# Information Visualization As A Knowledge Integration Tool

**Diana Burley, The George Washington University, Ashburn, VA USA**

ABSTRACT:

This article advances the argument that information visualization is a valuable tool for knowledge integration activities; including information exploration, integration, and analysis; within complex organizational data environments. In order to effective meet the knowledge integration demands of organizations and their constituents, information visualization tools must be adopted within the context of practical considerations such as the type of data to be examined, the type of visualization desired, the usability of the tool and the capability of the users. Properly considered, information visualization tools will prove to be a valuable knowledge integration tool.

Keywords: *Knowledge integration, Information visualization, Data exploration*

## 1.    Introduction

This article advances the argument that information visualization is a valuable tool for knowledge integration activities; including information exploration, integration, and analysis; within complex organizational data environments. The discussion places the use of information visualization tools within the context of 21st century information and knowledge management; and includes an examination of the strengths and weaknesses of visualization tools, the opportunities their use exploits, and notable challenges to their practical use and implementation for knowledge integration.

## 2.    Information Visualization

Information visualization is rooted in the practice of data exploration. Data exploration is a process through which large amounts of data, located in disparate databases, are examined for structure, patterns and other characteristics to understand relationships and trends. The practice enables users to perform interactive investigations that lead to new insights about relationships in complex data. Data visualization is "the science of visual representation of 'data', which has been abstracted in some schematic form, including attributes or variables for the units of information" (Friendly, 2008, p. 2).

As a category of data exploration, information visualization is focused on "the use of computer-supported, interactive visual representations of data to amplify cognition" (Card et al., 1998). As a field of study, visualization "…uses interactive graphical tools to explore and present digitally represented data that might be simulated, measured, or archived" (Stone, 2009, p. 44). Information visualization enables the creation and exploration of large collections of data and allows for interactive exploration (Stone, 2009). The primary goal for information visualization is "to bring to light meanings in data that might remain hidden from view if displayed in other ways" (Few, 2008, p. 15). Visualization is particularly useful for exploratory analyses that provide the basis of knowledge integration activities. It provides researchers with a mechanism to explore patterns, test hypotheses, discover exceptions and explain their findings. Information visualization strengthens the ability of users to uncover phenomena that have been previously hidden.

Information visualization combines several different research areas including scientific visualization, human-computer interaction, data mining, information design, cognitive psychology, visual perception and computer graphics (Kerren et al., 2008). Although information visualization developed its roots from the field of scientific visualization there are key distinguishing features that set the two apart. Scientific visualization deals with "three-dimensional physical objects and processes such as blood flowing through heart valves, tornado formation, crystal growth, protein structures, and oil reservoir

shapes" (Bederson & Schneiderman, 2003, p. ix). Scientific visualization also tends to look at realistic renderings of volumes, surfaces, and illumination sources (Friendly, 2008).

Information visualization generally references large scale collections of non-numerical information and focuses on abstract phenomena such as social relationships, political polls and economic trends. The users of information visualization look primarily at the relationships among categorical variables and help to identify patterns or gaps in the data, while the users of scientific visualization focus on continuous variables such as density, temperature or pressure (Bederson & Schneiderman, 2003). With information visualization, graphical models may represent abstract concepts and relationships that do not necessarily have a counterpart in the physical world (Ferreira de Oliveira & Levkowski, 2003); allowing people to think more effectively about information in order to fully understand it (Few, 2008). This is critical for knowledge integration.

## 2.1.  Information Visualization For Knowledge Integration

Information visualization can be seen as the applied science that examines large amounts of data through visual representation (Friendly, 2008). Through visualization techniques, users are able to analyze and gain a better understanding of the data (Russell et al., 1993).

Information visualization amplifies human cognition in six basic ways. It increases cognitive resources by using a visual resource to expand human working memory; reduces search by representing a large amount of data in a small space; enhances the recognition of patterns; supports the easy perceptual inference of relationships that are otherwise more difficult to induce; allows for the perceptual monitoring of a large number of potential events; and provides a malleable medium that, unlike static diagrams, enables the exploration of a space of parameter values (Thomas & Cook, 2005; Wong & Thomas, 2004).

Information visualization is most effective in heightening the understanding when data contain an underlying structure that supports the inference that proximate items can be inferred as being similar, users are unfamiliar with the contents of the data or have limited understanding of the data structure, users have difficulty verbalizing necessary underlying information, and information is easier to recognize than describe.

However captivating, questions regarding the usefulness of information visualization for more analytical tasks (e.g. knowledge integration), continue to dog the field. Indeed, moving beyond visualization toward analytical interpretation is where significant knowledge integration support may lie. The emerging field of visual analytics addresses this task.

## 2.2.  Visual Analytics

Visual analytics is defined as "the science of analytical reasoning supported by the interactive visual interface" (Klein et al., 2006, p.72). It combines automated analysis with visualization in order to have a more effective understanding and reasoning with the existing data sets (Keim et al., 2008). As such, many decision makers are turning toward this field in order to address complex problems. Visual analytics enables users to synthesize information and gain insight from vague and sometimes conflicting data. It provides an avenue to deliver timely, defensible and comprehensible assessments and helps combine "new computational and theory-based tools with innovative interactive techniques and visual representation to enable human-information discourse" (Keim et.al., 2008, p.155).

With its focus on human interaction within massive, dynamically changing information spaces, visual analytics research concentrates on support for perceptual and cognitive operations that enable users to detect the expected and discover the unexpected in complex information space. Analytical reasoning is central to the analyst's task of applying human judgments to reach conclusions from a combination of evidence and assumptions (Thomas & Cook, 2005). Visual analytics seeks to marry techniques from information visualization with techniques from computational transformation and analysis of data. Information visualization itself forms part of the direct interface between user and machine. This process is akin to the knowledge integration process in which multiple ideas are synthesized into a

single representation that is larger than the sum of its parts. It is likely then, that the capabilities of information visualization, combined with computational data analysis, can be applied to analytic reasoning to support the knowledge integration process.

## 3.      Visualization Tools

Visualization tools are used for a variety of tasks common tasks including: list generation and graphical display (e.g. histograms), concept grouping and mapping, adding temporal components to maps, citation analysis, clustering, categorizations, grouping and extraction of text, federated searching. Whereas some tools use statistics analysis techniques as their basis, others use semantic analysis algorithms to analyze data (Yang et al., 2008). Types of textual data can be divided into three categories: structured content (e.g. database records), unstructured or abstract content (e.g. full-text documents, email, webpage contents), and hybrid content that can include both structured and unstructured data (Yang et al., 2008). Accordingly, tools can be distinguished by their use of different types of data inputs (structured, unstructured, and hybrid), and different forms of visual outputs (e.g. graphs, histograms, lists, maps, cluster diagrams).

Not surprisingly, the commercial space is filled with a plethora of off the shelf visualization tools that range from simple desktop software packages to complex, web-based enterprise applications. This section presents several of the more prominent visualization tools based on the type of content used as input and highlights their relative strengths.

### 3.1.    Structured Content

The first category of visualization tools uses structured data as input. Content from data sources such as databases and bibliographies are included in this category.

> Example RefViz  (http://www.refviz.com): RefViz is a visualization tool that supports the analysis of structured data. It is designed to simultaneously search multiple online databases (such as Web of Science, Library of Congress), perform statistical and linguistic analyses to analyze the results for themes and patterns, and display thematic content through interactive visualizations.

> Example VantagePoint (http://www.thevantagepoint.com): VantagePoint is another visualization tool that supports the analysis of structured data. VantagePoint uses natural language processing algorithms to rapidly navigate through structured text to discover hidden relationships and patterns. One key strength of VantagePoint is the ability to ask who, what, where, and when questions that enable organizational profiling and technology assessment (Yang et al., 2008).

### 3.2.    Unstructured Content

The second category of visualization tools uses unstructured or abstract data as input. Content from data sources such as documents, email, reports, published articles, news stories, all are included in this category.

> Example OmniViz (http://www.biowisdom.com/content/omniviz): OmniViz is a visualization tool that supports the analysis of numerical and categorical information, and textual data. It is capable of analyzing millions of documents or numeric data points, and can perform case sensitive text analyses to differentiate compound names from common words. OmniViz combines sophisticated statistical and textual analysis algorithms with a range of visualizations to facilitate a deeper understanding of data. Through the integrated analysis environment, users can visually analyze textual, numerical, categorical and sequential data. OmniViz has been used in educational and a variety of research environments.

> Example TEMIS (http://www.temis.com/):TEMIS is another visualization tool that supports unstructured data analysis. TEMIS uses a symantic approach to extract, categorize and cluster textual data into contextual groups.

### 3.3. Hybrid Content

The third category of visualization tools uses both structured and unstructured data as input.

> Example Aureka
> (http://thomsonreuters.com/products_services/legal/legal_products/intellectual_property/Aureka):
> Aureka is a visualization tool that supports the analysis of both structured and unstructured data,
> and facilitates the organization and management of intellectual property (Yang et al., 2008).
> Aureka enables the research of full text data and uses concept mapping to reveal trends in the
> data. (Note: Aureka is specifically designed to analyze patent data; though the principles are
> transferable to other environments.)

### 4.    Practical Considerations

A major limitation of most visualization tools reveals either a limited connection to a theoretical
background, or an over-reliance on theory without a direct connection to user needs (Wohlfart et al.,
2008). It is the later, that presents a major concern for knowledge integration efforts. In fact, a common
"criticism of visualization research is that it presents techniques that are technically interesting but do
not provide solutions to real problems" (Stone, 2009, p. 47). Connecting the available tools to an actual
environment might help to make these tools more practical for knowledge integration tasks.

### 4.1.    Sample Data Management Environment

Consider a data management environment characterized by an increasing volume and complexity of
data on program activities, a heightened interest in evidence-based reports on program impact, and a
growing number of ad hoc information requests from various constituents. Although much of this
information is static (historical), periodic updates to project data are made at regular (and irregular)
intervals.

Coupled with the immense volume and varied sources, the data are a hybrid collection of structured and
unstructured (e.g. audio files) content, and are housed in disparate locations both within and outside
(e.g. private institutions) of the agency's control. Further, because different individuals (e.g. project
staff, external evaluators) create project specific data at different points in time, the data are not easily
combined.

Current information sharing practices are sufficient to provide a basic understanding of data in this type
of environment. They do not, however, facilitate knowledge integration. Nor do they support the
efficient response to requests for information regarding program impact, strategic decision-making, or
comprehensive program planning. Arguably, a data management environment with 21$^{st}$ century
challenges but not 21$^{st}$ century information and knowledge management practices, is limited. The use
of advanced data exploration and visualization tools might address the limitations of current knowledge
integration practices.

### 4.2.    Matching The Tool With The Environment

Although many features of commercial off the shelf tools are similar, they do vary in strengths and
weaknesses. In general, tools that support unstructured data analysis and visualization are the most
flexible (Yang, 2008). However, the use of symantic algorithms in these tools can require significant
staff training and an upfront investment in time and money. Table 1 below summarizes the strengths
and weaknesses of the tools highlighted above.

**Table 1:  Perceived Strengths And Weaknesses Of Existing Tools (Adapted from Yang, 2008)**

|  | Perceived Strength | Perceived Weakness |
| --- | --- | --- |
|  |  |  |

| RefViz | Bibliographic reference focus | Additional configuration required for data sources not included in application |
| --- | --- | --- |
| VantagePoint | Analytic capabilities | List cleaning of large datasets can be challenging |
| OmniViz | Interactive visualization | Focused primarily on visualization, not analysis |
| TEMIS | Data extraction method | Limited visualization capabilities |
| Aureka | Mapping, clustering, and citation analysis | Usability – difficult to understand labeling process |

In addition, other factors limit the practical usability of information visualization tools. In the past two decades there has been a significant focus on improving the speed of visualization tools so that users can perform interactive searches faster, explore bigger spaces and implore greater results more quickly (Chen et al., 2009). However, tools are often criticized because they do not embody basic usability constructs. For instance, basic visual design principles that guide the use of font size and color choices are often ignored (Stone, 2009).

Moreover, the context neutral visualization provided by the bulk of off the shelf tools is often in-adequate to meet the needs of the users in complex and volatile environments with a high volume of data (Chen et al., 2009). In these situations, general-purpose tools are largely unable to provide the usability features required of a complex data environment such as the one described above. Further, the lack of knowledge about visualization (i.e. what it can and cannot do) by the user is often a major obstacle when deploying visualization techniques (Chen et al., 2009). When the user does not have the proper training or expertise to specify certain functions or sufficient time or navigation skills to search all possible viewing positions, performance suffers and the value of the visualization tool is questioned. As with all new technologies, expectations must be managed and properly calibrated with the functionality of the tool.

### 4.3. Suggested Implementation Steps

Given the potential value for information visualization tools to positively impact knowledge integration activities, care should be taken during the implementation process. Implementation should begin with a data audit. The purpose of the audit is to answer key questions about the data, systems, and management environment. Generally, the audit is designed to answer questions such as: What data currently exist? Where are the data stored? How have the data been managed to date? Which of these data need to be integrated for the visualization? What are the challenges associated with this integration?

Several different methodologies exist to guide the data audit process. In general, however, a data audit procedure contains four primary phases (planning, identifying, assessing, and reporting):

*Planning the audit:* The two main objectives of the planning stage are to prepare for the audit in order to optimize on-site knowledge acquisition time; and to obtain organizational support through the establishment of a strong business case. By establishing expected outcomes with the organization's leaders the data auditors can determine the exact scope and focus of the audit.

Through the process of collecting background research the auditor is able to minimize demands placed on the data technicians, managers and users. Through this stage several key functions will take place including selection of an auditor(s), establishment of a business case, initial research to plan the audit and set up of the audit.

*Identifying and classifying data assets:* The purpose of the second phase is to establish what data assets exist and classify them according to their value for the organization. The auditor will inventory the data assets through a mapping exercise. Classification schemas will be established and modified as the process progresses. The classification will determine the scope of the future activities of the audit. Through this stage several key functions will take place including analysis of documentary sources, preparation for data asset inventory, data collection through interviews and questionnaire and approval and finalization of asset classification.

*Assessing the management of data assets:* Within the third stage the objective is to collect additional information about the data assets. It is important to find the key data assets that are central to the workings of the organization. By assessing the management of the data assets the auditor will be able to assess the current level of resources the organization has and determine if they are sufficient. Information collected will identify the data management practices. Based on the result of the audit, the gap analysis (difference between the current and desired states) will provide an accurate picture of required tasks for the development of the proposed customized visualization tool.

*Reporting findings and recommending change:* During this final stage of the audit, the results of the data audit are put together for a final report. The report will include recommended actions to improve the data management system and suggestions of appropriate information visualization tools that may be needed to enhance the organization's knowledge integration practices.

## 5.     Summary

Given the complexity of 21st century data management environments, and the capabilities of current information visualization tools, it is no surprise that analysts seek information visualization tools that facilitate information sharing and knowledge integration. Indeed, for many other contemporary organizations with vast amounts of increasingly complex data, the use of this type of tool has become a necessity rather than an option (Chen et al., 2009).

In today's climate, information visualization tools have risen to the forefront of conversations about how to effectively explore and display data. In fact, evidence of the increasing prominence of information visualization tools is apparent every time we turn on the news (e.g. the CNN magic election results board). However novel, in order to effective meet the knowledge integration demands of organizations and their constituents, tools must be adopted within the context of practical considerations such as the type of data to be examined, the type of visualization desired, the usability of the tool and the capability of the users. If properly considered, information visualization tools will prove to be a valuable knowledge integration tool.

## 6.     References

Bederson, B., Shneiderman, B. (Eds.). (2003). The craft of information visualization: Readings and reflections. San Francisco, CA: Morgan Kaufmann.

Bertin, J. (1983). Semiology of graphics: Diagrams, networks, maps (William J. Berg, Trans.). Madison, WI: University of Wisconsin Press. (Original work published 1967).

Card, S.K., Mackinlay, J., & Shneiderman, B. (1999). Readings in information visualization: Using vision to think. San Francisco, CA: Morgan Kaufmann.

Chen, M., Ebert, D., Laramee, R., van Liere, R., Ma, K., Ribarsky, W., Scheuermann, G., & Silver, D. (2009). Data, information, and knowledge in visualization. IEEE Computer Graphics and Applications, 29(1), 12-19.

Ferreira de Oliveira, M. and Levkowitz, H. (2003). From visual data exploration to visual data mining: A survey. IEEE Transactions on Visualization and Computer Graphics, 9(3), 378-394.

Few, S. (2008). A ménage a trois of data, eyes and mind. DM Review, 18(1), 14-34.

Friendly, M. (2008). Milestones in the history of thematic cartography, statistical graphics, and data visualization. Retrieved July 15, 2009 from http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf.

Keim, D., Andrienko, G., Fekete, J., Gorg, C., Kohlhammer, J. and Melancon, G. (2008). Visual Analytics: Definition, Process, and Challenges. in A. Kerren, J. Stasko, J. Fekete, & C. North (Eds.), Information visualization: Human-centered issues and perspective. Heidelberg, Germany: Springer.

Kerren, A., Stasko, J., Fekete, J. North, C. (Eds.). (2008). Information visualization: Human-centered issues and perspectives. Heidelberg, Germany: Springer.

Klein, G., Moon, B., & Hoffman, R. (2006). Making sense of sensemaking 1: Alternative perspectives. IEEE Intelligent Systems, 21(4), 70-73.

Klien, D., Andrienko, G., Fekete, J., Gorg, C., Kohlhammer, J., & Melancon, G. (2008). Visual analytics. In A. Kerren, J. Stasko, J. Fekete, & C. North (Eds.), Information visualization: Human-centered issues and perspective (pp. 154-175). Heidelberg, Germany: Springer.

Ribarsky, W., Stone, M., Dill, J., Fisher, B., MacEachren, A. (2005). The science of analytical reasoning: Perception and cognition. in J. Thomas & K. Cook (Eds.), Illuminating the Path-- Research and Development Agenda for Visual Analytics: IEEE Press.

Russell, D.M., Stefik, M., Pirolli, P., & Card, S. (1993). The cost structure of sensemaking, Proceedings of the INTERACT '93 and CHI '93 conference on human factors in computing systems (269-276). New York, NY: ACM.

Stone, M. (2009). Information visualization: Challenge for the humanities. In Working together or apart: Promoting the next generation of digital scholarship (43-56). Washington, DC: Council on Library and Information Resources.

Thomas J., & Cook, K. (2005). Illuminating the path: The research and development agenda for visual analytics. Los Alamitos, CA: IEEE Computer Society.

Wohlfart, E., Aigner, W., Bertone, A. & Miksch, S. (2008). Comparing information visualization tools focusing on the temporal dimensions. In IV 2008 - 12th International Conference on Information Visualization (69-74). London, UK: IEEE Computer Society.

Wong, P. , & Thomas, J. (2004). Visual analytics. IEEE Computer Graphics and Applications, 24(5), 20-21.

Yang, Y., Akers, L., Klose, T. & Yang, C. (2008). Text mining and visualization tools – Impressions of emerging capabilities. World Patent Information, 30, 280-293.

---

**About the Author:**

Dr. Diana Burley is Associate Professor in the Department of Human and Organizational Learning at The George Washington University. She serves as Co-Director of The George Washington University Institute for Knowledge and Innovation and the Vice-Chair of the ACM Special Interest Group on Computers and Society. Dr. Burley's research interests include knowledge management, IT workforce development and electronic governance.

Dr. Diana Burley, Department of Human and Organizational Learning, The George Washington University, 44983 Knoll Square, Suite 147, Ashburn, VA 20147; Tel: 703-726-3761; Email: dburley@gwu.edu