Title
Data Quality for Situational Awareness during Mass-Casualty Events

Authors
Demchak, Barry
Griswold, William G
Lenert, Leslie A

Publication Date
2006-04-11

Peer reviewed

Draft, submitted to AMIA 2006

Data Quality for Situational Awareness during Mass-Casualty Events

Barry Demchak, BA[2] William G. Griswold, PhD[1,2] Leslie A. Lenert, MD[1,3]
[1]California Institute for Telecommunications and Information Technology and
[2]Computer Science and Engineering
University of California, San Diego, La Jolla, CA
[3]Veterans Affairs San Diego Healthcare System, San Diego, CA

Incident Command systems often achieve situational awareness through manual paper-tracking systems. Such systems often produce high latencies and incomplete data, resulting in inefficient and ineffective resource deployment. The WIISARD system collects much more data than a paper-based system, dramatically reducing latency while increasing the kinds and quality of information available to incident commanders. Yet, the introduction of IT into a disaster setting is not without problems. Most notable is that temporary system component failures delay the delivery of data. The type and extent of the failure can have varying effects on the usefulness of information displays. We describe a small, coherent set of customizble information overlays to address this problem and discuss reactions to these displays by a medical commander.

INTRODUCTION

Traditionally, medical command makes decisions using a paper-and-radio system. Triage, treatment, and transport officers fill out tallies and create written reports while also performing their primary duties. These reports are periodically hand-carried or called into the medical command, where they are manually reviewed, summarized, and posted. This labor intensive process creates variable latencies, and incomplete and low-resolution data.

The WIISARD system combines state-of-the-art data collection and display devices, database services, and 802.11 wireless communications to produce a consistent, real-time view of the disaster scene (1). It further improves on the paper system by automatically integrating data into map- and graph-based displays, giving the command insight for making better decisions with greater confidence. It tracks patient and provider status, medication inventories, ambulance and hospital bed status, hot zone and plume locations, law enforcement zones, and device locations. This surfeit of continuously streamed data creates new technical challenges.

WIISARD's command center overcomes the limitations of scarce display space by role-tailored maps and graphs in a tiled layout. Medical command can choose and arrange displays as the dynamics of the situation may require (See Figure 1). Map displays are zoomable and scrollable, and medical command can select any of several overlays to both increase data density and reveal important correlations. Graph displays show both summarized and correlated data in easy-to-read standard formats. Graph layout and coloring is optimized for use in harsh sunlight.

This design addresses the information overload naturally created by the WIISARD system, but it does not cope with the issue of data quality. Some portion of the mobile devices and the network deployed at a disaster scene will fail intermittently. In such situations, the information arriving at, say, a command center display will be stale to some degree; how much is difficult to tell. Yet, medical command and field supervisors should be able to continue making decisions using these information tools.

We take an information visualization approach to this problem, providing information overlays on situational displays that convey the nature and extent of data staleness. An additional graph conveys information about the underlying system failure. A commander can then make informed decisions with confidence. Depending on the facts conveyed in the overlays, this could be "business as usual" with some checking by radio, engaging the IT group, or reverting to paper-and-radio until integrity is restored.

The contributions of this paper are a set of requirements for data quality management in mass-casualty response, an information visualization solution to the problem, and an initial paper evaluation of this solution. The requirements and evaluation are the result of deploying a prototype WIISARD command center at the San Diego MMST Drill at the Del Mar Fairgrounds in Nov. 2005, as well as an in-depth interview with a medical commander.

BACKGROUND AND REQUIREMENTS

The initial response to a disaster at a site secures the scene. Medical teams then move in, establishing areas for triaging patients, decontamination (if necessary), subsequent treatment, and transport to area hospitals. The basic workflow is a pipeline, with patients moving through the stages in succession.

The standard information tools of response include: Simple Triage Rapid Treatment (START) tags that record each patient's condition and any critical treat-

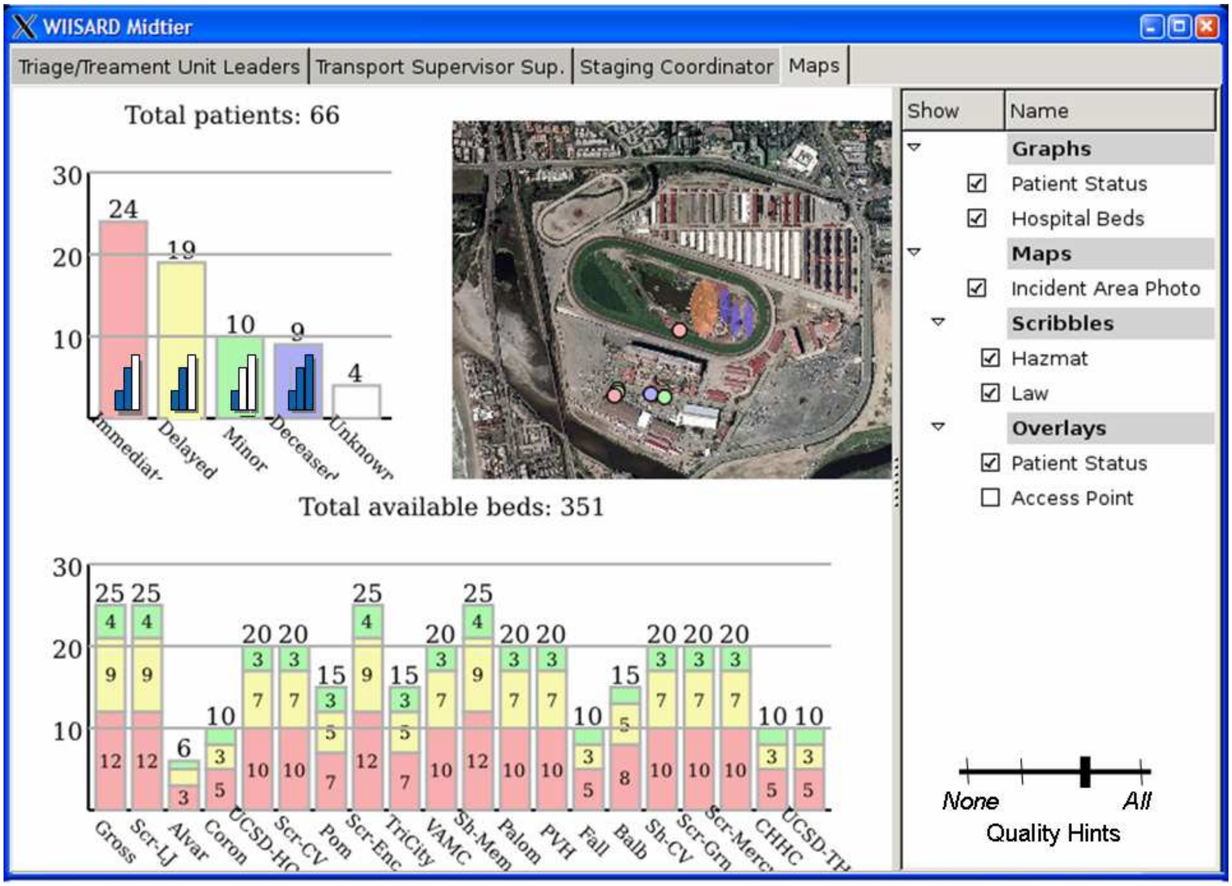

Figure 1. The WIISARD command center display.

ments (such as Mark I kits for nerve agents); clipboards for supervisors to track information such as transport logistics; and whiteboards and easels for commanders to track the status of the response. Information is moved among these tools by word of mouth, often by radio. This system of information management is mature, but slow and information-poor. Patients in need of immediate help may be just out of sight, dangers are not apparent, and key trends may be invisible to the command.

Mobile IT in the disaster setting can automatically propagate information among providers and construct information displays appropriate to each provider's role (e.g., triage provider, triage supervisor, medical command, etc.). With the WIISARD system, front-line providers carry wireless personal digital assistants (PDAs) with integrated barcode scanners, supervisors carry Tablet PCs, and medical command employs large-display devices. Patients are tagged with a wireless smart tag (2) or a traditional paper START tag. Network communication is supported by a portable 802.11b mesh network (3).

A triage provider, for example, places a tag on the patient, scans the tag's barcode, and then, assessing the patient, clicks several items on a screen that looks like a START tag. (If just setting the patient's status, then triage can be performed using buttons on the smart tag.) The patient's information is automatically distributed, for example to the triage supervisor's patient list and the medical command's graphs and map display. No pause for human conversation is necessary, and the communication is essentially instantaneous. START triage can be completed in about 30 seconds, roughly half the time of the traditional way. The smart patient tag continuously updates the patient's location, and if connected to a pulse/oximiter, continuously monitors patient status.

Because of the harsh dynamics of the disaster scene, devices will fail, networking will be interrupted, and providers or patients will walk beyond the network's range. Typically these problems will be resolved in a short amount of time. The challenge for the medical command (and field supervisors) is to understand what data on their displays is out of date, to what extent, and what it means to them. Perhaps a few provider PDAs have reported for a few minutes; perhaps a small number of PDAs and patient tags in an isolated area have not reported in 20 minutes.

Information displays that just show the last data reported will not convey these nuances. Without some cue as to the quality of the reported data, the commander is left to blindly (mis)trust the data.

Medical command is experienced in assessing data quality with paper-and-radio based systems, as well as taking data quality into account when making decisions. Yet, addressing the lack of quality information faces stiff requirements. Commanders are already overwhelmed with data, so providing more information can be a negative. The reported quality data can itself be suspect, since it is captured and delivered over the same infrastructure. Deploying a redundant infrastructure to manage data quality is out of the question. What redundancy that can be afforded is invested directly into the core IT infrastructure to minimize failures in the first place.

PRIOR WORK

The research in this area can be roughly divided into ontology, metrics, and uncertainty visualization.

Wang et al. constructed an early ontology of data quality, as well as primitive metrics for the various definitions (8,9). Bouzeghoub et al. refined Wang's timeliness metrics and exponential decay of quality over time by introducing notions of data volatility and delay of data delivery (4). These works point to the idea that data quality is a function of time and the data's likely rate of change.

Mackinlay et al. explored uncertainty visualizations involving variants of error bars (6). Pang et al (7) developed aggressive uncertainty visualization overlay techniques, exploiting 3D and color. MacEachren (5) explored uncertainty visualization overlay techniques for spatial data, many employing blurring, graying, and shading. Our approach depends heavily on overlays.

THEORY AND DESIGN

We take an information visualization approach to managing data quality. Given data quality visualizations, experienced commanders will be able to make the same quality assessments using WIISARD as they do today with their existing systems.

When looking at a typical command center display with an eye for data quality, three questions come to mind: which (or how many) data points are out of date, how much time has passed since the last report for these data points, and how quickly are these particular data points likely to go out of date?

Medical command is usually interested in the bottom line—*is this data good enough to be used to make decisions?*—meaning that the time passed since the last report is not of direct interest. However, the likely usefulness of the data can be assumed to be a monotonically decreasing function over time (8,9). The quality of data that is considered volatile will diminish more quickly than the quality of slow-changing data (4).

Of course, volatility varies across different *types* of data—say patient status versus location on the scene. But it also varies across *different values* of the same type. For example, the condition of a severely injured patient is likely to change more quickly for the worse than the condition of one with minor injuries.

For uniformity, we normalize quality as a fraction between 0 (useless) and 1 (entirely useful). We use the notation $Q(d, t_d)$ to denote the quality function for a datum d that was reported at time t_d. The notation $Q(d)$ denotes the quality for datum d based on its time of capture, t_d. An open question is how the function $Q(d, t_d)$ is determined for the data of disaster response. Lacking a validated theory, we are consulting emergency medicine professionals to help us construct these functions on a case-by-case basis.

With these thoughts in mind, we can characterize the quality for a given display as follows. Given:

D_G, the data points displayed in a graphic G, and

$D_{G,T}$, the data points of type T displayed in a g G,

we define the overall quality of a graphic, $Q(G)$, as the average quality of the data points in G, and $Q_T(G)$ as the average quality of the data points of type T in G. Replacing T with v defines $Q_v(G)$, the average quality of the data points of value v in G.

Information Visualization Design Principles

We cite four principles that we applied in our approach. These are not entirely unique to data quality, but are still the main drivers of our design:

Overlay quality metadata. The human brain is excellent at comparing visually juxtaposed information and drawing conclusions. Overlaying a visualization with quality attributes instantly identifies the quality attribute with its data. It is also efficient in its use of screen space, a scarce resource.

Avoid shading or color gradients. The visual contrast of computer displays is low in outdoor settings, so subtle distinctions in color saturation or tone are difficult to see. This takes an important set of visualization techniques off the table, especially as relates to overlays. For example, it is not practical to convey poor data quality for an item by dimming it.

Support customization. Different situations result in different information needs. Because information overload is a concern, the ability to selectively enable overlays and ancillary displays is critical.

Support exploration. The questions answered by one display can lead to other questions. Ideally, those questions should be explorable with a few clicks rather than a call to the IT group or field personnel.

Approach

Our information visualization approach employs several types of information displays, depending on (1) whether the quality attribute applies to an individual data item (e.g., a single patient) or an aggregate (e.g., all "immediate" patients), (2) what visual dimensions are already in use in a graphic (e.g., a row of 1-D colored bars in a bar graph versus a 2-D black-and-white map), (3) how much detail on data quality is desired, and finally, (4) whether an overlay is feasible or not. We will not explore the complete design space here, but will instead use examples that convey how the constraints therein drove the design.

The overall concept of operation is illustrated in Figure 1. A slider in the right-hand area controls the amount of quality information displayed in the selected graphs, displayed on the left. Running the slider to the left amounts to flying blind, but it can be used temporarily to get an uncluttered look at detailed graphs. Moving the slider to the next tick mark (the default setting), adds basic data quality overlays of the form *Q(G)*. The next tick breaks down the quality information with overlays of the form $Q_T(G)$ *and* $Q_v(G)$. The last tick adds overlays that display, within reason, *Q(d)* for every datum *d* in the display.

For displaying *Q(G)*, $Q_T(G)$ *and* $Q_v(G)$, we overlay a display reminiscent of the signal-strength bars shown on mobile phones (Figure 1, patient status bar graph). We chose this visual cue because it is a familiar quality indicator, thus both easy to learn and reinforcing that the indicator is about quality. Also, it does not depend on subtle color or shading, and can be drawn fairly small.

For displaying all *Q(d)* in a graphic, we create a 2-D graphic with the data elements enumerated on one dimension, and the quality in the other (Figure 2). To make the display easy to read, the elements are sorted according to data quality. As shown in Figure 2, the patients are enumerated vertically, from highest quality at the bottom to lowest quality at the top. The quality is drawn as a line, left to right, whose proportional length is the quality. In the case of Figure 2, the quality lines are colored in the base color of the graphic, and the unfilled portion is drawn in a very light shade of the base color to maintain high contrast. A black line is drawn between the two regions to increase their visual distinctness.

Many interesting conclusions are possible from such a quality graph. Looking again at Figure 2, and focusing on the "immediate" patients on the left, we can surmise that about one third of the immediate

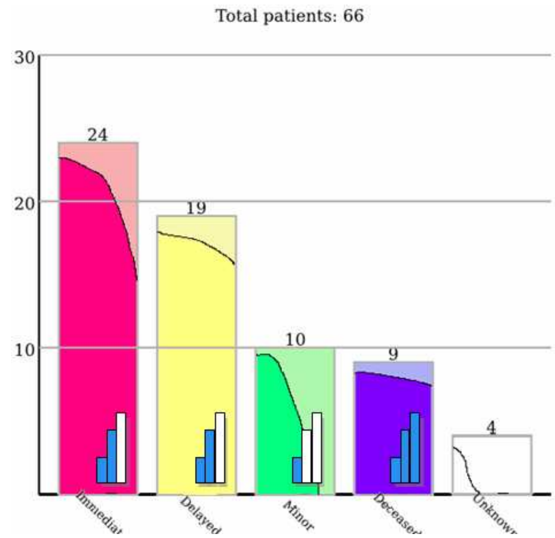

Figure 2. Displaying *Q(d)* for all patients.

patient data has less than perfect quality, as the "quality frontier" meets the right side of the bar about one third of the way down. We can also see that one patient has for all intents and purposes completely disappeared, a serious concern. Perhaps four patients total are of some concern at this point. Looking at the "minor" patients in the middle, we might be concerned that 80% have not reported in a while. Perhaps these patients have taken it upon themselves to leave the scene, which can be a serious problem.

In the example shown in Figure 2, we have in essence embedded a 2-D graph inside a 1-D display. The width of the bars in a bar graph normally carries no intrinsic meaning; it just enhances readability. There is no such obvious embedding trick for 2-D graphics like maps or scatter plots. Instead, we can create a miniature standalone version of our desired graphic and overlay it in an unused part of the graphic, much like we do with our "signal" bar graph. If it is desired to break down the data by type or value, as in Figure 2, then multiples of this graph can be overlaid, space permitting.

These *Q(d)* overlays convey a lot of information in a small amount of space, and potentially can clutter a graphic. We anticipate that these overlays will be used intermittently, when the medical command needs to drill down to answer a particular question when data quality is not at its best. Most of the time the "signal" bar graphs will prove sufficient.

Assistance for diagnosing poor data quality

Knowing the underlying cause for degrading data quality is useful in decision making, especially for fixing the problem. The details of diagnosis are beyond the scope of this paper, but a few information visualizations for basic diagnosis are discussed.

To support basic diagnosis, the system's components need to be modeled by the system just like patients, medicines, etc. This makes them available for graphical display in the command center as summative bar graphs and as objects on the map display.

A display that captures useful information is a "stacked" bar graph. A complex example of such a graph appears at the bottom of Figure 1, which is an inventory of multiple bed types at area hospitals. For inventorying system components, we wish to have a bar for each component type—PDAs, patient tags, Tablet PCs, and network nodes. The bottom part of each component's bar is the number of reporting components, and the top is the number of components that have long since stopped reporting since the inception of the response. Quality data can then be overlaid on these bars to provide hints of what reporting components may soon be non-reporting, etc.

All of these devices can be selectively shown on the map as well. A question, however, is how quality data can be shown. Markers for patients already use the triage colors to convey medical status. Adding a separate marker for their devices could significantly clutter the display, but may be possible. On the other hand, using color overlays to indicate the health of network nodes is neither ambiguous nor cluttering.

Finally, there is the question of the command center's own network connection. If lost, all the displayed data's quality will decay together. Rather than depending on this indirect, albeit alarming, cue, it makes sense to provide an additional connection indicator. However, this is not a data quality indicator, but rather a health indicator, so using one of our data quality visualizations could be confusing.[1] An adequate cue would be to display a green dot for connected and a red dot for disconnected, with the accumulated (dis)connection time overlaid.

EVALUATION

To gain an early assessment of our approach, we constructed a set of graphics like those in this paper, and shared them with a fire captain responsible for medical command in small-to-medium events. After a brief training, the captain was able to interpret the overlays and formulate remediation plans. Looking at the minor patients in Figure 2, the captain immediately inferred that patients were leaving the scene and that there were problems with controlling the scene.

The captain said that a command center would have to list the disconnected patients and highlight them on a map. This would facilitate manually checking up on the patients and diagnosing the connection problem. Quick navigation from the graphics to matching patient lists was desired.

CONCLUSION

The gathering of field data in disasters with mobile devices and networks necessitates understanding how their occasional misoperation is affecting data quality in visual displays. An information visualization approach, using a variety of overlays on existing graphics, can help commanders and supervisors make informed decisions. A San Diego MMST drill, planned for August 2006, will permit an empirical evaluation.

Acknowledgements: This work was supported by contract N01-LM-3-3511 from the National Library of Medicine. We thank Captain Daniel Saner of San Diego company 26 for his time in using our visualizations.

REFERENCES

1. Chan TC, Killeen J, Griswold W, Lenert L. Information technology and emergency medical care during disasters. *Acad Emerg Med.* 2004; 11(11):1229-36.

2. Palmer DA, Rao R, Lenert LA. "An 802.11 Wireless Blood Pulse-Oximetry System for Medical Response to Disasters", *Amer. Med. Inf. Assoc. 2005 Annual Symp.*; 2005.

3. Arisoylu M, Mishra R, Rao R, Lenert L. "802.11 Wirless Infrastructure To Enhance Medical Response to Disasters", *Amer. Med. Inf. Assoc. 2005 Annual Symp.*; 2005.

4. Bouzeghoub M, Peralta V. A Framework for Analysis of Data Freshness. *IQIS 2004*. 2004.

5. MacEachren AM. Visualizing Uncertain Information. *Cartographic Perspective*, 1992 Fall, p. 10-19.

6. Olston C, Mackinlay J. Visualizing Data with Bounded Uncertainty. Technical Report CS-2002-20, Computer Science, Stanford University. 2002.

7. Pang AT, Wittenbrink CM, Lodha SK. Approaches to Uncertainty Visualization. Technical Report UCSC-CRL-96-21, UC Santa Cruz. Sept 1996.

8. Pipino LL, Lee YW, Wang R. Data Quality Assessment. *Communications of the ACM*, Apr 2002, p. 211-218.

9. Wand Y, Wang R. Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*, Nov 1996, p. 86-95.

[1] Think of it this way: If another device were gathering data on the health of the command center, and such reports were delayed, then a data quality indicator would be overlaid on the out-of-date health status.