Scientific Storytelling using Visualization

Kwan-Liu Ma, University of California, Davis, CA, USA
Isaac Liao, University of California, Davis, CA, USA
Jennifer Frazier, The Exploratorium, San Francisco, CA, USA
Helwig Hauser, University of Bergen, Norway
Helen-Nicole Kostis, NASA Goddard Space Flight Center, Greenbelt, MD, USA

"I like a good story well told. That is the reason I am sometimes forced to tell them myself."
—Mark Twain

What springs to mind when you hear the word "storytelling"? For most of us, it conjures up images of children gathered in front of a rocking chair, rapt with attention as an elder narrates a fairy tale to them. Unencumbered by the inhibitions of older years, they are not afraid to interrupt and ask for details to satisfy their curiosity, or clamor for more when the story ends. How can we, as visualization researchers and practitioners, elicit this same engagement and wonder in our viewers? How can we aid Mr. Twain in his plight, and ensure that he is not the exclusive purveyor of good stories well told?

All stories are sequences of causally-related events. However, the *good* ones tend to share several important features. First, they take time to unfold, and their pacing matches the audience's ability to follow them. Second, they hold the audience's attention by having interesting settings, plots, and characters. Finally, they leave a lasting impression, either by piquing the audience's curiosity and making them want to learn more, or by conveying a deeper meaning than your normal everyday run-of-the-mill sequence of causally-related events.

When using visualization to tell a story, what does it mean for a visualization to have "good pacing"? A story that is well paced exhibits deliberate control over the rate at which plot points occur. However, any given pace may feel too fast or too slow to different audiences, depending on their attention spans and personal preferences. Similarly, in designing visualizations, it is crucial to gauge the intended viewers' familiarity with both the subject matter and with visualization conventions. For a given dataset, distributing data across multiple line charts might prove to be the most suitable approach for a general audience, but domain experts might prefer to combine data into a single parallel coordinates diagram to facilitate comparison.

What are the settings, characters, and plots of visualizations? First, the setting of a visualization is all of the background information a viewer needs to know in order to contextualize and comprehend the visualization. In theatrical productions, the stage is generally set before the curtains rise; similarly, viewers should be introduced to the subject matter before seeing a visualization of it. In addition, visual elements representing data points are the characters and centers of attention in visualizations – they're the stars of the show. Finally, the plot, and the dramatic tension, of a visualization arises from the juxtaposition of its visual elements, how they interact and compare with one other, and how they evolve over time.

Armed with these notions, let us discuss how visualization can be used to tell a good story, and tell it well. In particular, we emphasize *scientific storytelling* – that is, telling stories using scientific data – which is a topic that the visualization research community has paid little attention to so far. In contrast, the subject of storytelling in *information visualization* has been the topic of several recent workshops and panels, and provides a starting point for the discussion of scientific storytelling.

Storytelling in Information Visualization

At VisWeek 2010 in Salt Lake City, there was a day-long workshop called "Telling Stories with Data: Using visualization to create narratives and engage audiences." Hosted by Matt McKeon, Joan DiMicco, and Karrie Karahalios, this workshop featured a diverse range of speakers, including journalists, bloggers, literary analysts, and developers of information visualization software. Throughout the day, we saw numerous examples of how stories are told with data: a casualty map ("Home and Away") from CNN.com; a political blogger (Matthias Shapiro, aka "10000Pennies") using pennies to explain that a budget cut of $100 million, while sounding impressive, is actually a tiny fraction of the deficit; a collaborative visualization website called CommentSpace that allows users to create, share, and comment on views of datasets. Visualization creators shared their goals and design decisions, and breakout sessions allowed for discussion in smaller groups.

Two examples of scientific storytelling. **Top:** Science on a Sphere, a presentation to Queen Elizabeth II and Prince Philip at NASA Goddard Space Flight Center. **Bottom:** Visitors at the Exploratorium, a science museum in San Francisco, examining oceanographic data projected onto a topographical relief map.

Several interesting points emerged over the course of the day. There was a general consensus that framing data as a narrative makes it more interesting and memorable. Why might this be? Cognitive science postulates the existence of separate types of memory for storing different types of information: semantic memory, for remembering disconnected facts, and episodic memory, for remembering sequences of events. By presenting themselves as narratives, visualizations can tap into episodic memory and establish themselves as cohesive entities. In addition, the issue of interactivity in visualizations came up repeatedly. The style of storytelling present in static visualizations, such as info-graphics, is fundamentally different from that of interactive visualizations, in which users are allowed to navigate and modify views of data. Making a visualization more interactive gives users more freedom to explore, but lessens the amount of control visualization designers have over how the story is told. In the end, we concluded that the interactivity of a visualization should be carefully balanced against the need to guide the viewer through the data. A useful compromise might be to start the visualization in a non-interactive mode, ensuring that the most salient features of the dataset are presented, and then allow users to explore the rest of the dataset afterwards.

The visualizations presented and discussed in this workshop fell squarely in the domain of *information visualization*, which tends to use more abstract representations and are usually targeted towards more general audiences. By contrast, what challenges does *scientific visualization* face in storytelling?

Scientific Storytelling

Visualization has become an important tool for scientists in their daily work. Scientists create visualizations for various purposes: to validate experiments, to explore datasets, or to communicate findings to others. If appropriately presented, such visualizations can be highly effective in conveying narratives. Thus, using the above criteria, let us explore the possibility of telling stories using scientific visualizations.

The narrative impact in information visualization stems from visual comparisons using simple, abstract representations of data: bar charts show differences in length, scatterplots show differences in position, treemaps and pie charts show differences in area, and heatmaps show differences in color and intensity. As such, information visualization stories are about comparison or change: "Look at how much bigger A is than B," or "look at how C has grown over time." In contrast, much of the narrative impact in scientific visualization comes from being able to see real data that is normally invisible. At its best, scientific visualization acts as an extension of our senses, allowing us to perceive and manipulate data at otherwise impossible scales and perspectives, such as vector fields in weather systems, isosurfaces in supernova simulations, and layers of human anatomy rendered semitransparently. Whereas information visualizations are allegories – abstractions and summaries of raw data – scientific visualizations are more literal; they strive for realism and spatial accuracy, sacrificing details only to facilitate understanding.

In some ways, scientific visualization has it easy. Usually, the intended viewers are the scientists who generated the data, and others in the same field. Thus, they need very little introduction – in terms of our storytelling metaphor, they're already familiar with the setting, and all that is left is to identify the characters (for instance, what glyphs represent, and how color is used). In fact, when we design scientific visualizations, the scientists are usually the ones setting the stage for us! Additionally, the fact that the data are already highly relevant to them increases the likelihood that visualizations will leave a lasting impression in their minds.

However, difficulties arise when scientific visualizations are introduced to broader audiences. Even the best visualizations are incomprehensible if their concepts are alien, and scientific visualizations are often designed assuming viewer familiarity with the subject matter. Moreover, time constraints and limited attention spans often preclude the possibility of full explanations. How can we address these issues?

In 2010, a one-day workshop on scientific storytelling was held at the University of California, Davis. Participants included visualization researchers and practitioners as well as experts in animation, scientific journalism, and science museum exhibition. The rest of this article presents highlights and findings from this workshop.

Production Visualization at a Scientific Research Center

The use of visualizations to tell scientific stories is a routine practice at the U.S. National Aeronautics and Space Administration (NASA). Observational data – that is, data that can be recorded by instruments and sensors – are continuously collected, archived, and processed from NASA airborne missions and experiments. As of 2011, there are 64 airborne missions operating within the Science Mission Directorate of NASA (Earth: 19, Heliophysics: 16, Astrophysics: 15, and Planetary: 14) [1]. Each of these missions is usually equipped with multiple sensors and instruments, whose purpose is to acquire and transmit data sets daily, hourly, or even every few minutes. Data acquisition is an ongoing process, and lasts for the duration of the mission. The majority of airborne missions are operational for more than a year, and some can be operational for more than a decade (e.g., LandSat satellites).

NASA scientists, who are sometimes the principal investigators of missions, need to process and visualize data acquired from airborne science missions to advance their research and to support outbound communication and scholarly work, such as by publishing in scientific journals. Data visualization is also needed for education and public outreach activities, to engage and educate the public about NASA's research and science efforts. Scientists and mission teams have their own tools to process and analyze data, but cannot easily develop and produce high-quality visualizations for the following reasons: 1) complexity and volume of data; 2) complexity of the tools and technology required to perform high-quality visualization production; 3) lack of expertise in the fields of art, visualization, and storytelling production.

The Scientific Visualization Studio (SVS) [2] at NASA Goddard Space Flight Center (GSFC) facilitates scientific inquiry and outreach within NASA programs through visualization. The SVS works closely with scientists in the creation of visualization products, systems, and processes in order to promote a greater understanding of Earth and Space Science research activities at GSFC and within the NASA research community. The Studio also provides expertise in data visualization and science storytelling, and it is part of the bigger Earth Science Storytelling team, which comprises three entities: The SVS, the Conceptual Image Laboratory (concept animators; non-data-driven products), and the Goddard TV Multimedia (a team of producers, science writers, editors, and web and social media experts).

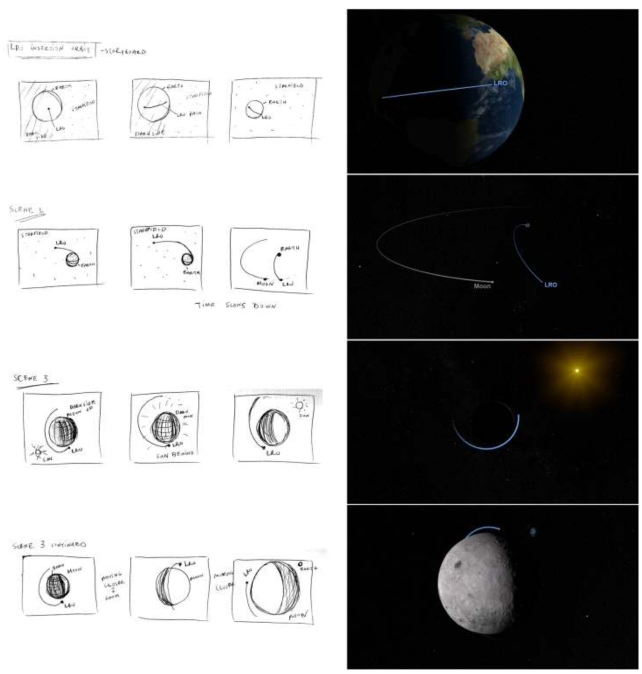
Data visualizations produced and developed at the SVS are cinematic-quality computer graphics short films, similar to productions by Hollywood computer animation studios. The main characteristics of SVS visualizations are: scientific integrity, preservation of data, seamless blending of multi-resolution data from different sources, aesthetics, and a solid story that engages the public. The successful production of such visualizations depends on free-form collaboration among members from all three teams and requires the following ingredients:

1) communication between all parties involved, including scientists;
2) data availability and transparency regarding limitations or problems in the datasets;
3) a context that makes the science story relevant and interesting to the public;
4) availability of resources for the production of a visualization story;

5) ability of participants to shift roles, wear multiple hats, and work collaboratively.

Storytelling is a key component in the production of every visualization at the SVS. Although storytelling manifests itself differently in various art forms, whether they be literary, performance-based, aural, visual, or interactive, a storyteller should know the story's audience and take ownership of the story. In general, all forms of stories are made up of the following ingredients: perspective, characters, imagery, and language. All of these ingredients are combined in a structure that defines the story from beginning to end. Visual storytelling, and specifically, storytelling for animation, borrows from the conventions of photography, cinema, episodic comics, and performance arts. The structure in storytelling for animation is established by camera work (visual perspective, time and space of framing, composition, point of view, lighting, color, form, and style), audio work (with or without, tone), and the visual, aural, and editorial rhythm of the animation.

Visualization-driven end products are archived in the SVS repository [2], which is a free and publicly accessible database with more than 3,000 entries (as of June 2011). The products span many visualization forms including 2D, 3D-Stereoscopic, Science-On-A-Sphere, Hyper-wall, Dome Show and even touch-display. Each production includes various formats including frame sets, still images, movies and, when appropriate, data in a wide gamut of resolutions. Upon release the products may take on lives of their own, since the public can use them freely.



From a storyboard to the story of NASA's Lunar Reconnaissance Orbiter (LRO) Mission, told as a sequence of images. **Left**: Storyboard. **Right:** Selected images from top to bottom: LRO launched from Cape Canaveral FL, USA; LRO approaching the Moon; LRO orbit trail shown with the sun and the dark side of the Moon; LRO moving into orbit around the Moon. [3]

Although there is a streamlined process in place for the production of visualizations, there are always challenges that might compromise the quality, structure, and story of the end product. These challenges are often rooted in data issues – for example, gaps in the data, insufficient amounts of data, low resolutions, or even data that do not show the expected phenomena. Other times, new visualization techniques are required in order to highlight important, necessary information. The need for new techniques can occur either within the technical infrastructure of the visualization production pipeline (e.g., modified shaders, transitions between different coordinate systems, or the development of a new pipeline) or within the design domain (e.g., finding the best ways to map complex data to visual models). In short, the amount of resources and effort required to produce high-quality visualizations for scientific storytelling can overwhelm any individual scientist.

The SVS is one example of a successful scientific storytelling and visualization studio. To create visualizations suitable for consumption by the general public, unique challenges must be overcome, and a dedicated team comprised of versatile, talented individuals is needed. In short, scientific storytelling is not a trivial endeavor, and the creation of successful visualizations requires the collective effort of many specialists working together.

Production Visualization at a Science Museum

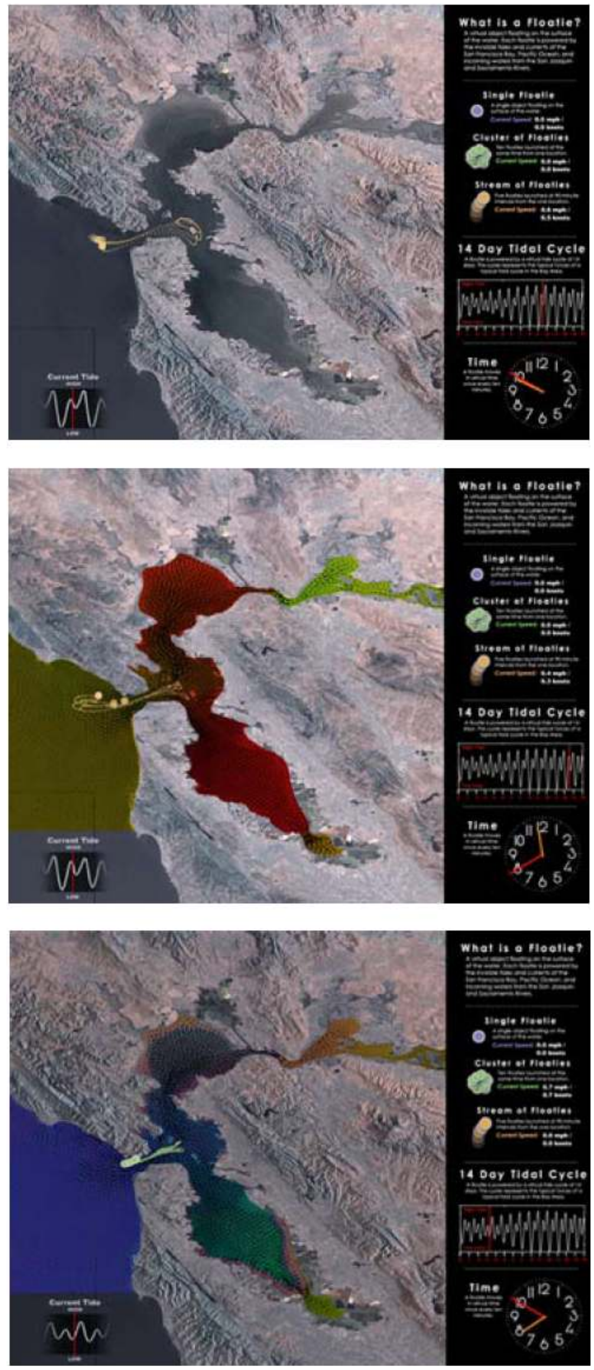
Science museums are places where people can experience science in ways they can't at school or at home. Museum visitors can swing on a giant pendulum, stand under a life-size *T. Rex* fossil skeleton, or watch the birth of a galaxy in 3-D. Museums tell the stories of science, and – perhaps more importantly – they also provide a unique venue for people across generations to play together, interact with scientists, and use scientific tools.

Museums have long used visualizations to show the public things they can't normally see, such as evolutionary relationships or the structure of DNA. But visualizations are an increasingly critical medium for science museums. As the volume of data collected by scientists expands exponentially, visualization is the tool that allows them to make observations or detect patterns. Whether comparing genomes, mapping the structure of a virus, or developing new models of Earth's climate, most scientists now do some – if not all – of their work using visualized data. To tell the stories of modern science, museums must use visualizations.

The growing importance of visualizations in science presents an exciting opportunity for museums. Scientific visualizations can provide stunning images, engaging the public with phenomena they've never seen before. Visualizations can be displayed on large, dynamic interfaces, providing new ways for the public to participate in interactive, social learning. Visualizations can also be used to create authentic tools for the public to make their own discoveries, analogous to microscopes or telescopes. But for scientific visualizations to have any significant meaning for the public, they must be carefully interpreted or designed.

Interpretation, or explanation, is essential when using scientific visualizations in museums. Visualizations often show complex and abstract phenomena at extreme size scales using colors that have no inherent meaning. For instance, a research study conducted at the Exploratorium, a science museum in San Francisco, showed that many visitors grossly misinterpreted the scale and use of color in a nanoscale image [4]. Similar studies have documented the difficulty of learners in interpreting visualizations from fields as disparate as genetics and astrophysics. As science museums increase their use of scientific visualizations, they are providing more interpretation through labels, videos, and live explanations. A complementary, though less common, strategy is to redesign scientific visualizations with the public in mind. Phipps and Rowe [5] conducted a study in which students were better able to interpret visualizations of oceanographic data that had been redesigned with more intuitive color schemes and recognizable (though unscientific) landmarks.

The most significant challenge for museums is finding ways to transcend the use of visualizations as explanatory animations or pretty pictures. In many museums, visitors can watch stunning simulations of Earth's climate or the collapse of a star, but they



The *Bay Model* at the Exploratorium in San Francisco allows visitors to interact with a scientifically accurate virtual model of how tides, currents, and rivers combine to create the complex water flows of the San Francisco Bay estuary. Using a touchscreen, visitors place virtual floats into a video image projected onto a three-dimensional geographic model of the Bay Area. After a float is launched, visitors watch how currents move the float to different locations according to predicted tide and river flow cycles. Color coding highlights varied water conditions during tidal phases. [6]

cannot control or explore these visualizations. Such direct interaction would allow visitors to control their experience and to make discoveries with data the way that scientists do. The Exploratorium is addressing this challenge by creating visualization tools for the public, where they can ask and answer their own questions with real scientific data. In a pilot project funded by the National Science Foundation, the Exploratorium is collaborating with visualization researchers. By tailoring their development process to different end users (the public) and iterating through intensive prototype testing with visitors, the Exploratorium hopes to create one of the first examples of a visualization tool that allows the public to explore real scientific data without mediation.

In summary, the increasing role of visualization in scientific discovery presents a tremendous opportunity for science museums to engage the public with stunning images, novel interfaces, and authentic tools. However, it is essential to realize that transforming the rapidly growing number of scientific visualizations into meaningful experiences for the public requires thoughtful interpretation, design, and collaboration.

Storytelling using Interactive Visualization

Following the publication of the NSF Report on *Visualization in Scientific Computing* in 1987, the early development of the field of scientific visualization was largely driven by the need to gain insight into large, complex datasets arising in science and medicine. This led to the invention of many new visual abstractions, rendering methods, and interaction techniques. Visualizations are used in *scientific storytelling,* but are generally created after the fact, separate and independent from the data exploration step. There is no storytelling model built into the visualization process; that is, there is no direct support for creating a story based on the visualizations made in the process of data exploration and knowledge discovery.

The concept of incrementally creating a story by depicting the forward progress of the visualization process is intuitive and powerful. The scientist is immersed in the data domain and assembles pieces of the story together as she learns more and more about the data. AniViz [7] realizes this concept by allowing the user of a visualization system to do exactly that, and to present the story as an animation. As the user interactively explores the data, he or she is able to locate intrinsic views of the data, specify some of these views as keyframes of an animation, review the animation constructed so far, add annotations and voiceovers, and edit the keyframes and transitions as needed until the exploration is complete and the resulting animation is satisfactory.

AniViz can effectively support storytelling within a visualization system based on the well-known keyframe approach from the field of computer animation. It is possible to think of this keyframe approach as a story model if it is more intuitive to the user. In earlier work [6], just such a story model is described, which is composed of:

(a.) *Story nodes* (major steps, or milestones, in a story, where we can imagine the story coming to a brief halt, perhaps for the purpose of interactive exploration by the story consumer, before resuming), and

(b.) *Story transitions* (which smoothly connect story nodes, leading from one node to the next).

It makes sense to consider story transitions as being composed of sequential or parallel actions. Based on this model, it is possible to create visualization stories that conform to the visual information-seeking mantra of Ben Shneiderman. A visualization story could begin with an overview of the data, followed by a focusing transition, leading the user to a more detailed visualization of some particular aspect, and conclude with a guided sequence of images that substantiate the message to be communicated. Other possible visualization stories could be constructed around the aim of comparative visualization (building up, for example, a side-by-side comparison during the story) or iterative visualization (such as the sequential visualization of all relevant features in a selected region, following a repetitive pattern like "zoom onto a particular feature, rotate around it, show context, then continue").

While storytelling, by nature, is not completely interactive, we ponder how interactive storytelling can be facilitated. How can we stimulate the participation of story consumers? Can we let them influence not only how the story is told, but the outcome of the story itself? For example, adventure games allow users to interact with and affect a premade game story. Also, in science museums, we find many hands-on activities, which may be considered a form of interactive storytelling. However, once spectators become "spect-actors" (the terminology of Augusto Boal, in "Theater of the

Oppressed" [8]), a conflict of control emerges: the spect-actor diverts the course of the story from the original plan. This is also known as the "narrative paradox," and different suggestions have been presented regarding how to address it (e.g., by the use of emergent narratives, as described by Aylett [9]).

A taxonomy of four different modes for splitting the control between the story author and the story consumer to varying degrees has been proposed [10]. First, the traditional *passive story telling* mode prohibits any interaction on the part of the story consumer; the story author has full control over all domains. Second, in *story telling with interactive approval*, passive storytelling pauses at certain points and allows spect-actors to take temporary control. Users can change the view, representation, and even content of the visualization. Once users are satisfied with this interactive exploration, storytelling continues as originally intended. Third, in *semi-interactive story telling*, the story consumer is allowed to take control not just for an interim excursion, but for an entire section of the story. Finally, in *total separation from the story*, story consumers may completely detach from the story and engage in interactive visualization with total freedom.

In terms of storytelling, interactive visualization has the potential to help with three issues that are important in the context of communication: comprehensibility, credibility, and involvement. First, by incrementally building up a story, enhancing it with labels and annotations, and allowing the user to interrupt the story and take control of it, the risk of presenting an overloaded visualization and having it poorly understood is reduced, thus improving comprehensibility. Additionally, the credibility of a visualization can be improved if observers are allowed to interact with it and verify that the visualization actually shows what it claims. Finally, allowing viewers to interact with visualizations "breaks the fourth wall," transforming them from passive observers to active participants. By actively participating in the process of visualization, viewers will feel a greater sense of engagement with the data being presented.

Conclusion

Clearly, there is a need to consider how storytelling and visualization can make scientific findings more comprehensible and accessible to the general public. Scientific visualization has much to learn from information visualization in this regard. Consider that information visualizations are aimed at the general public, and that they draw attention to differences and changes in visual elements. Perhaps scientific visualizations can take a similar approach in order to reach broader audiences. If we focus on important features by emphasizing how they change across time or experimental conditions, we may be able to tell a compelling story without having to explain extraneous details.

In addition, thinking about visualizations in a narrative context can help make them more comprehensible, memorable, and credible to the general public. Whether we use visualizations to tell a story, or use a story model to make visualizations more compelling, we cannot neglect the fundamentals of how to tell a good story. First, know your audience – assess their level of domain knowledge and familiarity with visualization conventions. Next, set the stage – make sure they have enough background knowledge about the specific dataset being visualized to make sense of your visualization. Introduce the characters – show them the visual elements and what they represent. Develop the plot – arrange your visual elements in a way that tells an interesting and compelling story. Finally, leave the audience with a lasting impression by showing them how the story is relevant to them, and its greater implications.

Scientific storytelling using visualization is not easy, and the successful examples highlighted in this article are the exception rather than the rule. Much work remains to be done in establishing guidelines and principles for successful storytelling. As visualization designers, we must ask ourselves how we can better support the efforts of the scientific community in reaching out to the general public. Scientists have amazing stories to tell, and we can help ensure that they are not – to paraphrase Mark Twain – forced to tell them themselves.

References

1. http://science.nasa.gov/missions
2. http://svs.gsfc.nasa.gov
3. http://svs.gsfc.nasa.gov/goto?3603
4. Joyce Ma. *Visitors' Interpretations of Images of the Nanoscale*. NISE Net Project Report, San Francisco: Exploratorium. 2008-07-30.
5. Molly Phipps and Shawn Rowe. *Seeing satellite data*. Public Understanding of Science, Volume

19, Number 3, May 2010, pp. 311-321.

6. http://www.exploratorium.edu/outdoor/#/exhibit/bay-model/
7. Hiroshi Akiba, Chaoli Wang, and Kwan-Liu Ma. *AniViz: A Template-Based Animation Tool for Volume Visualization*. IEEE Computer Graphics & Applications, 30(5), September/October 2010, pp. 61-71.
8. Augusto Boal. *Theatre of the Oppressed.* New York: Theatre Communications Group, 1979.
9. Sandy Louchart and Ruth Aylett. *Narrative theory and emergent interactive narrative.* International Journal of Continuing Engineering Education and Lifelong Learning, 14(6), 2004, pp. 506-518.
10. Michael Wohlfart and Helwig Hauser. *Story Telling for Presentation in Volume Visualization*. In EuroVis 2007 Proceedings, pp. 91-98.