Cooperative visualization: A design case

Article *in* Library Hi Tech · June 2013

DOI: 10.1108/07378831311329112

1 author:

Nathan Prestopnik
Ithaca College

Cooperative visualization: a design case

Nathan Prestopnik
School of Information Studies, Syracuse University, Syracuse, New York, USA

Received 21 February 2013
Revised 21 February 2013
Accepted 21 February 2013

Abstract

Purpose – In this design case, a participatory approach to visualizing a complex computational pipeline was adopted, with the goal of exploring what benefits might be derived when groups of people visualize complex information for themselves.

Design/methodology/approach – Several visualization artifacts were developed to support collaborative process at the Laser Interferometer Gravitational Wave Observatory (LIGO). Researchers adopted a participatory approach, engaging directly in LIGO activities and drawing together explicitly codified data from the LIGO computational pipeline as well as structural knowledge tacitly held by project scientists. Both sources of information were critical to producing meaningful visualizations and progressing design and research efforts.

Findings – This design case revealed several benefits realized when individuals or groups visualize information for themselves, especially improved communication and enhanced understanding of complex systems of information.

Originality/value – This design case demonstrates how cooperatively creating visualizations can enhance understanding and support group activities and goals. It is also a call to move beyond data, technologies, and techniques to introduce more human-centered approaches within visualization scholarship.

Keywords Information visualization, Interaction, Design, Cooperative work, Information management

Paper type Case study

Introduction

In the current body of information visualization (IV) literature, there are two broad emphases: representation and interaction (Yi *et al.*, 2007). Representation research explores ways to visually represent information on display devices, including algorithms, techniques, and technologies (e.g. Card *et al.*, 1991; Feiner and Clifford, 1990; Fekete and Plaisant, 2002; Furnas, 1986; Mackinlay *et al.*, 1991; Robertson *et al.*, 1991). Interaction research studies the dialog that occurs between users of a visual information system and the system itself. Interactions might include filtering data to order it in various ways, drilling down through a display to different levels of detail, zooming, panning, or otherwise manipulating the visual display to achieve the view or

Published with the kind permission of iSchools, http://ischools.org

A version of this article was originally presented at the iConference 2013 held in Fort Worth, Texas, 12-15 February 2013.

The author would like to acknowledge Duncan Brown, Peter Couvares, Steven Rowe, and Howard Turtle, co-designers involved in the visualization process described in this work. The author would also like to thank the Syracuse University Gravitational Wave Group and the Syracuse University Center for Natural Language Processing for participation and insights over the course of developing this design case. This work represents the author's view and has not been reviewed by the LIGO Scientific Collaboration.

Library Hi Tech
Vol. 31 No. 2, 2013
pp. 371-390

0737-8831
DOI 10.1108/07378831311329112

perspective that the user is interested in. Of the two emphases, interaction and representation, representation has received by far the most scholarly attention from IV researchers (Yi *et al.*, 2007; Ellis and Dix, 2006; Tory and Möller, 2004; Thomas and Cook, 2005; Chen and Czerwinski, 2000; Chen, 2005), but interaction is of growing interest in the IV community (Chen, 2005). The focus of this present research is on interaction.

Visualization interactions can be split into three categories:

(1) representational interactions;

(2) cognitive interactions; and

(3) creative interactions.

Representational interactions include the many types already mentioned; they provide the user opportunities to modify the visual display and organization of information on the screen. Cognitive interactions, described through a variety of models and theories of visualization (e.g. Spence, 2001, 2007; Ware, 2004; Chen, 2003, 2006; Card *et al.*, 1999), are the purely mental activities that a user will undertake when working with a visualization, setting aside any affordances the visualization artifact itself might have for manipulating the representation of data.

Creative interactions are defined by human involvement in the creative activity of generating visualizations, with the term "creative" indicating the act of creation – the transformation of raw information into a visual representation. Creative interactions can be highly complex and require a great deal of cognitive effort (for example, manually transforming qualitative or quantitative information into something that can be visualized). Creative interactions are less well studied than other forms of interaction, but they are important. Interactions that result in the formation of a visualization will necessarily require individuals to cognitively and representationally interact with raw information and the visualization itself. The same is true of groups of individuals who cooperate together to produce visualizations, but with additional possible benefits, including the use of collaboratively developed visualizations to foster discussion and achieve group goals.

Broadly, this present research explores the following: What do groups of people gain by visualizing information for themselves? There are many possibilities, including cognitive or analytical benefits, the ability to cooperate and share information more effectively, and the ability to draw new connections or see new patterns within one or more data sets. Using participatory research techniques, these possibilities are explored in a design case: the documented effort to visualize a complex computational pipeline used by astronomers for signal/noise processing of gravitational wave detector data. The different visualizations described in this design case were produced under the aegis of ongoing development of an information retrieval (IR) system to be used by gravitational wave physicists as part of their work. Key users of the visualizations included the astronomers involved in the Laser Interferometer Gravitational Wave Observatory (LIGO) collaboration as well as the IR researchers responsible for producing the retrieval system for this collaboration and the IV researcher (the author) tasked with studying various use scenarios and creating the visualizations themselves.

The LIGO design case revealed a number of concepts that are important elements of creative visualization interactions. Tacit and explicit knowledge (Collins, 2007;

Polanyi, 1966; Miller, 2008) turned out to be useful framing concepts for understanding the differences between information that different individuals bring to a creative visualization exercise themselves vs the information that may already be codified in the data to be visualized. Underlying data from the LSC's compact binary search pipeline explicitly contained important information to be visualized, but the scientists who use this pipeline and originally designed various aspects of it also tacitly held knowledge that was critically necessary in order to meaningfully visualize the entire computational process. Only in combination could tacit and explicit information actually be used to produce meaningful visual artifacts.

Intermediary visual artifacts also emerged as a key aspect of the creative visualization interaction. LIGO scientists, IR researchers, and IV researchers all collaborated to produce a variety of rough sketches, incomplete diagrams, and temporary chalkboard drawings of the compact binary search pipeline. Though inaccurate in and of themselves, these images were critical for orienting participants and enabling further discussions of the visualization activity at hand, a process known as image-enabled discourse (Snyder, 2009a, b, 2012a, b, c).

This design case details how, over the course of a participatory research and design process, creative visualization interactions were critical to the development of visualization tools for LIGO. Indeed, it demonstrates how under some circumstances these kinds of human-centered approaches to visualization are not merely desirable, but actually critical to the ultimate achievement of the visualization task.

Problem overview

The LIGO Scientific Collaboration

The Laser Interferometer Gravitational Wave Observatory (LIGO) is a National Science Foundation funded project operated by Caltech and MIT. The LIGO Scientific Collaboration (LSC) and its French-Italian partner Virgo (Virgo Collaboration, 2009) form an international scientific collaboration that uses the LIGO and Virgo observatories to detect gravitational waves from astrophysical sources. (e.g. "inspiral" waves produced when two closely neutron orbiting stars or black holes collide and merge with each other, or "stochastic" waves, which are the detectable effects of the cosmic gravitational wave background) (Barrish and Weiss, 1999; Abadie *et al.*, 2011). To date, no gravitational wave has yet been directly detected.

Laser sensor systems, called Michelson interferometers, are used to detect gravitational waves. The LIGO interferometers consist of two lasers in a vacuum, arranged perpendicular to each other in an "L" shape and pointed down very long (4 km) "arms" at corresponding photodetectors. Gravitational waves can have a measurable physical effect on these laser instruments, and data from several of these instruments, located at various locations in the USA and around the world, is typically combined for analysis (Harry *et al.*, 2004; Barrish and Weiss, 1999). Because of the extreme sensitivity of the laser interferometer instrumentation, the data produced is very noisy. For example, a member of the LIGO collaboration indicated that a train passing several miles from a detector would be enough to render instrument data from that time unusable; interference effects combined with gravitational wave signals that are faint to begin with produce a low signal/noise ratio (Barrish and Weiss, 1999).

To search for gravitational waves from inspiralling neutron stars and/or black holes, LSC and Virgo scientists use a complex computational pipeline, the subject of this research: the LSC compact binary search pipeline (Brown *et al.*, 2007; Babak *et al.*, 2012). This pipeline runs on multi-CPU/GPU server clusters (2011) to filter the large amounts of raw data available from the interferometer instruments. These computer processes include steps to "veto" bad data, look for "triggers" (signals that may be worth examining more closely), match triggers to coincident data from other detectors, and evaluate whether a trigger actually matches the signature of a gravitational wave event. The process of filtering signal from noise within LIGO data is computationally expensive because of the quantity of data and the many steps needed to process it. LIGO scientists have an interest in avoiding duplicated computational runs by producing information tools that:

- inform other collaboration members of what interferometer signals have previously been processed and how; and
- provide ways for collaboration members to select and incorporate portions of previously completed processing runs into newly established processing runs.

Information retrieval task

An information retrieval (IR) tool to search previous computational processing activities was proposed. Members of the Syracuse University Gravitational Wave Group (S.U., 2011) (a member group of the LIGO scientific collaboration looking specifically for "inspiral" gravitational wave events) therefore began working closely with IR researchers to produce such a tool.

Visualizations of the computational pipeline itself soon became an intended component of the proposed IR tool. Such visualizations were expected to be useful for several reasons. Scientists using the IR system could use the returned visualizations of previous computational runs to help make decisions about the value of those runs for their own work (use case 1). Furthermore, such visualizations could eventually be developed into interfaces for LIGO scientists to easily select portions of previous runs as a basis for future work (use case 2).

It is understood that knowledge construction and communication can be enabled by visualization tools (Fischer *et al.*, 2002), as well as sketches or drawings produced during conversation, a phenomenon sometimes called image-enabled discourse (Snyder, 2009a, b). Even rough, incomplete, or inaccurate diagrams of LIGO's complex computational processes would be expected to enable more detailed discussion, and thus the generation of progressively more detailed and accurate visualizations and a more useful IR system overall. Though not considered to be a final deliverable for this project, intermediary visualization artifacts became an important incidental use case (use case 3).

Visualization task

Use cases 1 and 2 were dependent on successfully developing visualizations for use case 3. That is, before dynamically generated visualizations depicting specific computational runs could be produced for the IR system, static, hand-crafted visualizations of the pipeline in its general form would be needed to improve communication and planning (e.g. preparation for the tool to be built). The majority of this design case addresses visualization artifacts produced for use case 3.

Challenges

Creating visualizations for case 3 was a non-trivial task. The LSC's compact binary search pipeline is highly complex. It is documented using a wiki system that relies on Gravitational Wave Group scientists' own initiative to contribute and has very little overarching structure or organizational control. The wiki web site describes pipeline functionality, but not a systematic or complete way. Most experienced Gravitational Wave Group scientists become familiar with the pipeline through their use of it over time, and the executable files of the system itself become the main form of documentation. This is problematic for new inductees to the project, including new researchers and graduate students. It was similarly problematic for the IR and IV researchers involved in planning and developing IR and visualization tools, who had no knowledge of how the pipeline worked. For the Gravitational Wave Group scientists themselves, reliance on executable files and their own knowledge can break down when the need arises to undertake infrequently performed tasks.

IR researchers, having no familiarity with LIGO or the computational pipeline, began by automatically producing a diagram of the pipeline, drawing on the many thousands of intermediary files that the pipeline executes during a run. Executables within the pipeline have multiple input and output files, and these could be visually reproduced, along with their connections to one another (see Figure 1).

In most respects, this first visualization attempt was a failure. It was massive in scope, and overwhelmingly complex. As an automatically generated diagram based on simple input/output links between executables, it was also an exercise in pure representation that excluded any kind of human interaction with the data or the visual output. The diagram was static and printed on paper. It had no affordances for representational interactions. Users could not easily change or manipulate links between nodes, seeking patterns or meaningful connections. Furthermore, though this visualization did support cognitive interaction, its complexity and scale acted as more of a barrier rather than an aid to understanding. This is to say that the diagram's massive and seemingly patternless network of links and nodes, as well as the fact that no human knowledge of the pipeline and its workings had helped to shape it, conspired to generate more confusion than clarity. It was clear from discussions with Gravitational Wave Group scientists that some underlying organizational structure had been excluded from the visualization, resulting in its misleading scale and complexity, but it was unclear as to what this structure should look like or how it could be visualized.

Method

Having failed with an automatically generated visualization, IR and IV researchers directed their attention toward human-centered visualization – that is, creative visualization interactions. By seeking rich knowledge of LIGO practices and activities and creating different visual artifacts to foster discussion, IR and IV researchers hoped to unlock the underlying structure of the compact binary search pipeline. Participant observation, a qualitative, action-oriented method of research, would be used. DeWalt and DeWalt (2002) describe this method as follows:

> Participant observation is a method in which a researcher takes part in the daily activities, interactions, and events of a group of people as one of the means of learning the explicit and tacit aspects of their life routines and their culture." This method, "draws on the insights

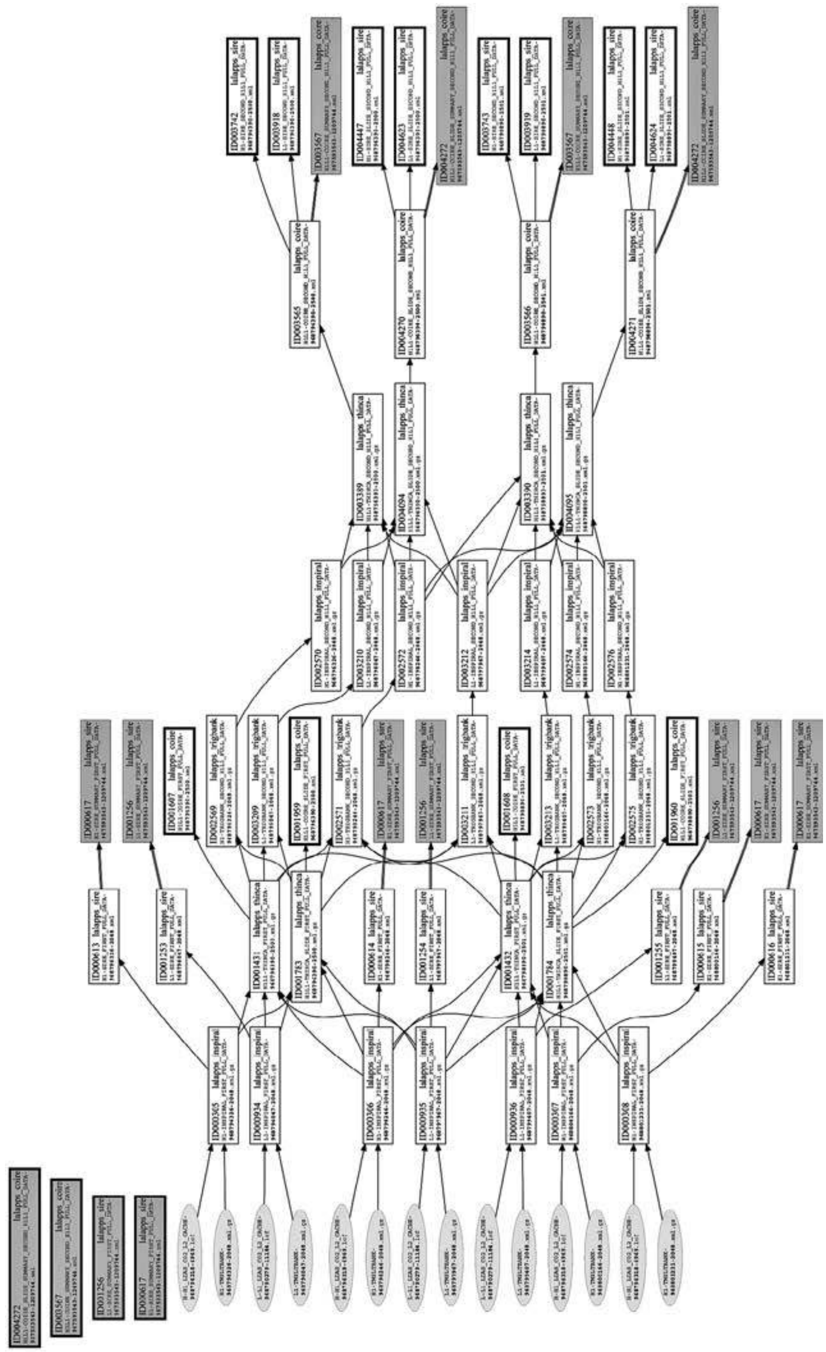

Figure 1.
The first LSC compact binary search pipeline visualization

> gained through the use of participant observation for gaining greater understanding of phenomena from the point-of-view of participants (DeWalt and DeWalt, 2002).

Active researcher participation of this kind is sometimes frowned on because of the potential to introduce researcher bias (Yin, 2009). However, active involvement can also be beneficial and preferable in some cases. According to DeWalt and DeWalt (2002):

> The participating observer seeks out opportunities to spend time with and carry out activities with members of communities in which he or she is working. Because enculturation takes place at the same time (it is hard to avoid), we believe that a tacit understanding of the experience is also being developed. It is an understanding that is not easily articulated or recorded, but that can be mobilized in subsequent analysis.

A series of ten meetings were held over the course of three months with Gravitational Wave Group members and IR/IV researchers. Gravitational Wave Group members attended six of these meetings, while IR and IV researchers were present at all ten. Meetings were typically between two and four hours in length.

Gravitational Wave Group participants included two scientists (a gravitational wave physicist and a high performance computing specialist), each of whom had high levels of experience in the use and design of the computational pipeline. In addition, three graduate students in physics participated in one meeting; their experience ranged from high levels of proficiency (a senior student who was close to defending his dissertation) to intermediate (students who had been in the physics program for a few years). IR group participants included the director of the Syracuse University Center for Natural Language Processing (CNLP), as well as a technical specialist and programmer who would ultimately be directly responsible for developing the IR system. The author of this design case, an IV researcher, was also considered to be a member of the IR group because of the close relationship between the visualization task and other IR-specific tasks.

In accordance with participant observation and action-oriented research norms, IR/IV researchers did not assume a purely observational role, but rather were partners in the project, fully engaged contributors to the various topics under discussion. They guided conversations to points where knowledge held by the Gravitational Wave Group scientists could be extracted and codified, and sought to understand LIGO activities through a collaborative design experience. For example, IR/IV researchers participated in instantiating a pipeline "run," choosing run parameters and learning from the physicists who normally undertake such work. By participating in normal pipeline operations themselves, IR/IV researchers had a frame of reference for how the pipeline system worked and what kinds of information and decisions went into a successful run.

Detailed notes in the ethnographic research tradition (Emerson *et al.*, 1995) were taken at all meetings. In addition, photographic records were made of important visual artifacts resulting from meeting discussions. These included chalkboard and whiteboard notes, sketches, diagrams, and hand-drawn notations made on previously produced visual material. Gravitational Wave Group and IR/IV researchers also exchanged regular e-mails that expanded on and clarified information discussed during meetings. These e-mails became an additional source of data.

Findings

The LSC's compact binary search pipeline: overview

Even when the decision was made to abandon the preliminary, automatically-generated visualization, progress seemed not to have improved. One reason was the use of specialized terminology by Gravitational Wave Group members. LIGO and its associated organization, instrumentation, scientific goals, and the pipeline itself are complex, so it took time for IR and IV researchers to become comfortable with a glossary of highly specialized terms: "DAX," "DAG," "Condor," "Pegasus," "veto," "cat," "event," "frame," "workflow," etc. More challenging, it soon became apparent that members of the Gravitational Wave Group defined these terms differently. No standardized glossary of terms existed, so each scientist used a mixture of commonly-understood and personalized terminology. This led to many disagreements and discussions about specific meanings for a host of different terms.

Members with high levels of experience held relatively accurate views of the pipeline, while newer group members had less accurate and sometimes underdeveloped views of the pipeline. One PhD student at an intermediate experience level described the pipeline as a "black box" and described previous visualization efforts that emphasized only the computational aspects of the pipeline as "next to useless." Yet thoroughly understanding the pipeline and its various components and relationships was important to this student: "Knowing what this [the pipeline] is doing... that is the biggest thing." An advanced doctoral student further indicated that, "Intermediate data products are indecipherable until you know the whole system." Similarly, IR and IV researchers needed terminology to be semantically defined in detail, but also depicted visually in relationship to other terms and components of the pipeline. No such visualization existed, so after four meetings, IR and IV researchers attempted to make one on their own, producing a whiteboard diagram over a two-hour discussion session (see Figure 2).

Despite the risks of attempting a visualization without Gravitational Wave Group members in attendance, this intermediary visualization artifact was a moment of clarity for the IR and IV researchers who had produced it. Since it had been produced by members of the collaboration for themselves (rather than automatically generated or provided by outsiders), it was simpler and clearer than previous automated attempts, uncluttered by masses of irrelevant data. It was also a representation of what IR and IV researchers thought they knew, rather than a representation of whatever was possible to produce from pipeline data alone. This distinction turned out to be invaluable.

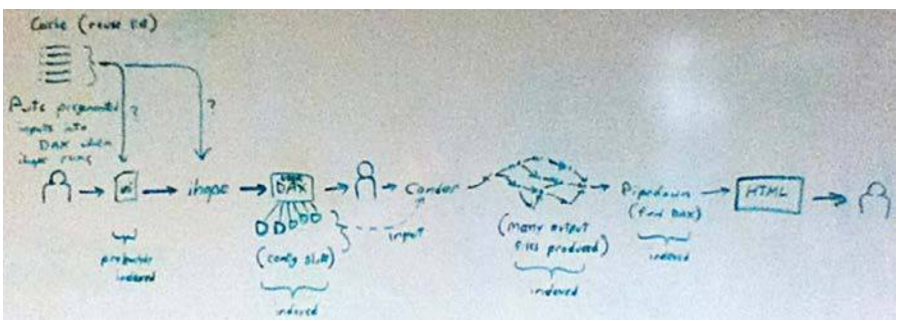

Figure 2.
Early LSC compact binary search pipeline overview (whiteboard)

Nonetheless, the visualization highlighted key doubts held by the IR and IV researchers, as indicated by quotes from the meeting where it was produced: "I think that is what they were saying," and, "Does this [diagram] seem to resemble what we have been talking about?" Some nodes of the pipeline structure were highly uncertain. For example the processes "Condor" and "Pipedown" (Couvares *et al.*, 2007; Tannenbaum *et al.*, 2002) were known to be inaccurately represented, but were depicted as they were because of the need to create some foundational visual artifact for discussion.

An identical (albeit tidier) version of this diagram was prepared digitally and printed for discussion in a follow-up meeting with Gravitational Wave Group members. This "discussion version" of the diagram had an immediate effect. Because the diagram was not a depiction of truth or data, but merely an attempt to make sense of many prior discussions, collaborators felt free to mark and correct it at will. IR and IV researchers knew the diagram was inaccurate, but hoped that Gravitational Wave Group members could identify those specific inaccuracies and correct them. This turned out to be the case, as the collaborators held a visually-enabled discussion, sometimes referred to as image-enabled discourse (Snyder, 2009a, b, 2012a, b, c). Some aspects of the diagram were certified as accurate, but many others were revised (see Figure 3).

The overview visualization and subsequent discussion established that the processes called "Pegasus," "DAGman," and "Condor," (Couvares *et al.*, 2007; Tannenbaum *et al.*, 2002) interacted in the pipeline to parse through an XML formatted hierarchy of interrelated computational jobs (a "DAX"), establish job dependencies,

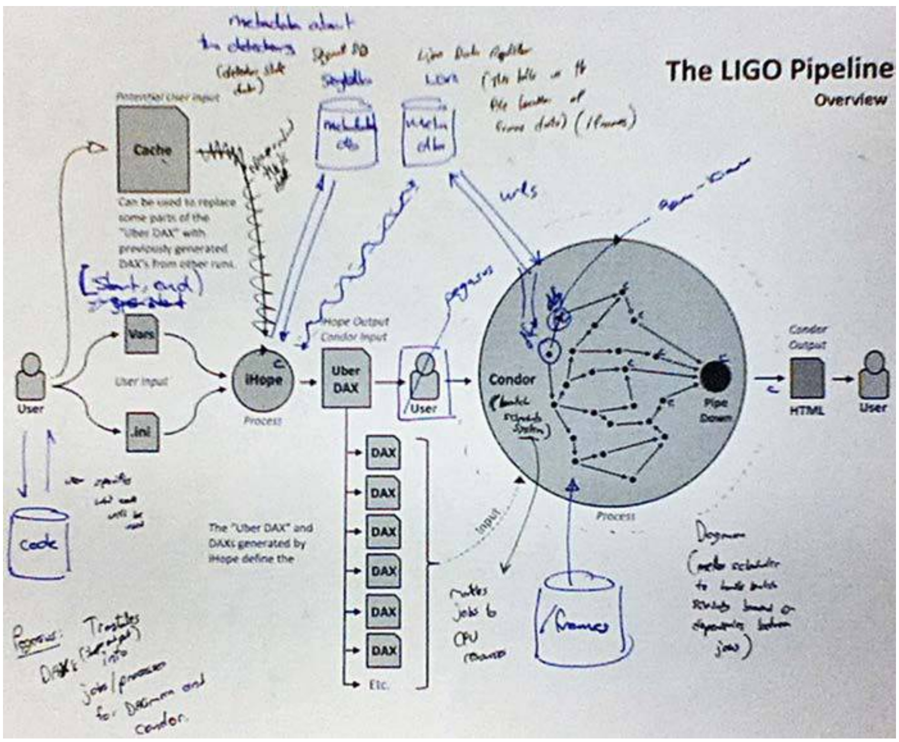

Figure 3.
LSC compact binary search pipeline overview (discussion version)

and feed those jobs to processing resources. This highly complex series of events was made more so by the addition of antecedent processes such as "iHope" that form the DAX based either on user input or the "cache" where users can substitute parts of previous runs into the pipeline data to avoid unnecessary duplication of computational effort.

A major flow of the automatically produced diagram was that it could not show variations that occur between different pipeline runs. Each run produces data only about itself, and the automated visualization was produced out of data from just one run. In contrast, the discussion diagram was a holistic view produced from human understanding of the pipeline. It showed the activities surrounding the pipeline, including human activities that are not computational at all. It also showed various points where modifications or data insertions might be made by Gravitational Wave Group members. Though it took many hours of hard work, discussion, and correction to achieve, the visualization produced through creative visualization interactions among several collaborators was a more useful tool by far than the automatically generated visualization. A senior member of the Gravitational Wave Group described the overview visualization as, "probably the clearest picture we've ever had of this process" (see Figure 4)

The pipeline overview was a necessary, but not sufficient, first step toward producing visualization-enabled IR tools that could accommodate the typical functionalities of overview, filter, and detail (Shneiderman, 1996). Since all members of the collaboration now had a reasonable understanding of the basic pipeline structure, conversation turned to the details of the "DAGman" and "Condor" processes. In the overview diagrams, this portion of the pipeline was initially represented by a series of nodes and arrows depicting the flow of data through a series of interdependent computational jobs. It was not yet clear how this specific structure was organized.

DAGman and Condor

The process for unpacking "DAGman" and "Condor" (Couvares *et al.*, 2007; Tannenbaum *et al.*, 2002) was similar to that for creating the overview visualization. A series of discussions led to informal visualization attempts (usually during meetings), followed by collaborative corrections and a final, refined diagram. Perhaps the most critical piece of information discerned in discussions about "DAGman" and "Condor" was the role of time in the interferometer data itself.

The LIGO interferometers, when turned on, provide a steady stream of data, similar in some respects to an audio microphone. Analyzing this data requires that scientists define a past timeframe to analyze, be it just a few seconds, several minutes, or even hours at a time. The interferometer data is subdivided into frames, consisting of 16 seconds each, and these frames constitute the minimum actionable unit of data that is addressed by jobs within the pipeline. Members of the Gravitational Wave Group eventually codified this via a chalkboard visualization showing how the interferometer data became partitioned into frames. This visualization was produced over the course of a two-hour collaborative discussion (see Figure 5).

This drawing led to the key finding for this design problem: that the LSC compact binary search pipeline analyzes many frames of data simultaneously, and these frames are all subjected to the same relatively simple computational process. A second

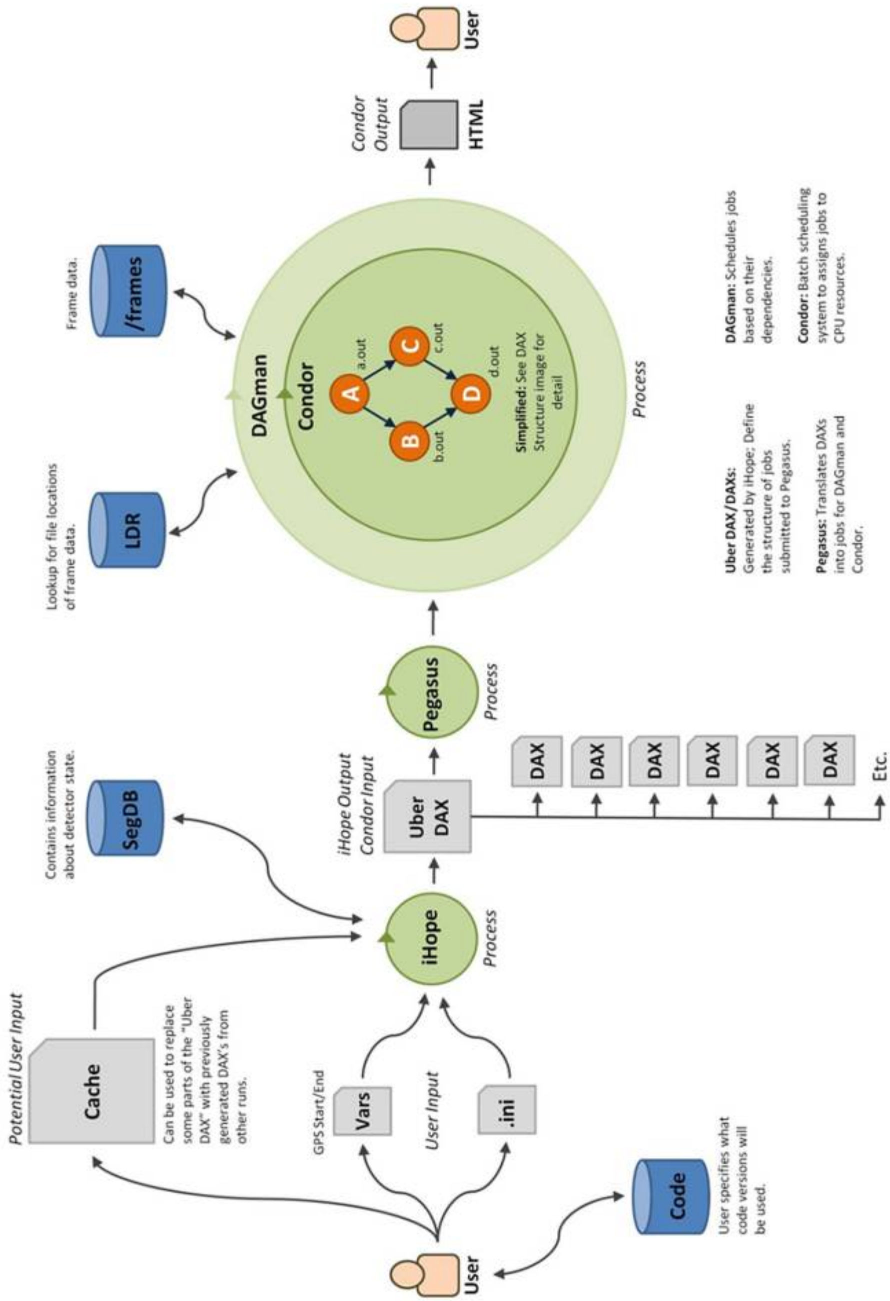

Figure 4.
The final LSC compact binary search pipeline overview visualization

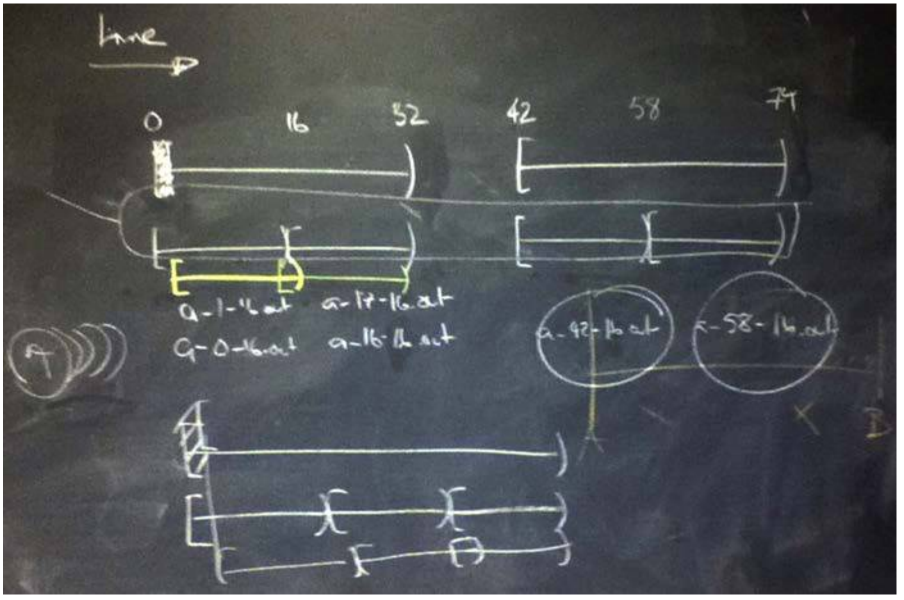

Figure 5.
Interferometer data represented as 16 second frames

chalkboard diagram showed how each frame of data would be submitted to a computational node (labeled as "A"). From there, output data would be submitted to two new nodes ("B" and "C"), and their output then combined by a fourth node ("D") (see Figure 6).

Again, a creative interaction was necessary to make this organizing structure fully clear. One of the Gravitational Wave Group members described the "fatal flaw" of the initial, automatically-generated visualization: "the graphing code can't figure out that there are a bunch of nodes that are really the same job repeated *n* times with different GPS start and end times." In this instance, human input to the visualization activity was necessary to reveal both the problem and its solution.

Interferometer frames were one important dimension of visualizing the pipeline. A second key dimension was the number of interferometers that are included in a run. There are several interferometer instruments located around the world, including instruments that are no longer operational but have archived data sets that are still useful. LIGO scientists can choose to include data from one or several of these instruments in a given run (typical runs use more than one data source to more easily filter signal from noise). The pipeline subjects each frame of data from each interferometer to the same series of computational transformations, so the number of interferometers examined also affects how many of each type of job will occur.

The third key dimension for visualization was known from the start: processing time. Frames of data progress through the pipeline in a relatively linear fashion, with the output of one job acting as an input to others. The time it takes for frames of interferometer data to progress through the pipeline is dependent on the power of the computational hardware that the pipeline is running on. Jobs that occur later in the pipeline are generally not able to be started until previous jobs have completed.

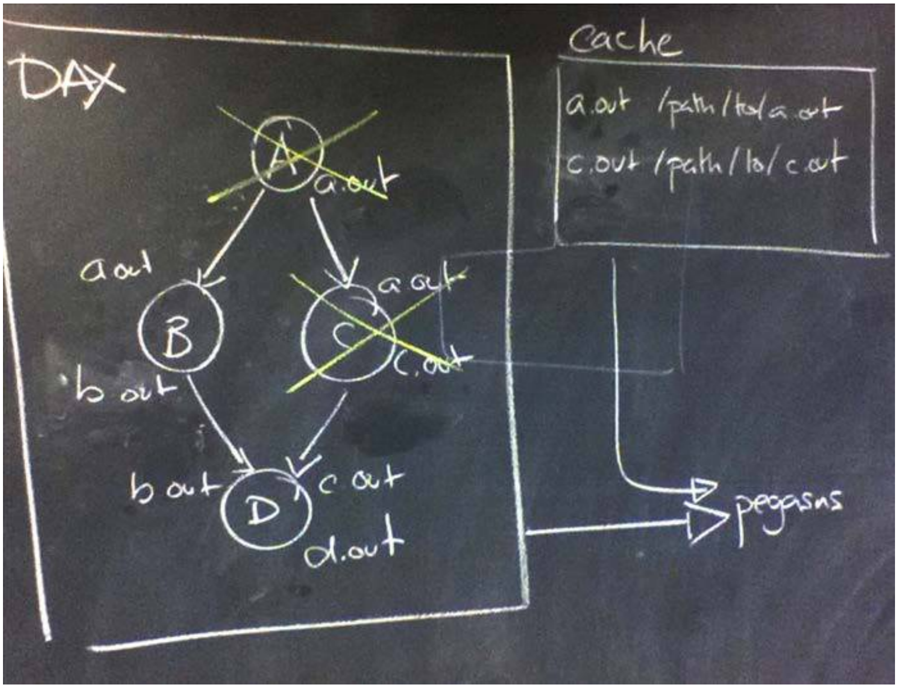

Figure 6.
DAGman and Condor in chalk, used to enable discussion

With knowledge of these three dimensions for visualization – frames, number of instruments, and processing time – IV researchers manually produced a new 3D diagram showing an abstracted view of the pipeline structure (see Figure 7).

This new diagram captured the essence of the pipeline structure, though it encompassed only part of the full process and lacked accuracy in several respects. Over the course of the remaining discussions it became apparent that visualizing the frame dimension of the pipeline (e.g. its 3D "depth"), while useful as a conceptual aid for those unfamiliar to the pipeline, was unimportant to project scientists in their real-world activities. A finished visualization tool for LIGO could collapse this dimension, the LIGO scientists being better served by numerical representations indicating how many frames had been processed rather than visualized depth.

The final visualization produced for the LIGO design case was a version generated automatically from real data, this time incorporating the knowledge gathered over the course of many round of creative visualization activity: the frame, instrument, and time-to-process dimensions, the pipeline structure, and the final use scenarios for the visualization itself (see Figure 8).

Discussion

What do groups of people gain by visualizing information for themselves? The LIGO design case reveals a number of interesting possibilities that are likely to be generalizable to other visualization problems.

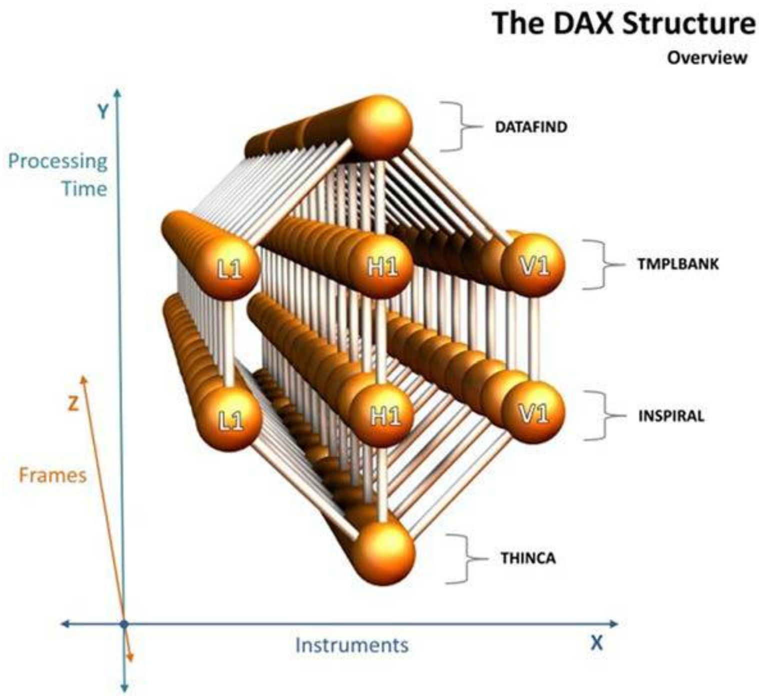

Figure 7.
The pipeline in three dimensions: frame, instrument, and processing time

Cooperative sensemaking

Many prominent IV researchers (e.g. Spence, 2001, 2007; Ware, 2004; Chen, 2003, 2006) emphasize the benefits of visualizations for human cognition. The basis for most cognitive theories of IV is simple: the human mind is not very good at holding many objects in short term memory all at once. IV tools are an aid to human memory, visually storing objects for us and arranging them in a fashion that will improve our ability to understand and draw conclusions about them (Card *et al.*, 1999). Spence (2001) describes IV as the formation of a mental model through cognitive interaction with a visual display of data. Mental models help users to form a, "better understanding of the artifact, scheme, or situation to which the data refers, and to be able to interpret that model in some useful way, perhaps to make a decision," (Spence, 2001). Ware (2004) emphasizes the perceptual aspects of cognition, describing many of the ways human beings see and interpret different kinds of visual data. The notion of image-enabled discourse (Snyder, 2009a, b) is related to this, conceptualizing, as it does, visualization artifacts as communication tools (Fischer *et al.*, 2002).

These benefits certainly applied to the visualizations produced by IR and IV researchers. These individuals came to the collaboration with virtually no knowledge of LIGO, the computational pipeline, sub-processes within the pipeline, or any other aspect of the gravitational wave research being undertaken. Gravitational Wave Group members and non-physicist scholars found themselves talking at cross-purposes, mixing vocabulary, and misunderstand fundamental terminology and concepts as they strove to understand core elements of the problem at hand. Four meetings passed with

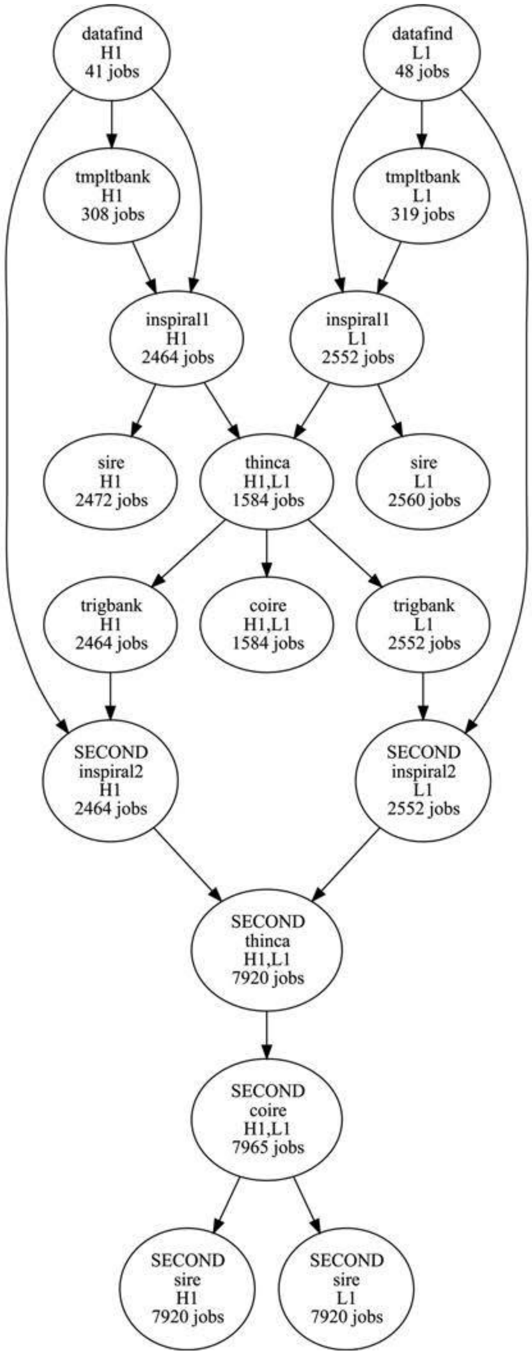

Figure 8.
The computational pipeline

minimal progress made; IR and IV researchers could not see beyond the massively complex and disorganized early visualization of the pipeline, while Gravitational Wave Group members seemed to be suggesting that in reality, the pipeline was far more structured and simple than the diagram implied.

By producing a series of visualizations for themselves, physicists, IR, and IV researchers managed to bring sense to something so complex as to seem senseless. Intermediary and even inaccurate visualization artifacts served an important purpose, anchoring discussion and directing it toward the improvement of the diagram. This is a key benefit of intermediary visualization artifacts and of cooperatively creating visualizations: the visualization and the act of creating it are mechanisms to explore the knowns and unknowns of the underlying information. The act of creation requires the different individuals involved to cognitively interact with data, filtering and manipulating it, contributing to it, exploring it, and ultimately representing it in a visual form that makes sense to everybody involved.

Codifying tacit knowledge

In the LIGO design case, as with many visualization activities, data manifested in two forms. Some data was concretely represented and accessible, such as pipeline input and output files. Other data was known, often in partial form, only by certain individuals. This data was not concretely represented or accessible to others, even in existing project documentation. It proved helpful to conceptualize these two types of data through the constructs of explicit and tacit knowledge (Collins, 2007; Polanyi, 1966; Miller, 2008; Nonaka, 2007; Nonaka and Takeuchi, 1995).

Explicit knowledge is codified knowledge, information that has been recorded or can be observed and explained in some concrete fashion (for example, mathematical formulae or a written record of a person's observable daily routine). Tacit knowledge is personal and specialized. It is un-codified knowledge that resides only within the individual and may be difficult to translate into explicit form. Because of its personal nature, tacit forms of knowledge can only be accessed by human beings (Collins, 2007; Polanyi, 1966; Miller, 2008; Nonaka, 2007; Nonaka and Takeuchi, 1995). For example, Collins (Collins, 2007) describes riding a bicycle as an expression of tacit over explicit knowledge: it is possible to read and understand codified explanations of the physics and mechanics of riding a bicycle, but this is not the same as knowing how to ride a bicycle. The physics and mechanics are explicit knowledge, while the understanding of how to ride is tacit knowledge.

In visualizing the pipeline, matters of explicit vs tacit knowledge were highly important. Computational artifacts (input, output, and executable files) were explicitly documented by the pipeline as it ran. These artifacts contained a great deal of information about the pipeline itself: file names reflected important details about relationships to or dependencies on other files, as well as information about which "jobs" (executable code) the various input and output files were associated with. However, early attempts to visualize the pipeline using only explicit, codified information showed that deeper knowledge of the pipeline would be required to produce anything meaningful. IR and IV researchers knew from early conversations with Gravitational Wave group physicists that the pipeline had an underlying structure that was somehow not being reflected through use of explicit, data-oriented knowledge alone. Visualizing this underlying structure required that IR and IV researchers gain access to the tacit knowledge of it held by the Gravitational Wave Group physicists. This tacit knowledge included details of the pipeline that were not, but could have been, made explicit. It also included knowledge of how to execute a computational run and how to read, interpret, and use results in the overarching context of the LIGO organization. A model for using visualizations to connect tacitly held knowledge to explicitly held knowledge has been described by (Jeong *et al.*, 2008), though this model assumes a conversion of explicit knowledge to tacit knowledge. In the LIGO visualization task, this flow was reversed; the design process extracted tacitly held knowledge from a variety of collaborators and structurally applied it to explicit knowledge sources through visualization.

Intermediary artifacts (the pipeline overview diagrams; the chalk sketches of "DAGman" and "Condor") allowed all members of the collaboration to begin explicitly codifying their interpretation of information under discussion. Even when these visual interpretations were incorrect, the fact that they were codified made it far easier for others with different – possibly more accurate – interpretations to weigh in and discuss corrections. This visually-enabled process collaboratively improved all participants' tacitly held knowledge and simultaneously resulted in an accurate and explicitly codified visual representation of that knowledge.

Beyond data, technology, and technique

In the current IV literature, much has been made of data, technology, and technique as determiners of a visualization's success or failure. Topics of interest include how different kinds of data or the scale of a data set (e.g. Keim, 2001; Fekete and Plaisant, 2002; Aigner *et al.*, 2007; Shneiderman, 1996) require different approaches to visualization. 3D technologies (Robertson *et al.*, 1991, 1993), graphics processing hardware (e.g. Rhyne, 2000), and mobile platforms (e.g. Chittaro, 2006) are also of interest. Finally, many visualization scholars have explored different approaches to visualization, including "Fisheye Views" (1986), virtual worlds (1990), 3D rooms (1991), "Cone Trees" (Robertson *et al.*, 1991), and many more.

The LIGO design case reveals how this tripartite scholarly emphasis on data, technology, and technique frequently overlooks an important fourth dimension: human-centeredness and the very act of creating visualizations. In this design case, neither data nor technology proved very helpful by themselves in achieving a meaningful and useful visualization; indeed, in some respects these were an obstacle to progress, as massive amounts of disorganized data and automated visualization techniques added to the confusion of the collaborators. In the same way, specific visualization techniques were not of particular interest. A variety of techniques and media were used: whiteboard, chalkboard, paper + pen, 3D, and 2D. The selection of these was made based on convenience and availability and was not, *per se*, the driving force behind the visualizations that were produced.

On the other hand, the direct involvement of many different collaborators in the creation of visualizations for themselves was a critical and important element. Only by casting aside automated shortcuts to deeply immerse themselves in the Gravitational Wave Group's activities, difficulties, and objectives, were IR and IV researchers able to begin composing visualizations with any real utility. Ultimately, it was enhanced human understanding of the pipeline and the objectives of the Gravitational Wave Group itself that led IR and IV researchers back to automatic visualization, this time with highly successful results.

Conclusion

Creative interactions with visualizations can have many benefits and advantages. In this design case, collaborators worked together to build several visualizations of a complex computational pipeline. Their process helped to transform tacitly held knowledge into actionable forms of explicit, codified knowledge. Without this transformation process, generating meaningful visualizations of the LSC compact binary search pipeline would not have been possible. As such, the visualizations generated over the course of this study served as both completed interim products of research and as tools to foster image-enabled discourse for future tool planning and development.

References

Abadie, J., Owen, B., Reitze, D., Lazzarini, A., Kondrashov, V., Stuver, A., Brady, P., Robinson, M., Daw, E., Ugolini, D., Cavaglia, M. and Ingram, D. (2011), LIGO Scientific Collaboration Web site.

Aigner, W., Miksch, S., Müller, W., Schumann, H. and Tominski, C. (2007), "Visualizing time-oriented data – a systematic view", *Computers & Graphics*, Vol. 31, pp. 401-409.

Babak, S., Biswas, R., Brady, P.R., Brown, D.A., Cannon, K., Capano, C.D., Clayton, J.H., Cokelaer, T., Creighton, J.D.E., Dent, T., Dietz, A., Fairhurst, S., Fotopoulos, N., Gonz'aleZ, G., Hanna, C., Harry, I.W., Jones, G., Keppel, D., McKechan, D.J.A., Pekowsky, L., Privitera, S., Robinson, C., Rodriguez, A.C., Sathyaprakash, B.S., Sengupta, A.S., Vallisneri, M., Vaulin, R. and Weinstein, A.J. (2012), "Searching for gravitational waves from binary coalescence", *Physical Review*, Vol. 87 No. 2, pp. 24-33.

Barrish, B. and Weiss, R. (1999), "LIGO and the Detection of Gravitational Waves", *Physics Today*, Vol. 54, October, pp. 44-50.

Brown, D.A., Brady, P.R., Dietz, A., Cao, J., Johnson, B. and McNabb, J. (2007), "A case study on the use of workflow technologies for scientific analysis: gravitational wave data analysis", *Workflows for e-Science*, Springer, New York, NY, pp. 39-59.

Card, S.K., MacKinlay, J.D. and Shneiderman, B. (1999), *Readings in Information Visualization: Using Vision to Think*, Morgan-Kaufmann, San Francisco, CA.

Card, S.K., Robertson, G.G. and MacKinlay, J.D. (1991), "The information visualizer, an information workspace", *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology. New Orleans, Louisiana, United States*, ACM, New York, NY.

Chen, C. (2003), *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*, Springer-Verlag, New York, NY.

Chen, C. (2005), "Top 10 unsolved information visualization problems", *IEEE Computer Graphics and Applications*, Vol. 25 No. 4, pp. 12-16.

Chen, C. (2006), *Information Visualization: Beyond the Horizon*, Springer-Verlag, London.

Chen, C. and Czerwinski, M. (2000), "Empirical evaluation of information visualizations: an introduction", *International Journal of Human-Computer Studies*, Vol. 53 No. 5, pp. 631-635.

Chittaro, L. (2006), "Visualizing information on mobile devices", *Computer*, Vol. 39 No. 3, pp. 40-45.

Collaboration (2009), *Collaboration Advanced Virgo Baseline Design*, VIR A 27.

Collins, H. (2007), "Bicycling on the moon: collective tacit knowledge and somatic-limit tacit knowledge", *Organization Studies*, Vol. 28 No. 2, pp. 257-262.

Couvares, P., Kosar, T., Roy, A., Weber, J. and Wenger, K. (Eds) (2007), *Workflow in Condor*, Springer Press, New York, NY.

Dewalt, K.M. and Dewalt, B.R. (2002), *Participant Observation: A Guide for Fieldworkers*, AltaMira Press, Walnut Creek, CA.

Ellis, G. and Dix, A. (2006), "An explorative analysis of user evaluation studies in information visualisation", *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, Venice, Italy*, ACM, New York, NY.

Emerson, R.M., Fretz, R.I. and Shaw, L.L. (1995), *Writing Ethnographic Field Notes*, The University of Chicago Press, Chicago, IL.

Feiner, S.K. and Clifford, B. (1990), "Worlds within worlds: metaphors for exploring n-dimensional virtual worlds", *Proceedings of the 3rd Annual ACM SIGGRAPH Symposium on User Interface Software and Technology, Snowbird, Utah, United States*, ACM, New York, NY.

Fekete, J.-D. and Plaisant, C. (2002), "Interactive information visualization of a million items", *INFOVIS 2002. IEEE Symposium on Information Visualization INFOVIS 2002. IEEE Symposium on Information Visualization INFOVIS 2002 Proceedings*.

Fischer, F., Bruhn, J., Gräsel, C. and Mandl, H. (2002), "Fostering collaborative knowledge construction with visualization tools", *Learning and Instruction*, Vol. 12 No. 2, pp. 213-232.

Furnas, G.W. (1986), "Generalized fisheye views", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, Massachusetts, United States*, ACM, New York, NY.

Harry, G. *et al.*, (2004), "Detector description and performance for the first coincidence observations between LIGO and GEO", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. 517, pp. 154-179.

Jeong, D.H., Chang, R. and Ribarsky, W. (2008), "An alternative definition and model for knowledge visualization", paper presented at IEEE Visualization 2008 Workshop on Knowledge Assisted Visualization.

Keim, D.A. (2001), "Visual exploration of large data sets", *Commun. ACM*, Vol. 44 No. 8, pp. 38-44.

MacKinlay, J.D., Robertson, G.G. and Card, S.K. (1991), "The perspective wall: detail and context smoothly integrated", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching through Technology, New Orleans, Louisiana, United States*, ACM, New York, NY.

Miller, K.D. (2008), "Simon and Polanyi on rationality and knowledge", *Organization Studies*, Vol. 29 No. 7, pp. 933-955.

Nonaka, I. (2007), "The knowledge-creating company", *Harvard Business Review*, No. 85, pp. 162-171.

Nonaka, I. and Takeuchi, H. (1995), *The Knowledge Creating Company*, Oxford University Press, New York, NY.

Polanyi, M. (1966), *The Tacit Dimension*, Anchor Day Books, New York, NY.

Rhyne, T.M. (2000), "Computer games' influence on scientific and information visualization", *Computer*, Vol. 33 No. 12, pp. 154-159.

Robertson, G.G., Card, S.K. and MacKinlay, J.D. (1993), "Information visualization using 3D interactive animation", *Commun. ACM*, Vol. 36 No. 4, pp. 57-71.

Robertson, G.G., MacKinlay, J.D. and Card, S.K. (1991), "Cone Trees: animated 3D visualizations of hierarchical information", *Proceedings of the SIGCHI Conference on Human Factors in*

Computing Systems: Reaching through Technology, New Orleans, Louisiana, United States, ACM, New York, NY.

S.U. (2011), Gravitational Wave Group.

Shneiderman, B. (1996), "The eyes have it: a task by data type taxonomy for information visualizations", IEEE Symposium on Visual Languages, Boulder, CO.

Snyder, J. (2009a), "Image-enabled discourse: a preliminary descriptive investigation", iConference, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Snyder, J. (2009b), "Investigating the creation of ad hoc information visualizations", *Proceedings of the American Society for Information Science and Technology*, Vol. 46 No. 1, pp. 1-7.

Snyder, J. (2012a), "Image-enabled discourse: investigating the creation of visual information as communicative practice", School of Information Studies, Syracuse University, Syracuse.

Snyder, J. (2012b), "Let me draw you a picture: coordination in image-enabled conversation", *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion, Seattle, Washington, USA*, ACM, New York, NY.

Snyder, J. (2012c), "Preliminary findings: image-enabled discourse and the creation of visual information", *Proceedings of the American Society for Information Science and Technology*, Vol. 48 No. 1, pp. 1-4.

Spence, R. (2001), *Information Visualization*, ACM Press, Harlow.

Spence, R. (2007), *Information Visualization: Design for Interaction*, ACM Press, Harlow.

Tannenbaum, T., Wright, D., Miller, K. and Livny, M. (Eds) (2002), *Condor – A Distributed Job Scheduler*, The MIT Press, Cambridge, MA.

Thomas, J.J. and Cook, K.A. (2005), *Illuminating the Path*, IEEE, Los Alamitos, CA.

Tory, M. and Möller, T. (2004), "Human factors in visualization research", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 10 No. 1, pp. 72-84.

Ware, C. (2004), *Information Visualization: Perception for Design*, Morgan Kaufmann, San Francisco, CA.

Yi, J.S., Kang, Y.A., Stasko, J.T. and Jacko, J.A. (2007), "Toward a deeper understanding of the role of interaction in information visualization", *Visualization and Computer Graphics, IEEE Transactions on*, Vol. 13 No. 6, pp. 1224-1231.

Yin, R.K. (2009), *Case Study Research: Design and Methods*, Sage Publications, Thousand Oaks, CA.

About the author
Nathan Prestopnik is currently involved in research which explores the relationship between design and scholarship within the HCI discipline. As part of this work, Nathan has designed several interactive information systems to study key issues in HCI, including visualization tools to support history education and purposeful games to engage the general public in crowdsourced science. Before beginning his doctoral study, Nathan was a web developer, graphic designer, and UX designer in Los Angeles, CA, Clifton Park, NY, and Syracuse, NY. He holds bachelor degrees in Television-Radio-Film Production and US History, as well as an MS in New Media from Syracuse University. Nathan Prestopnik can be contacted at: napresto@syr.edu

This article has been cited by:

1. Mu-Yen Chen, Edwin David Lughofer. 2013. Smart space technology innovations. *Library Hi Tech* **31**:2, 197-200. [Abstract] [Full Text] [PDF]