeners, the study included 25 participants self-reporting their activities over six days via Revelation. The recruited cohort included five participants in their 20s, ten in their 30s, five in their 40s and five in their 50s. In total, participants completed 441 sets of responses and activities, some of which were repeated multiple times, generating text and images for the design team to analyze. To show how Nineteen was of value in this scenario, we will discuss the use of Nineteen as both a monitoring and reading device as data came in, and as an early-phase analytic tool used with the data generated.

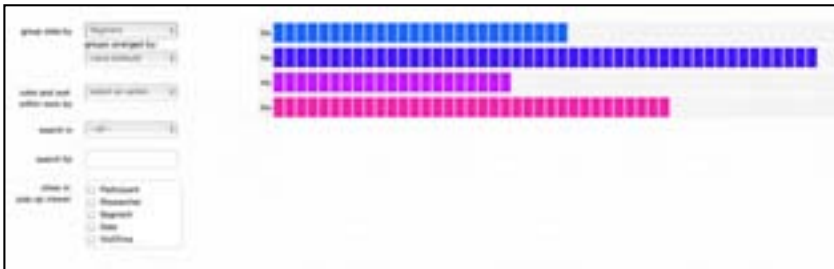
First, Nineteen proved useful to the team as a data-monitoring tool. Each day, team members would export data from Revelation (which has limited support for data analysis) to monitor responses as they came in. Nineteen proved useful as a data monitoring tool because the visual environment allowed researchers to shift between a high-level aggregate view of the data to a close inspection of individual responses without changing screens. This simultaneous paring of abstracted and detailed views allowed the team to identify specific participants who needed additional engagement to improve response rates or quality. It also allowed the team to build and evolve an understanding of the participants, both as individuals and as a collective. In addition to a browsing tool, a number of students chose to use Nineteen as a reading and viewing interface because of its ability to present the text and image responses of *all* participants to a specified task (see Figure 4).

In the analysis phase, one early task for Nineteen was simply to see whether participant responses to a given activity balanced across individuals and segments. A quick visual assessment of Figure 5 reveals a response rate proportional to the number of people in each segment.

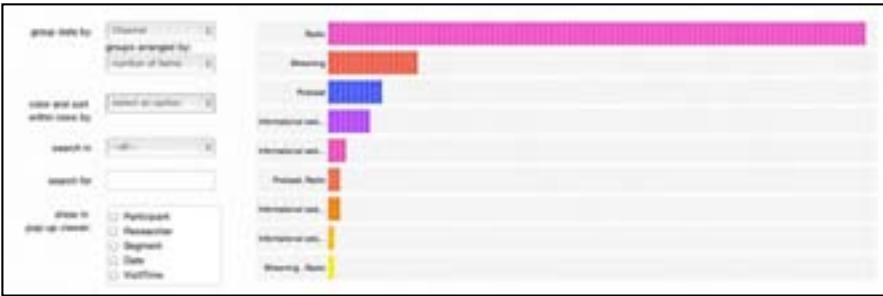
Early expectations among WBEZ staff and the design team were that its website might be a major point of interaction for participants. However, a quick look at usage log entries by channel (Figure 6) told a different story—it was clear that radio was by far the most common point of contact. This was true regardless of participant age (Figure 7).



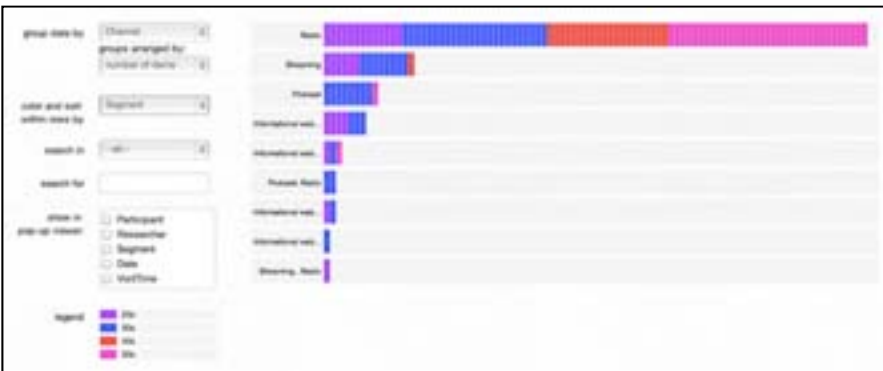
**FIGURE 4.** Keeping pace with the data: Each square represents one participant's response to an assigned activity. Here the researcher is browsing all responses to the question, "What does WBEZ mean to you" and the resulting text and images.



**FIGURE 5.** Checking on the data: here we see the total participant responses to the WBEZ usage log, organized by segment. Each entry in the log is represented by a rectangle. A quick look tells us that, given the number of participants in each segment, the number of entries per segment is roughly even—a good sign.



**FIGURE 6.** Despite early theories that recruiting participants through social media might make the website a dominant touchpoint among the group, results of the Usage Log show that most entries were about radio experiences, not web experiences.



**FIGURE 7.** When the visualization is recolored to reveal participant age, we see that radio is the central experience point of the brand, regardless of age.

Usage Log responses viewed through Nineteen also allowed the team to explore how WBEZ users interact across media. For example, the team sought to understand Facebook's role in driving contact with WBEZ. Nineteen's full-text search feature revealed many entries (Figure 8) that include the word "Facebook." Using the pop-up feature to read just those log entries revealed two things. First, entries that matched "Facebook" and used the website as a channel came from only two participants; however, those entries all mentioned being driven to the WBEZ website by Facebook and not the other way around. Second, in most of the other cases, participants noted Facebook usage occurring *while* listening to content from WBEZ. This suggested a possible theme that eventually emerged from the study as a whole: WBEZ is often a constant companion during listeners' days, sometimes as the focus of attention, and other times as a background companion as participants multitask.



**FIGURE 8.** Still viewing Usage Log entries by media type, the search function highlights those containing the word Facebook, now readable as a collection.

Is time a revealing lens? Each user's entry in the Usage Log data contained timestamps, allowing Nineteen to display how those entries fall out over the course of a day. In figure 9, we see that entries are logged relatively evenly until around 7pm; this was reflected across all age segments as well. This visualization also helped correct one researcher's early impression that participants were using podcasts more during the evenings, but a quick re-colouring by channel revealed this to be unsupported by the data.

The use of Small Packages like Nineteen to poke at data does not replace the need for in-depth analytical efforts. But as demonstrated by this case study, and as suggested by Given and Olson (2003) and Erwin (2011), more work up front can speed later analysis by allowing researchers to more quickly form and test models of their data. In this case, the use of Nineteen directed the WBEZ team's attention to two provisional insights that emerged more strongly during close reading of the entries and developed into design principles for prototype development: first, that *WBEZ is radio* and their users value it for the qualities it entails; second, *WBEZ is a constant companion* for some listeners: at home, at work, in the car, WBEZ is the soundtrack to their day. The team used those two insights as part of a network of ideas that were instrumental in developing the WBEZ Broadcast Browser concept<sup>1</sup> presented to the WBEZ board in spring of 2012.

### Methodological implications of Small Packages

What might be the impact of engaging Small Packages in an analytic process? The most substantive implication for analysis is also the most intentional: Employing focused, loosely coupled, task-driven tools can allow researchers to be more responsive and diverse when

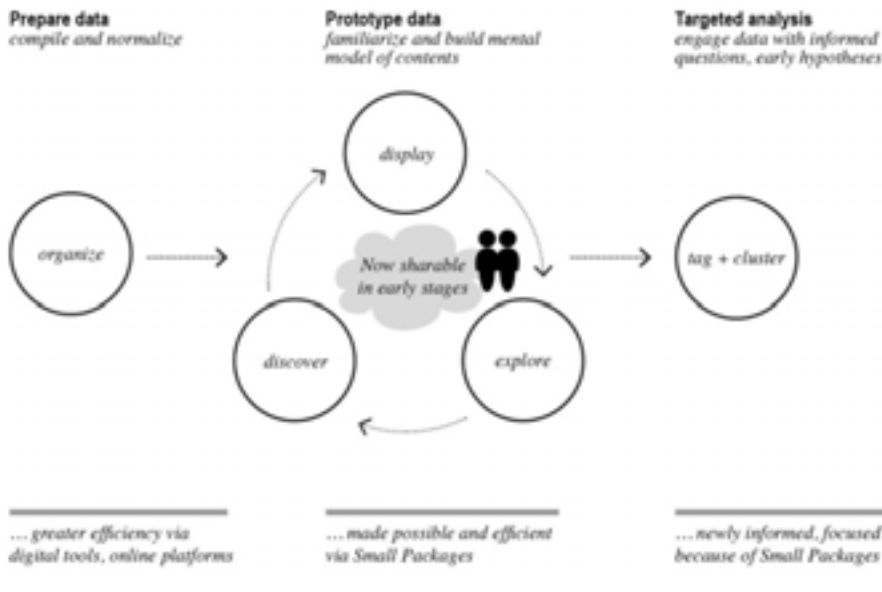
<sup>1</sup> A video demo is available at <https://vimeo.com/43589121>

crafting their analytic processes. If the impact of similar tools in the Digital Humanities is any indication, a Small Packages approach in qualitative research is likely to generate new analytic approaches and new forms of knowledge that complement and inform traditional code-focused approaches. Our own early experiences using these tools in team settings produced analytic processes that became more iterative, non-linear and exploratory, and created insights that then focused and accelerated the inductive analytic processes that followed. As a result, we suggest that Small Packages be used upfront, before inductive coding, to help the researcher create a structural understanding of the data from multiple viewpoints (see Figure 10) and proceed from there.



**FIGURE 9.** All entries in the Usage Log are arrayed by time of entry. Note the small early morning peak and relatively constant logging during morning, afternoon hours.

Small Packages has other procedural implications: Because they are by design decoupled from the larger analytic process, Small Packages can allow researchers to engage in early explorations of data almost as quickly as it arrives. This enables parallel work processes of data collection and analysis, and more closely mirrors the traditional notions of grounded theory in supporting an interplay between data collection and analysis processes (Bowen 2008, p. 13). Enabling researchers to be immersed in the data as it unfolds also means they are able to recognize potential problems with their data collection tasks or with specific participants, and opens up the possibility of addressing them quickly enough to minimize the impact on the research in progress, instead of discovering issues well after data collection had ended.



**FIGURE 10.** *The small packages concept proposes a new step in the analytic process, one that is fast, visual, iterative, and exploratory. This new step allows researchers to engage more deeply with their data earlier in the process to drive efficiencies in later coding efforts.*

A more speculative implication involves the inclusion of more people in the analytic process: The visual and dynamic nature of Small Package transforms data into a compact object that is then easier to show and share with others in the early stages of analysis, before codes have solidified or insights have been culled. Qualitative analysis is often a shared effort among immediate team members. However, engaging clients, stakeholders or subject matter experts—people who benefit from early inclusion—can prove difficult in the early stages, when data is unstructured and hard to conceptualize. Small Packages, then, can aid in the communication of research.

## CONCLUSION AND NEXT STEPS

We are witnessing a tremendous revolution in qualitative research. Never before have researchers in business and academic environments had the ability to collect so much data in such tight timeframes. While this is exciting, it poses clear challenges: traditional analytic tools have not enabled researchers to keep up with the increased volume of pre-organized, bite-sized chunks of data produced by online research platforms. Traditional computer

assisted qualitative data analysis software tools were built for different challenges and, as such, new tools must be developed to help researchers cope with this new paradigm.

The Small Packages for Big (Qualitative) Data concept is not a substitute for an active and engaged researcher. Instead, it aims to more quickly engage researchers with their data and their analytic processes by accelerating a structural understanding of what could otherwise be an insurmountable avalanche of data. To do this, the small packages concept builds on modular, loosely-coupled tools, emphasizing visualization and employing dynamic interfaces to enable parallelizable and responsive workflows. This enables researchers to quickly explore their data, raise questions about its nature, and aid the researcher in building a mental model of the dataset and its contents to carry into analysis.

As we've shown, a number of tools already exist that fit with the Small Packages paradigm, many coming from the Digital Humanities. We believe that qualitative researchers would be well served by fostering a cross-discipline development and sharing of analytic tools. The Small Packages paradigm sets the stage for this by advocating for tools that are focused in scope, and thus more amenable to being adopted outside of their original discipline.

Going forward, we recognize that one substantial challenge for tools that fit the Small Packages paradigm is to work on *robustness*—all three tools discussed above are still in development and are still considered beta or prototype software. Further, it is clear that more work must be done to make Small Packages a more complete suite of analytic tools. We hope that by identifying the need for this shift in paradigm, we will encourage a similar shift in the aims of researchers and programmers tasked with developing new tools. By explicitly embracing the principles of the Small Packages concept, we believe that more effective tools can be built, tested and put into use more quickly.

## REFERENCES CITED

- Bertin, Jacques  
1967 *Sémiologie Graphique: Les diagrammes, les réseaux, les cartes*. Paris: Gauthier-Villars. Translated by William J. Berg as *Semiology of Graphics: Diagrams Networks Maps* (The University of Wisconsin Press, 1983).
- Bowen, Glen  
2008 "Grounded theory and sensitizing concepts." *International Journal of Qualitative Methods* 5.3: 12-23.
- Cefkin, Melissa  
2013 "The Limits to Speed in Ethnography," *Advancing Ethnography in Corporate Environments*, 108-121. California: Left Coast Press.
- Cefkin, Melissa  
2010 *Ethnography and Corporate Encounter: Reflections on Research in and of Corporations*. Pp1-9. Berghahn Books.



**PAPERS / Session 1, Big Data/Ethnography or Big Data Ethnography**

- Cleveland, William S., and Robert McGill  
1984 "Graphical perception: Theory, experimentation, and application to the development of graphical methods." *Journal of the American Statistical Association* 79.387: 531-554.
- Ewin, Kim  
2011 "The Visual Coding Of (Big) Qualitative Data: New Analytic Methods And Tools For Emerging Online Research." In *Diversity and unity: Proceedings of LASDR2011 4th World Conference on Design Research*. Roozenberg, , L. Chen, and P. Stapperseds.. Delft, The Netherlands, October 31 - November 4, 2011.
- Few, Stephen  
2009 *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press.
- Given, Lisa and Hope Olson  
2003 "Knowledge organization in research: a conceptual model for organizing data." *Library and Information Science Research* 25: 157-176.
- Healey, Christopher G., Kellogg Booth, and James T. Enns  
1996 "High-speed visual estimation using preattentive processing." *ACM Transactions on Computer-Human Interaction (TOCHI)* 3.2: 107-135.
- Isaacs, Ellen  
2013 "The Value of Rapid Ethnography." *Advancing Ethnography in Corporate Environments*. Pp 92-107. California: Left Coast Press.
- Ladner, Sam  
2012 "Is Rapid Ethnography Possible? A cultural analysis of academic critiques of private sector ethnography"(guest blog entry). *Ethnography Matters*, January 13, 2012. <http://ethnographymatters.net/2012/01/26/is-rapid-ethnography-possible-a-cultural-analysis-of-academic-critiques-of-private-sector-ethnography-part-2-of-2/>.
- Malefyt, Timothy de Waal  
2009 "Understanding the rise of consumer ethnography: branding techno methodologies in the new economy." *American Anthropologist* 111 (2): 201-210.
- Miles, Matthew and Michael Huberman  
1994 *Qualitative Data Analysis: an expanded sourcebook*. Sage Publications.
- Moretti, Franco  
2000 "Conjectures on World Literature." *New Left Review* Jan-Feb: 56-57.

- Nickerson, Jeffrey, et al.  
2013 "Cognitive tools shape thought: diagrams in design." *Cognitive processing*: 1-18.
- Orton, J. Douglas, and Karl E. Weick  
1990 "Loosely coupled systems: A reconceptualization." *Academy of management review*, 15.2: 203-223.
- Rhea, Darrel and Lisa Leckie  
2006 "Digital Ethnography Sparking Brilliant Innovation." *Innovation: the IDSA Quarterly Design Journal* Summer 2006: 19-21.
- Ruecker, Stan  
2006 "Experimental interfaces involving visual grouping during browsing." *Partnership: the Canadian Journal of Library and Information Practice and Research* 1.1.
- Ruecker, Stan, Milena Radzikowska and Stéfán Sinclair  
2011 *Visual Interface Design for Digital Cultural Heritage*. Ashgate Publishing, Surrey UK.
- Sinclair, Stéfán and Geoffrey Rockwell  
2012 "Teaching computer-assisted text analysis." *Digital Humanities Pedagogy: Practices, Principles and Politics*. Brett D. Hirsch, ed. Open Book Publishers.
- Slone, Debra  
2009 "Visualizing qualitative information." *The Qualitative Report* 14.3: 489-497.
- Thrift, Nigel  
2000 "Performing cultures in the new economy." *Annals of the Association of American Geographers* 90 (4): 674-692.
- Tufte, Edward  
2001 *The visual display of quantitative information*. 2<sup>nd</sup> Edition. Cheshire, CT: Graphics Press.
- Tversky, Barbara.  
2011 "Visualizing thought." *Topics in Cognitive Science* 3.3: 499-535.
- Ware, Colin  
2008 *Visual Thinking for Design*. Morgan Kaufmann.