

## Cronfa - Swansea University Open Access Repository

---

This is an author produced version of a paper published in :  
*IEEE Computer Graphics and Applications*

Cronfa URL for this paper:  
<http://cronfa.swan.ac.uk/Record/cronfa179>

---

### Paper:

Chen, M., Ebert, D., Hagen, H., Laramée, R., Liere, R., Kwan-Liu, M., Ribarsky, W., Scheuermann, G. & Silver, D. (2009). Data, Information, and Knowledge in Visualization. *IEEE Computer Graphics and Applications*, 29(1), 12  
<http://dx.doi.org/10.1109/MCG.2009.6>

---

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.  
<http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/>



Min Chen,  
Swansea  
University

David Ebert,  
Purdue  
University

Hans Hagen,  
Technical  
University of  
Kaiserslautern

Robert S.  
Laramée,  
Swansea  
University

Robert van  
Liere,  
CWI,  
Amsterdam

Kwan-  
Liu Ma,  
University of  
California,  
Davis

William  
Ribarsky,  
University  
of North  
Carolina,  
Charlotte

Gerik  
Scheuermann,  
University of  
Leipzig

Deborah  
Silver,  
Rutgers  
University

## Data, Information, and Knowledge in Visualization

In visualization, we use the terms *data*, *information* and *knowledge* extensively, often in an interrelated context. In many cases, they indicate different levels of abstraction, understanding, or truthfulness. For example, “visualization is concerned with exploring data and information,”<sup>1</sup> “the primary objective in data visualization is to gain insight into an information space,”<sup>1</sup> and “information visualization” is for “data mining and knowledge discovery.”<sup>2</sup> In other cases, these three terms indicate data types, for instance, as adjectives in noun phrases, such as data visualization, information visualization, and knowledge visualization. These examples suggest that data, information, and knowledge could serve as both the input and output of a visualization process, raising questions about their exact role in visualization.

There are many competing definitions of data, information, and knowledge, in different aspects of computer science and engineering and in other disciplines such as psychology, management sciences, and epistemology (the theory of knowledge).<sup>3</sup> The use of the three terms isn’t consistent and is often conflicting. For instance, data and information are often interchangeable in computing (for example, data processing and information processing or data management and information management). From a systems perspective, however, data is referred to as bits and bytes stored on or communicated via a digital medium. So any computerized representations, including knowledge representations, are types of data. On the other hand, from the perspective of knowledge-based systems, data is a simpler form of knowledge.

Researchers have attempted to clarify the taxonomy of terms used in the visualization community (for example, in the work of Ed H. Chi,<sup>4</sup> Ben Shneiderman,<sup>5</sup> and Melanie Tory and Torsten Möller<sup>6</sup>). However, the terms data, information and knowledge remain ambiguous. This article doesn’t attempt to offer a different taxonomy for visualization. Instead, we differentiate these three terms from the perspective of visualization pro-

cesses. Furthermore, we examine the current and future role of information and knowledge in the development of visualization technology.

### Definitions of Data, Information and Knowledge

Since we can read data, grasp information and acquire knowledge, we must differentiate these terms in the perceptual and cognitive space. Because we can also store data, information and knowledge in the computer, we must also differentiate them in the computational space.

#### Perceptual and Cognitive Space

The data-information-knowledge-wisdom (DIKW) hierarchy<sup>7</sup> is a popular model for classifying human understanding in the perceptual and cognitive space. The origin of this hierarchy can be traced to the poet T.S. Eliot.<sup>8</sup> Table 1 shows Russell Ackoff’s definitions of data, information, and knowledge.<sup>7</sup>

Let  $\mathbb{P}$  be the set of all possible explicit and implicit human memory. The former encompasses the memory of events, facts, and concepts, and the understanding of their meanings, context, and associations. The latter encompasses all nonconscious forms of memory, such as emotional responses, skills, and habits.<sup>9</sup> We can thus focus on three subsets of memory,  $\mathbb{P}_{\text{data}} \subset \mathbb{P}$ ,  $\mathbb{P}_{\text{info}} \subset \mathbb{P}$ , and  $\mathbb{P}_{\text{know}} \subset \mathbb{P}$ , where  $\mathbb{P}_{\text{data}}$ ,  $\mathbb{P}_{\text{info}}$ , and  $\mathbb{P}_{\text{know}}$  are the sets of all possible explicit and implicit memory about data, information, and knowledge, respectively.

Despite the lack of an agreeable set of the definitions of data, information, and knowledge, a consensus exists that data isn’t information, and information isn’t knowledge. Without diverting from this article’s scope, here we simply assume that  $\mathbb{P}_{\text{data}}$ ,  $\mathbb{P}_{\text{info}}$ , and  $\mathbb{P}_{\text{know}}$  aren’t mutually disjoint and none of them is a subset of another. Without losing generality, we can generalize  $\mathbb{P}_{\text{know}}$  to include wisdom, and any other high level of understanding, in the context of the DIKW hierarchy.

### Computational Space

Let  $\mathbb{C}$  be the set of all possible representations in computer memory. Similarly, we can consider three subsets of representations,  $\mathbb{C}_{data}$ ,  $\mathbb{C}_{info}$ , and  $\mathbb{C}_{know}$ . However, data is an overloaded term in computing. For example, it's common to treat programs as a special class of data. In many cases, it isn't possible to distinguish programs from other data. Applying the same analogy, a computer representation of a piece of information or knowledge is just a particular form of data. A computer representation of visualization is also a form of visual data.

We hence propose to use the definitions in Table 2 for the following discussions. With such definitions, we have  $\mathbb{C}_{data} = \mathbb{C}$ ,  $\mathbb{C}_{info} \subset \mathbb{C}_{data}$ , and  $\mathbb{C}_{know} \subset \mathbb{C}_{data}$ . We can easily extend the definitions in Table 2 to include categories of raw data ( $\mathbb{C}_{rawdata}$ ), volume data ( $\mathbb{C}_{volume}$ ), flow data ( $\mathbb{C}_{flow}$ ), software ( $\mathbb{C}_{software}$ ), videos ( $\mathbb{C}_{video}$ ), mathematical models ( $\mathbb{C}_{mathmodel}$ ), visual data ( $\mathbb{C}_{image}$ ), and so forth. This also makes sense when using the category names as the adjectives in noun phases, such as volume visualization and software visualization.

Figure 1 shows a typical visualization process, illustrating instances of data, information, and knowledge in both computational space and perceptual and cognitive space. Hence, the need for visualization is based on the difficulties humans face in acquiring a sufficient amount of information ( $P_{info} \subset \mathbb{P}_{info}$ ) or knowledge ( $P_{know} \subset \mathbb{P}_{know}$ ) directly from a data set ( $\mathbb{C}_{data} \subset \mathbb{C}_{data}$ ). The process of creating visualization is a function that maps from  $\mathbb{C}_{data}$  to the set of all imagery data,  $\mathbb{C}_{image}$ . It transforms a data set  $\mathbb{C}_{data}$  to a visual representation  $\mathbb{C}_{image}$ , which facilitates a more efficient and effective cognitive process for acquiring  $P_{info}$  and  $P_{know}$ .

### A Visualization Process is a Search Process

Given a data set  $\mathbb{C}_{data}$ , a user first makes decisions

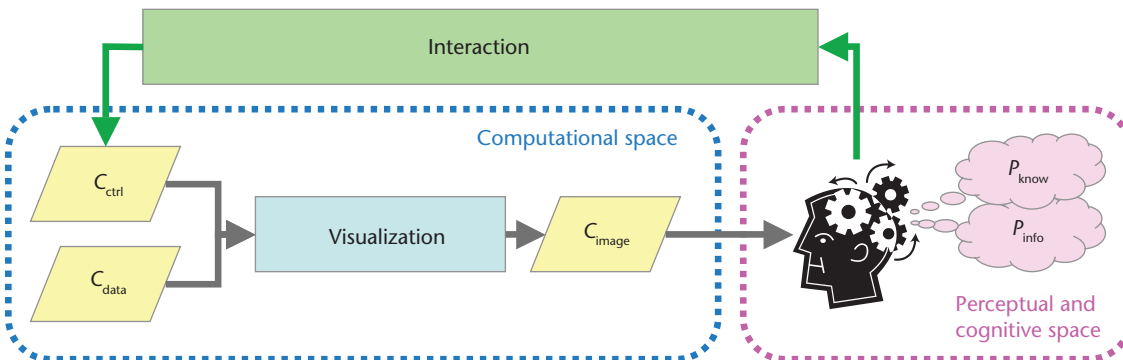


Figure 1. A typical visualization process, where interaction provides the primary means for reducing the search space in visual exploration.  $\mathbb{C}_{data}$ ,  $\mathbb{C}_{ctrl}$ , and  $\mathbb{C}_{image}$  denote input data, control parameters and visualization results stored in computer memory, respectively.  $P_{info}$  and  $P_{know}$  represent the information and knowledge acquired by the user.

Table 1. Russell Ackoff's definitions of data, information, and knowledge in perceptual and cognitive space.<sup>7</sup>

| Category    | Definition                                                                                              |
|-------------|---------------------------------------------------------------------------------------------------------|
| Data        | Symbols                                                                                                 |
| Information | Data that are processed to be useful, providing answers to "who," "what," "where," and "when" questions |
| Knowledge   | Application of data and information, providing answers to "how" questions                               |

Table 2. Building time as a function of data size.

| Category    | Definition                                                                                                                                                                                          |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data        | Symbols                                                                                                                                                                                             |
| Information | Data that represents the results of a computational process, such as statistical analysis, for assigning meanings to the data, or the transcripts of some meanings assigned by human beings         |
| Knowledge   | Data that represents the results of a computer-simulated cognitive process, such as perception, learning, association, and reasoning, or the transcripts of some knowledge acquired by human beings |

about which visualization tools to use for exploring the data set. The user then experiments with different controls, such as styles, layout, viewing position, color maps, and transfer functions, until he or she obtains a satisfactory collection of visualization results,  $\mathbb{C}_{image}$ . Depending on the visualization tasks, satisfaction can come in many forms. For example, the user may have obtained sufficient information or knowledge about the dataset, or may have obtained the most appropriate illustration about the data to assist others in the knowledge acquisition process.

Such a visualization process is fundamentally the same as a typical search process, except that it is usually much more complex than plugging a few keywords into a search engine. In visualization, the tools for the "search" tasks are usually application-specific (for example, network, flow,

Table 3. Examples of information used in visualization.

| Information categories                              | Examples                                                                                                 |
|-----------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| Information about the input data set                |                                                                                                          |
| Abstract geometric and temporal characteristics     | Skeletons, features, events                                                                              |
| Topological properties                              | Contour tree for volume data, vector field topology, tracking graph for time-varying data                |
| Statistical indicators and information measurements | Histogram, correlation, importance, certainty, entropy, mutual information, local statistical complexity |
| Information about the results                       | Color histogram, level of cluttering                                                                     |
| Information about the process                       | Interaction patterns, provenance                                                                         |
| Information about users' perceptions                | Response time, accuracy                                                                                  |

volume visualization). The parameter space for the “search” is normally huge (for example, exploring many viewing positions or trying out many different transfer functions). The user interaction for the “search” sometimes can be very slow, especially in handling very large datasets. Figure 1 depicts the process with a large interaction box that connects from the user to the control parameters,  $C_{ctrl}$ , which are also data.

In fact, over the past two decades, much emphasis has been placed on improving the speed of visualization tools, so users can carry out interactive “searches” faster, explore bigger parameter spaces, and hopefully find satisfactory results quicker.

However, with the growing amount of data and increasing availability of different visualization techniques, the search space for a visualization process is also expanding. Like the Internet search

problem, interactive visualization alone is no longer adequate.

### Information-Assisted Visualization

In recent years, researchers have introduced an assortment of techniques for visualizing complex features in data by relying on information abstracted from the data. Here, we consider  $C_{info}$  in the computational space as well as  $P_{info}$  in the perceptual and cognitive space. Figure 2 illustrates an information-assisted visualization process. Some techniques use information captured in the visualization process to improve visualization efficiency and effectiveness. Table 3 gives examples of such information.

In information-assisted visualization, the system provides the user with a second visualization pipeline (see Figure 2), which typically displays the information about the input dataset. But it can also present attributes of the visualization process, the properties of the results, or characteristics of the user’s perceptual behaviors. The user uses such information to reduce the search space for optimal control parameters, hence making the interaction much more cost effective.

Such techniques provide an intrinsic interface between the scientific-visualization and information-visualization communities. With the increasing size and complexity of data, the use of information to aid visualization will inevitably become a necessity rather than an option.

### Knowledge-Assisted Visualization

In a visualization process, the user’s knowledge is an indispensable part of visualization. For in-

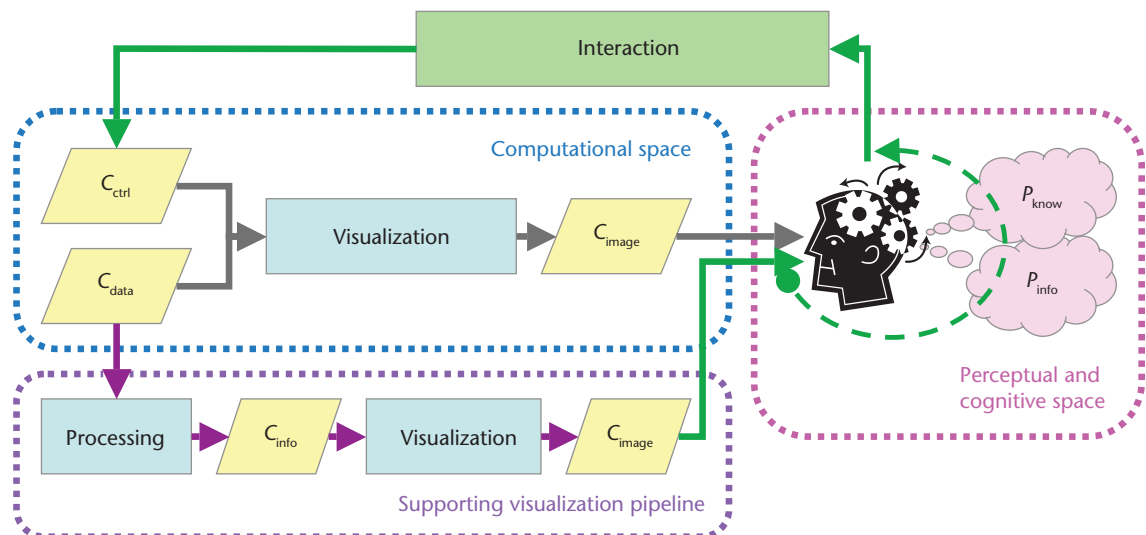


Figure 2. Information-assisted visualization, where an additional pipeline displays information about the input data to help the user reduce the search space in the main visualization process.

## Resolving Ambiguity Using the Set Notations

We can resolve the ambiguity in various statements that consist of the terms *data*, *information* and *knowledge* by tagging such terms using the set notations,  $\mathbb{P}$ ,  $\mathbb{C}$ , and their subsets. The following list of definitions from various publications is given below to help make these terms clearer:

Data [ $\mathbb{C}_{data}$ ]: a representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by human beings or by automatic means.<sup>1</sup>

Information [ $\mathbb{P}_{info}$  or  $\mathbb{C}_{info}$ ]: the meaning that is currently assigned [by human beings or computers] to data [ $\mathbb{C}_{data}$ ] by means of the conventions applied to those data [ $\mathbb{C}_{data}$ ].<sup>1</sup>

A useful definition of visualization might be the binding (or mapping) of data [ $\mathbb{C}_{data}$ ] to representations [ $\mathbb{C}_{image}$ ,  $\mathbb{C}_{auditory}$ ,  $\mathbb{C}_{tactile}$ , and so on] that can be perceived. The types of bindings could be visual, auditory, tactile, etc., or a combination of these.<sup>2</sup>

If researchers try to read the data [ $\mathbb{C}_{data}$ ], usually presented as vast numeric matrices, they will take in the information [ $\mathbb{P}_{info}$ ] at snail's pace. If the information [ $\mathbb{C}_{info}$ ] is rendered graphically, however, they can assimilate it at a much faster rate.<sup>3</sup>

[Visualization] transforms the symbolic [ $\mathbb{C}_{data}$ ] into the geometric [ $\mathbb{C}_{image}$ ], enabling researchers to observe their simulations and computations.<sup>4</sup>

[Information ( $\mathbb{C}_{info}$ ) visualization is] the use of computer-supported, interactive, visual representations [ $\mathbb{C}_{image}$ ] of abstract data [ $\mathbb{C}_{info}$ ] to amplify cognition.<sup>5</sup>

Information [ $\mathbb{P}_{info}$  or  $\mathbb{C}_{info}$ ] is born when data [ $\mathbb{C}_{data}$ ] are interpreted [by human beings or computers].<sup>6</sup>

Information [ $\mathbb{P}_{info}$  and  $\mathbb{C}_{info}$ ] has both qualitative and quantitative aspects.<sup>7</sup>

The amount of information [ $\mathbb{P}_{info}$  and  $\mathbb{C}_{info}$ ] conveyed in an event depends on the probability of the event.<sup>7</sup>

Knowledge [ $\mathbb{C}_{know}$ ] is the symbolic representation of aspects of some named universe of discourse ... We define

data [ $\mathbb{C}_{facts}$  or  $\mathbb{C}_{rawdata}$  but not  $\mathbb{C}_{data}$  because  $\mathbb{C}_{know} \subset \mathbb{C}_{data}$ ] as the symbolic representation of simple aspects of some named universe of discourse ... The amount of information [ $\mathbb{P}_{info}$ ] obtained by the receiver of a message is related to the amount by which that message reduces receiver's uncertainty about some aspect of the universe of discourse (Shannon).<sup>8</sup>

Knowledge [ $\mathbb{P}_{know}$ ]: understanding, awareness, or familiarity acquired through education or experience. Anything that has been learned, perceived, discovered, inferred, or understood. The ability to use information [ $\mathbb{P}_{info}$  and/or  $\mathbb{C}_{info}$ ].<sup>9</sup>

Knowledge base: the assembly of all the information [ $\mathbb{C}_{info}$ ] and knowledge [ $\mathbb{C}_{know}$ ] of a specific field of interest.<sup>9</sup>

---

### References

1. *Directory for Information Systems*, X3.172, American Nat'l Standards Inst., 1990.
2. J. Foley and B. Ribarsky, "Next-Generation Data Visualization Tools," *Scientific Visualization: Advances and Challenges*, L. Rosenblum et al., eds., Academic Press, 1994.
3. R.M. Friedhoff and T. Kiley, "The Eye of the Beholder," *Computer Graphics World*, vol. 13, no. 8, 1990, pp. 46–59.
4. B.H. McCormick, T.A. DeFanti, and M.D. Brown, eds., "Visualization in Scientific Computing," *Computer Graphics*, vol. 21, no. 6,.
5. S. Card, J. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann, 1999.
6. W. Stallings, *Data and Computer Communications*, 4th ed., Macmillan, 1994.
7. M.J. Usher, *Information Theory for Information Technologists*, Macmillan, 1984.
8. R.A. Frost, *Introduction to Knowledge Base Systems*, Collins, 1986.
9. E. Turban, *Decision Support and Expert Systems*, Prentice-Hall, 1995.

stance, the user might assign specific colors to different objects in visualization according to certain domain knowledge. The user might also choose different viewing positions because the visualization results can reveal more meaningful information or a more problematic scenario that requires further investigation.

Meanwhile, the lack of certain user knowledge is often a major obstacle in deploying visualization techniques. The user might not have received adequate training about how to specify transfer functions, or might not have sufficient time or navigation skills to explore all possible viewing positions.

Both scenarios suggest the need for knowledge-assisted visualization. The objectives of knowl-

edge-assisted visualization include sharing domain knowledge among different users and reducing the burden on users to acquire knowledge about complex visualization techniques. It also enables the visualization community to learn and model the best practice, so that powerful visualization infrastructures can develop and evolve.

Researchers and developers often incorporate some general or domain knowledge into visualization systems, either intentionally or unintentionally. For example, a default transfer function in a volume visualization system may capture the domain knowledge about a specific modality. If a visualization system could collect a large repository of such knowledge, it could then choose an ap-

## Examples of Visualization

While there are many examples of information-assisted visualization, the development of knowledge-assisted visualization is in its infancy. Here we selectively describe several examples of information-assisted visualization in the literature, while accentuating the use, or potential use, of knowledge in a few visualization systems. These examples are intended to reinforce the viewpoints of this article, rather than provide a comprehensive survey.

### Information-Assisted Visualization

The following examples show four different approaches to information-assisted visualization. The first three approaches make use of geometrical, topological and statistical information about the input data, respectively. The last approach guides the user by measuring the quality of visualization results.

#### *Curve-Skeleton*

Curve-skeletons are 1D geometrical representations abstracted from 3D objects in an input data set. Such information can be used to aid visualization tasks, including virtual navigation, reduced-model formulation, visualization improvement, and animation. For example, in virtual endoscopy, curve-skeletons can specify collision-free paths for navigation through human organs.<sup>1</sup>

#### *Isosurface Topology*

Isosurface topology, which is typically represented as a contour tree, provides abstract insight into the structural relationship and connectivity between isosurfaces in a data set. In volume visualization, such information can help users distinguish features in different topological zones, comprehend complex relationships between isosurfaces, and design effective transfer functions.<sup>2</sup>

#### *Local Statistical Complexity*

Local statistical complexity (LSC) is an information-theoretic measure that tells how much information from the local past is required to predict the dynamics in the local future. Given a time-varying data set, we can assign each data point an LSC value. Higher LSC values indicate regions that feature an extraordinary temporal evolution, whereas, lower values indicate temporal patterns that occur frequently in the data set.<sup>3</sup>

Figure A demonstrates how such information can help users generate a visualization that highlights temporally important features.

#### *Data Abstraction Quality*

Measuring the quality of visualization results, such as visual density and clutter, provides users with useful guidance in synthesizing the most effective visualization. One measurement is data abstraction quality, measuring the degree to which the visualization results convey the original data set. Such information enables users to determine the optimal abstraction level for a given visualization task. It can also help the user compare different visualization methods in terms of their capability of maintaining dominant characteristics of the original dataset while reducing the data's size and detail.<sup>4</sup>

### Knowledge-Assisted Visualization

The following examples represent visualization systems that exhibit some key features of knowledge-assisted visualization. They demonstrate both the feasibility and potential of knowledge-assisted visualization.

#### *Viewpoint Mutual Information*

From Figures 2 and 3 in the main article, we can observe that one transition path of information-assisted visualization to knowledge-assisted visualization is to automate reasoning about the information abstracted from the input data. Ivan Viola and his colleagues give a classic example of such a transition, where viewpoint mutual information (VMI) that measures the dependence or correlation between a set of viewpoints and a set of objects in a data set is used to determine the optimal viewpoint.<sup>5</sup> The fundamental difference between this approach and the above-mentioned examples of information-assisted visualization is that users don't make decisions according to the processed VMI. Instead, a relatively simple rule for minimizing VMI is used to determine viewpoint transformation automatically. Such a rule can be seen as a piece of knowledge hard-coded in the system.

#### *Predetermined Ranking*

Jock D. Mackinlay, Pat Hanrahan and Chris Stolte capture

appropriate transfer function based on the attributes of an input data set. Figure 3 (page xx) shows a visualization pipeline supported by a knowledge base ( $C_{\text{know}}$ ), that stores knowledge representations captured from expert users. The system can use rule-based reasoning to establish an appropriate set, or several optional sets, of control parameters which can significantly reduce the search space, especially for inexperienced users. The system

component for reasoning is commonly called an inference engine in knowledge-based systems (or expert systems).

The shortcomings of such a system include the difficulties in specifying comprehensively what knowledge to capture and the inconvenience of collecting knowledge from experts. This constrains the deployment of such a system to specific application domains.

a noticeable amount of generic knowledge as ranks of different visualization designs.<sup>6</sup> This enables the visualization system to automatically take users through a design process for creating a visualization. The stored ranks and ranking conditions are essentially a collection of expert knowledge.

### Ontology Mapping

The determination of visualization designs and parameters should depend on the input data. One approach is to extract semantic information from the input data and try to find the best match with the semantic information of visualization designs (for example, treemaps and graphs) and the associated parameters (for example, size and axes). Owen Gilson and his colleagues<sup>7</sup> use three ontologies, which are knowledge representations, to store (a) the domain-specific semantics about a class of input data, (b) the semantics about available visualization designs, and (c) the ontological mapping from (a) to (b). With these three ontologies, the system ranks different visualization designs dynamically according to the input data, and presents a set of highly-ranked visualization designs to the user automatically.

### Workflow Management

VisTrails is a visualization infrastructure that provides users with workflow management.<sup>8</sup> It can capture and store a huge amount of data about input data sets, user interaction, and visualization results in visualization processes. VisTrails exhibits some of the primary characteristics of the knowledge-supporting infrastructure shown in Figure 4 in the main article, though it currently has limited automated reasoning. Such an infrastructure has great potential to be developed into an infrastructure for knowledge-assisted visualization.

### References

1. N.D. Cornea, D. Silver, and P. Min, "Curve-Skeleton Properties, Applications, and Algorithms," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 3, 2007, pp. 530–548.
2. G.H. Weber et al., "Topology-Controlled Volume Rendering," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 2, 2007, pp. 330–341.

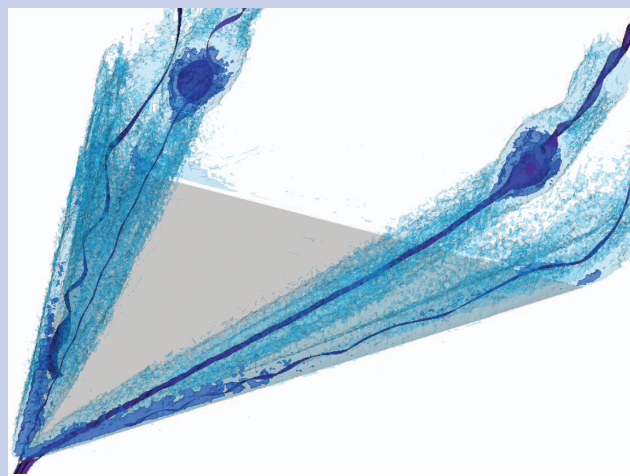


Figure A. The local statistical complexity (LSC) of flow around a delta wing (the gray triangle). Four stream surfaces indicate the vortices on top of the wing. The two isosurfaces in blue and light blue separate regions that hold LSC values within the range [14.7;15] and [11;15], respectively. High LSC values point the user to distinctive regions that might feature significant temporal events. (Source: Heike Jänicke, University of Leipzig, copyright 2008, used with permission.)<sup>3</sup>

3. H. Jänicke et al., "Multifield Visualization Using Local Statistical Complexity," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, 2007, pp. 1384–1391.
4. Q. Cui et al., "Measuring Data Abstraction Quality in Multiresolution Visualizations," *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 5, 2006, pp. 709–716.
5. I. Viola et al., "Importance-Driven Focus of Attention," *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 5, 2006, pp. 933–940.
6. J.D. Mackinlay, P. Hanrahan, and C. Stolte, "Show Me: Automatic Presentation for Visual Analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, 2007, pp. 1137–1144.
7. O. Gilson et al., "From Web Data to Visualization via Ontology Mapping," *Computer Graphics Forum*, vol. 27, no. 3, 2008, pp. 959–966.
8. L. Bavoil et al., "VisTrails: Enabling Interactive Multiple-View Visualizations," *Proc. IEEE Visualization*, IEEE Computer Society Press, 2005, pp. 135–142.

An alternative approach is to establish a visualization infrastructure, where the system can systematically collect, process and analyze data about visualization processes. Using case-based reasoning, it can infer knowledge from cases of successes and failures, common associations between data sets and control parameters, and many other patterns exhibited by visualization tasks, tools, users, and interactions. Such knowledge might include

a popular approach, commonly used parameter sets, the best practice, an optimization strategy, and so forth. Figure 4 (page xx) shows such an infrastructure.

Such an infrastructure is general purpose and can support multiple application domains. It can potentially enable applications to benefit from the best practice as well as software developed for other applications. Developers can build such an

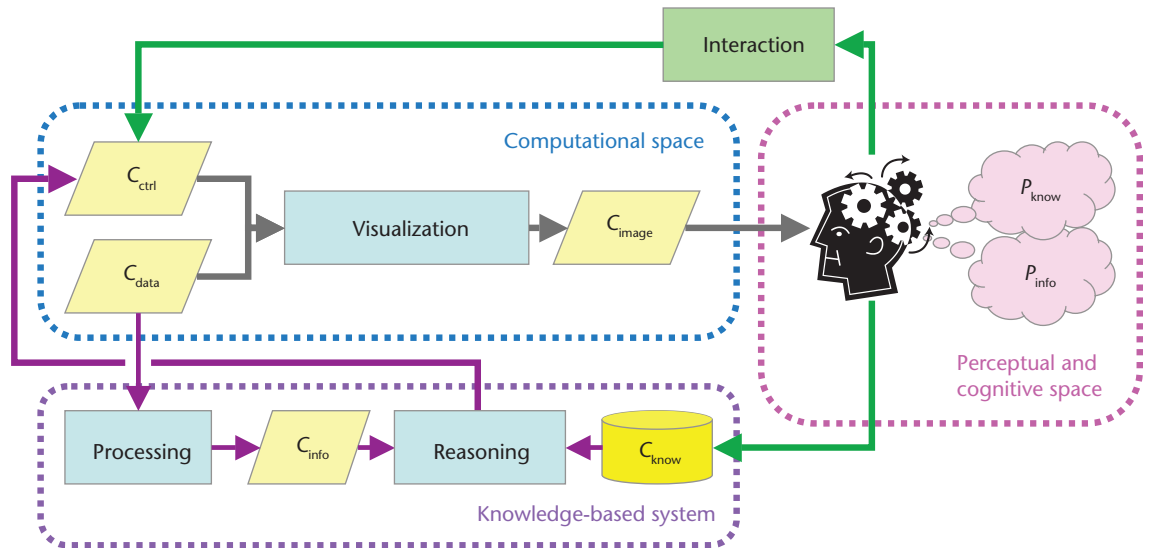


Figure 3. Knowledge-assisted visualization with acquired knowledge representations. The system stores expert knowledge about specific applications and complex visualization techniques, and uses such knowledge, in conjunction with rule-based reasoning, to automate part of a visualization process.

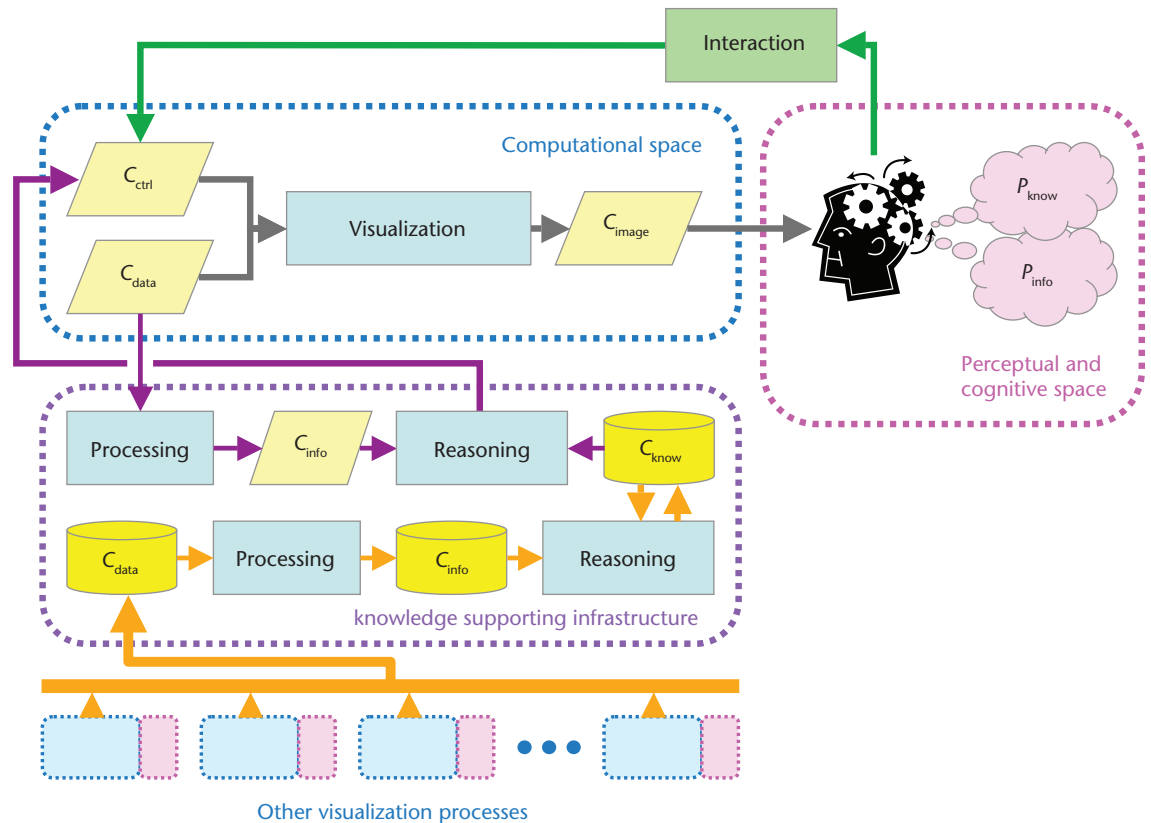


Figure 4. Knowledge-assisted visualization with simulated cognitive processing. The system makes use of the data passing through the visualization pipeline over time, and transforms such data to knowledge using case-based reasoning. This alleviates the difficulties of transcribing the knowledge of expert users.

infrastructure upon advances in other areas of computing technologies, including semantic computing, autonomic computing, knowledge-based systems, data warehousing, machine learning, and search engine optimization.

The development of visualization could follow a path similar to other computing technologies, such as speech processing, computer vision, and Web technology. Thus, one likely development path for visualization is

- from offline visualization,
- to interactive visualization,
- to information-assisted visualization, and
- to knowledge-assisted visualization.

Interactive visualization has matured, whereas information-assisted visualization is still undergoing a significant amount of development. With a large amount of information being collected locally and globally, a transition to knowledge-assisted visualization is inevitable.

As a discipline, visualization has thrived on helping application users transfer data ( $\mathbb{C}_{\text{data}}$ ) in the computational space to information ( $\mathbb{P}_{\text{info}}$ ) and knowledge ( $\mathbb{P}_{\text{know}}$ ) in the perceptual and cognitive space. As a discipline, we need infrastructures to collect data about visualization processes and to transfer this data to information and knowledge to further our understanding and enhance visualization technology..

#### Acknowledgments

*We thank Heike Jänicke and Gerhard Brewka at the University of Leipzig, and Phil Grant and John Sharp at Swansea University for their advice and comments on terminologies used in information theory, knowledge-based systems, and other aspects of computing.*

#### References

1. G.S. Owen et al., "Definitions and Rationale for Visualization," Apr. 2008; [www.siggraph.org/education/materials/HyperVis/visgoals/visgoal2.htm](http://www.siggraph.org/education/materials/HyperVis/visgoals/visgoal2.htm).
2. U. Fayyad, G.G. Grinstein, and A. Wierse, *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann, 2002.
3. K. Lehrer, *Theory of Knowledge*, Westview, 1990.
4. E.H. Chi, "A Taxonomy of Visualization Techniques Using the Data State Reference Model," *Proc. IEEE Symp. Information Visualization*, IEEE Press, 2000, pp. 69–75.
5. B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," *Proc. IEEE Symp. Visual Languages*, IEEE Press, 1996, pp. 336–343.
6. M. Tory and T. Möller, "Rethinking Visualization: A High-level Taxonomy," *Proc. IEEE Symp. Information Visualization*, IEEE Press, 2004, pp.151–158.
7. R.L. Ackoff, "From Data to Wisdom," *J. Applied Systems Analysis*, vol. 16, 1989, pp. 3–9.
8. N. Sharma, "The Origin of the 'Data Information Knowledge Wisdom' Hierarchy," Apr. 2008; [www-personal.si.umich.edu/~nsharma/dikw\\_origin.htm](http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm).

## CG&A Special Issue on Knowledge-Assisted Visualization

IEEE Computer Graphics and Applications seeks research articles in the area of knowledge-assisted visualization for an upcoming special issue. For further details, please visit [www.computer.org/portal/site/cga](http://www.computer.org/portal/site/cga).

9. E.E. Smith and S.M. Kosslyn, *Cognitive Psychology: Mind and Brain*, Pearson Prentice-Hall, 2007.

**Min Chen** is a professor at Swansea University. Contact him at [m.chen@swansea.ac.uk](mailto:m.chen@swansea.ac.uk).

**David Ebert** is a professor at Purdue University. Contact him at [ebertd@ecn.purdue.edu](mailto:ebertd@ecn.purdue.edu).

**Hans Hagen** is a professor at the Technical University of Kaiserslautern. Contact him at [hagen@informatik.uni-kl.de](mailto:hagen@informatik.uni-kl.de).

**Robert S. Laramée** is a lecturer at Swansea University. Contact him at [r.s.laramee@swansea.ac.uk](mailto:r.s.laramee@swansea.ac.uk).

**Robert van Liere** is a senior researcher at CWI, Amsterdam and a professor at the Technical University in Eindhoven. Contact him at [robert.van.liere@cwi.nl](mailto:robert.van.liere@cwi.nl).

**Kwan-Liu Ma** is a professor at the University of California, Davis. Contact him at [ma@cs.ucdavis.edu](mailto:ma@cs.ucdavis.edu).

**William Ribarsky** is a professor at the University of North Carolina, Charlotte. Contact him at [ribarsky@uncc.edu](mailto:ribarsky@uncc.edu)

**Gerik Scheuermann** is a professor at the University of Leipzig. Contact him at [scheuermann@informatik.uni-leipzig.de](mailto:scheuermann@informatik.uni-leipzig.de).

**Deborah Silver** is a professor at Rutgers University. Contact her at [silver@caip.rutgers.edu](mailto:silver@caip.rutgers.edu).