Geospatial Information Integration for Science Activity Planning at the Mars Desert Research Station

Daniel C. Berrios[1], Maarten Sierhuis[2], Richard M. Keller[3]
[1]University of California, Santa Cruz
[2]Research Institute for Advanced Computer Science
[3]Intelligent Systems Division, NASA Ames Research Center,
Moffett Field, CA USA 94035
dberrios@mail.arc.nasa.gov

***Abstract**. NASA's Mobile Agents project leads coordinated planetary exploration simulations at the Mars Desert Research Station. Through ScienceOrganizer, a Web-based tool for organizing and providing contextual knowledge for scientific datasets, remote teams of scientists access and annotate datasets, images, documents, and other forms of scientific information, applying pre-defined semantic links or meta-data using a Web browser. We designed and developed an experimental geographic information server that integrates remotely-sensed images of scientific activity areas with information regarding activity plans, actors, and data that had been characterized semantically using ScienceOrganizer. The server automatically obtains remotely-sensed photographs of geographic survey sites at various resolutions and combines these images with scientific survey data to generate "context maps" illustrating the paths of survey actors, and the sequence and types of data collected during simulated surface "extra-vehicular activities." The remotely located scientific team found the context maps were extremely valuable for achieving and conveying activity plan consensus.*

1 Introduction

Through the proliferation of high-speed communication networks and wide-spread availability of desktop computing systems, researchers in many different fields should now be able to conduct data gathering and analysis campaigns that involve larger groups of collaborating scientists separated by great distances. However, the design requirements for computer-based systems to support these efforts are not yet fully known. There is some evidence that functions such as synchronous electronic chat and data annotation capabilities (Olson et al., 1998) and geospatial displays for locating and planning scientific data collection (Ogren et al., 2004) can be quite useful. But roles for many other functions remain to be established or elucidated. For example, the relationship between the nature of data collected (e.g., qualitative vs. quantitative), collection and analysis methods employed (e.g., automated vs. non-automated), or the domain of investigation, and the optimal design of systems supporting scientific collaboration is still unclear. In this study, we provide our experience developing and deploying a system for generating geospatial and temporal traces for scientific data through dynamic integration of semantically tagged information.

For several years, the Mars Desert Research Station (MDRS) near Hanksville, Utah has been the location of planetary exploration simulation activities conducted by NASA's Mobile Agents project (Clancey et al., 2001). Part of the research activities conducted during these simulations includes studies of the process of scientific collaboration, activity planning, sharing and review of collected scientific data, and the design and development of computer-based tools to support these processes. We have had the opportunity to participate in these investigations through our work developing and deploying ScienceOrganizer (Keller et al., 2004) during Mobile Agents MDRS fields tests in 2003 through 2005.

ScienceOrganizer is a Web tool for managing contextual knowledge (Dey et al., 2001) of scientific datasets (Berrios et al., 2004) specifically developed to support the work of collaborating scientific and engineering teams. Through ScienceOrganizer, remote teams of scientists can access and annotate datasets, images, documents, and other forms of scientific information, supplying additional meta-data and interconnecting them through pre-defined logical relationships using any Web browser. Information stored thusly in ScienceOrganizer is semantically characterized along multiple dimensions, providing users with more precise "tags" with which to find data compared to traditional information storage systems. In addition, this semantic tagging can provide users with valuable cues regarding information purpose, provenance, and pedigree, and can assist in information navigation. Finally, semantic tags can be used to support two functions of advanced scientific information systems: information integration and inference.

During 2005, we specifically sought to study how temporal and geospatial meta-data can be used to organize and present scientific information for improved access and sharing. In this report, we discuss the development of an experimental geographic information server (GIS) that integrated remotely-sensed images of scientific activity areas (field sampling sites, geographic features, etc.) with information regarding activity plans, actors, and data that had been characterized semantically using ScienceOrganizer. The server automatically obtained remotely-sensed photographs of geographic survey sites at various resolutions and combined these images with scientific survey data to generate maps illustrating the geospatial paths of survey actors, and the temporal sequence and types of data collected during simulated surface "extra-vehicular activities" (EVA). This integration played key roles in support of scientific decision making for activity planning and execution prior to and during EVA.

2 Methods

The primary purpose of the GIS is to support the needs of remotely located scientists to help plan and monitor the activities of a geological survey team in the field. It leverages the semantic meta-data stored in ScienceOrganizer to formulate requests for geographic, topographic, and photographic information from a publicly available archive (Terraservice-USA, microsoft.terraserver.com). After obtaining this information through Web services, the GIS synthesizes temporo- and geospatial maps showing the precise sequence and location of scientific activities and data products collected during simulated EVA. .

2.1 Mobile Agents

NASA Ames' Mobile Agents Architecture is a distributed agent-based software architecture that integrates diverse mobile entities in a wide-area wireless system for simulated lunar and planetary surface operations (primarily EVA). Software agents, implemented in the Brahms multi-agent language, run in virtual machines onboard laptops integrated into space suits, robots, or located in habitats (Clancey et al., 1998, Sierhuis et al., 2002). "Personal agents" support the crew in the habitat and on the surface, who communicate with the agents via a speech dialogue system. All the actors (human and robotic agents) in the simulations are outfitted with high-precision global positioning devices that continuously track their locations. As data (e.g., digital images, voice recordings, sample measurements, etc.) are collected during EVA simulations, software agents transmit the data via a dynamically-configured wireless network to an installation of ScienceOrganizer located in the habitat. These agents generate and tag the data with a pre-defined set of meta-data that varies depending on the type of data collected. However, for collected scientific data, this meta-data always includes the GPS location of the agent that collected the data. The crew in the habitat can then view data and meta-data in real-time in ScienceOrganizer (Figure 1).

The participants in the 2005 field tests collected many images of sampling sites and surrounding areas and recorded voice notes describing major land features. While these data can provide a context for current and past activities, and help plan for future activities, it has proved difficult for off-site participants (in a "surface" habitat or in a mission control center) to relate data products to other data products or to activities temporally and geospatially. With advances in the remote-sensing capabilities of satellites orbiting earth, and the now-widespread availability of the high resolution images they produce, we sought to develop a system for dynamic integration of these image data with the data and meta-data collected at MDRS.

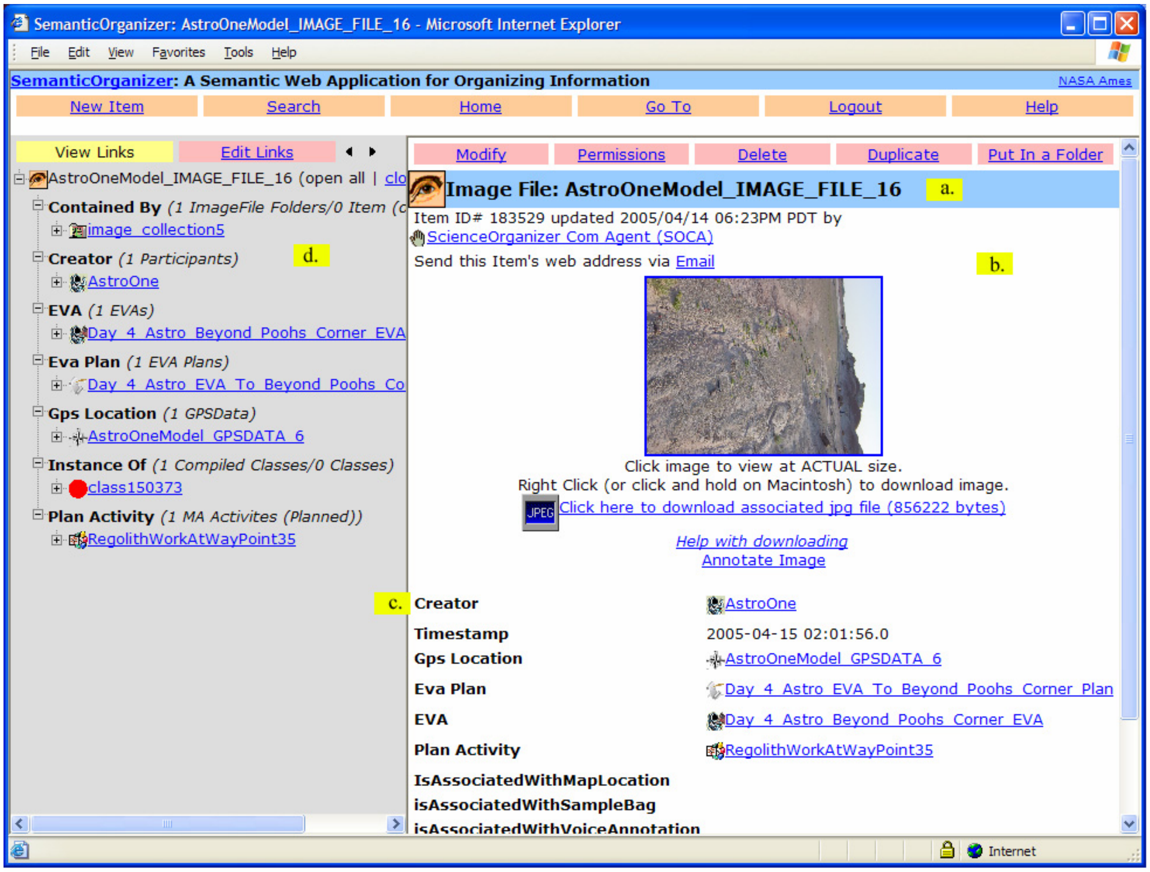

Figure 1. The Web application, *ScienceOrganizer*. This display shows details of one data item, an "Image File," including its name (a.), a "thumbnail" version of the image (b.), and a portion of its meta-data (c.) and semantic links (c, d.) to other data items.

2.2 TerraService USA

TerraService-USA (Terraservice.microsoft.com) is a Web service/site that provides access to remotely-sensed aerial photographs and topographic maps of the earth's surface (limited to the USA) for a range of spatial resolutions. A number of Web services are offered by TerraService-USA in addition to image access, including area/region identification and gazetteer functions. Image resolution varies by geographic area, with very high resolution imagery (the kind required for supporting scientific activities such as precise instrument deployment or manipulation) available only in limited (mostly urban) areas. Images are offered as fixed sized tiles (each identified by a unique set of metadata) from which client programs can compose larger images *ad hoc*.

2.3 Selecting and Integrating Information

The GIS uses the semantic data and meta-data generated by the ScienceOrganizer Communication Agent (a Brahms agent) and attached to various kinds of scientific data when it stores them in the ScienceOrganizer information system. These meta-data provide key contextual knowledge for collected data. For example, as a new survey panoramic image is stored in ScienceOrganizer, its "EVA" property is set to a reference to the particular EVA simulation during which it was collected, its "GPS data point" property is set to a reference to the latitude/longitude coordinates where the photo was taken, etc. The GIS gathers this contextual knowledge from ScienceOrganizer, then generates requests for imagery from the TerraService-USA using this knowledge. It then combines the returned imagery using the contextual knowledge of the region of the EVA to yield a context map image (Figure 2).

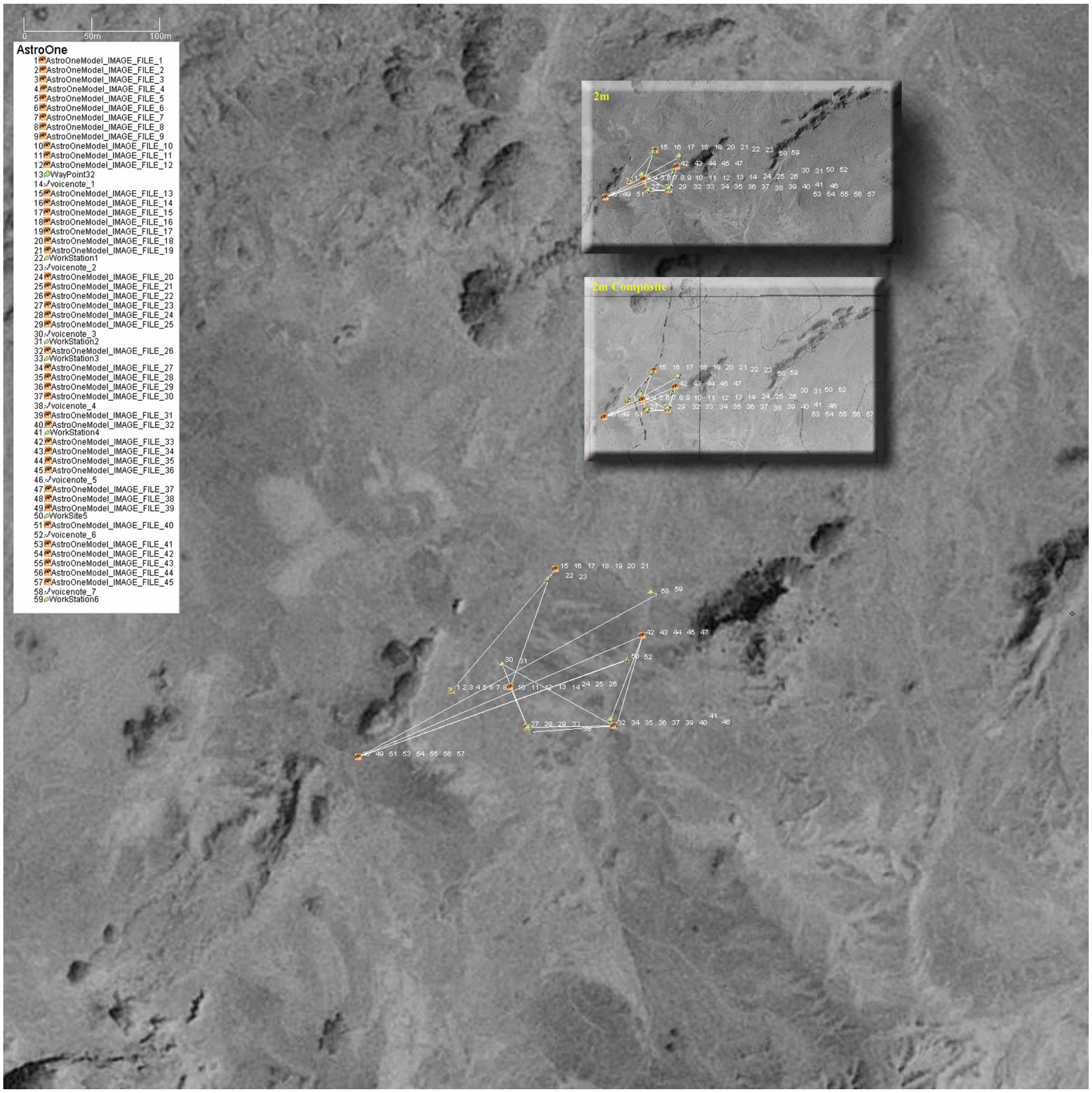

Figure 2. Context map image created by the GIS and stored in ScienceOrganizer. The larger image is a 1m-resolution photograph with overlaid EVA information. The two smaller inset images are 2m-resolution photographic map (top) and 2-m composite photographic/topographic map with the same information overlaid.

The contextual knowledge used by the GIS to generate context maps includes the scientific area-of-interest bounding box, inferred from the GPS coordinates of all the scientific data products collected during the EVA simulation (as of the time of map generation). The bounding box (plus additional area for displaying a map legend and scale) determines the size of final context map image. The GIS obtains the necessary tile images from the Terraservice-USA to span the entire bounding box, and then stitches these images together to form the base layer of the context map image. Next, sequential numbers are drawn at the locations of each of the data products according to their order of collection. Data products collected closely in time and space often can result in overlapping and unreadable numbers; in such cases, we chose to displace subsequent numbers horizontally (from left to right, as in Figure 3). Finally, points of collection are connected by (straight) lines to indicate rough traverse paths of agents. Future versions of the GIS could produce maps with more precise (non-linear) path traces, as actor GPS coordinates are ascertained every second.

Because actors involved in the simulations frequently cross paths, we programmed the GIS to use knowledge of the collecting actor to sort data and create a context map for each actor (Figure 3). This was done to minimize confusion between actor traces on the final map; however, these actor-specific map images do not show shifting spatial relations between actor over time, which may be important for certain collaborative scientific activities (e.g., collection tasks that require two or more actors working with a specified distance between them). Also, there are clearly some situations in which showing all path traces on a single map would prove valuable (e.g., for robot path planning). Finally, we repeated the map image creation process using aerial photographic images at spatial resolutions of 1m and 2m, using topographic maps at 2m resolution, and, finally, compositing 2m topographic and photographic images with semi-transparency (as shown in Figure 2, inset). We also transformed 1m photographic images by doubling their dimensions to simulate 0.5m context map images. These maps proved valuable in displaying simulations with tightly spaced activities, in spite of the reduced clarity of map images.

In addition to constructing these context map images, the GIS also generated HTML image maps, and combined them with the context map images to form "dynamic" (user-actionable) area maps that it also stored in ScienceOrganizer. Each data product shown on these maps was linked to its (unique) location (URL) in ScienceOrganizer so that users could readily move from context map image to hyperlinked data product (Figure 4).

Finally, the GIS stored all context map images and map documents in ScienceOrganizer as soon as they were generated, generating semantic links between the maps and the simulations shown on the maps. This created an efficient way for users to navigate from map to simulation to data products and back. As the EVA simulation progressed, the GIS updated these map images and documents, enlarging EVA area bounding box (and map) size, and adding additional data collection points, actor paths, and links to data products in ScienceOrganizer continuously. If new actors joined the simulation, new maps and documents displaying their collected data locations were created. Using these maps, remotely located ScienceOrganizer users were able to follow the activities of the field scientists in near real-time.

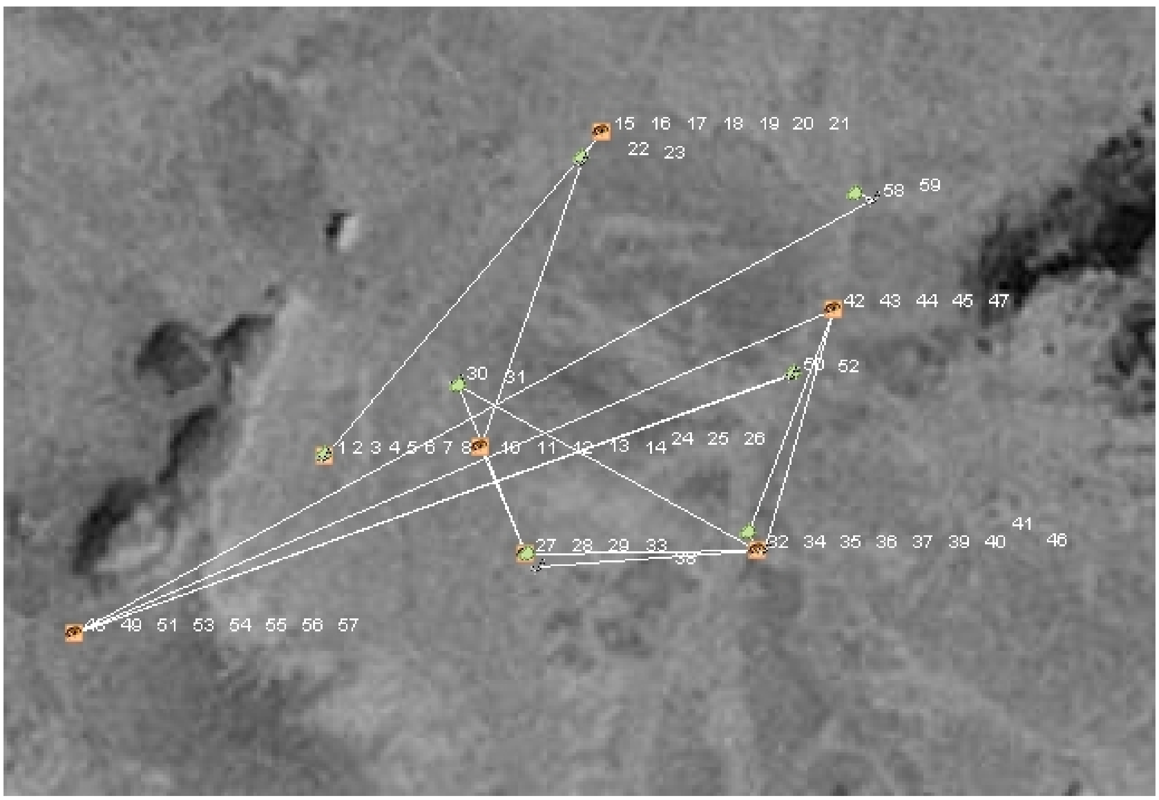

Figure 3. A one-meter resolution GIS-generated map showing the activities of a single actor, "Astro-One," during a simulation conducted April 14, 2005. "Rock Hill" is recognizable at left (see Figure 5). Collected data items are numbered sequentially (beginning due east of "Rock Hill"). Note evidence of imperfect determination of collection sequence (e.g., Nos. 32, 33, and 34), due to erroneous timestamps of large data files (see Results).

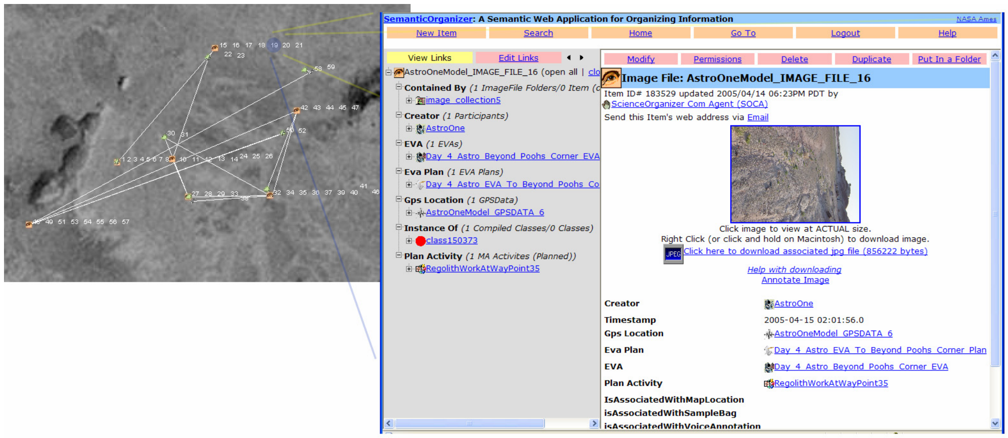

Figure 4. Illustration showing use of an image map to link to a data product collected during the EVA shown in Figure 2. In this case, the representation of a collected digital photograph (a type of "Image File") is selected by clicking on the ordinal number "19" on the image map (i.e., the 19th collected data product), after which the user is shown the item in ScienceOrganizer (right, including a thumbnail of the photograph, its meta-data and semantic links to other contextual information in ScienceOrganizer).

3 Results: MDRS 2005

We had the opportunity to test the performance of the GIS and the use of the context maps in EVA simulations conducted at MDRS in April 2005. Overall, generation of the map products was considered timely by the participants, although there was often a delay of one or two minutes from data collection to map generation. The GIS was programmed to search continuously for active simulations, and then create context maps so that users could track the progress of scientific activities in near real-time, overwriting any previous context maps (since the data on those maps was no longer "current"),. For the kinds of geological survey and sampling work performed at MDRS and the use of the maps by remotely located scientists, the delay in map generation was not deleterious. Periodically, the Web services offered by the Terraservice became unavailable. This resulted in map products being deleted but not replaced by the GIS with updated versions. Users complained that maps had mysteriously evaporated on a couple of occasions. About halfway through the field tests, we built into the GIS an adaptive capability to detect the availability of Terraservice-USA and abort map generation before deleting existing maps when the service was unavailable.

Having fine-tuned the performance of GIS, we then sought evidence of whether the context maps produced were useful to the remotely located team of scientists who were guiding the field geologists, specifically whether they aided them in planning the field team's next set of activities. One of the authors (Berrios) participated in the Remote Science Team (RST) planning meetings held before selected simulations during the field tests. Prior to the field tests, one of the field geologists had independently obtained aerial photographic maps of the sites likely to be surveyed, and annotated the maps with *ad hoc* feature names (Figure 5). The RST adopted these feature names and frequently used them to identify data products in planning-meeting discussions (e.g., "sample 3 taken at Red Hill"). The context maps proved extremely useful for coordinating agreed-upon features with specific data products; the team found it very difficult to distinguish and refer to data products based solely on collected GPS coordinates. Furthermore, the RST found that the context maps were extremely valuable for conveying consensus plans for the following day of simulation activities. In particular, they used the image-annotation capabilities in ScienceOrganizer () to draw these plans on the context maps themselves, indicating which areas in the region the field scientists should investigate next with geometric shapes (circles, ovals, etc.) and words (e.g., "Sample here").

The Brahms agent responsible for storing collected data (images, voice notes, etc.) in ScienceOrganizer was designed to tag those data with (time of collection) timestamps only after the data was completely received by the agent. This design of yielded occasional errors in the sequence of data products as shown on the context maps (see Figure 3), because larger data files required significantly longer transmission times over the wireless network in the field. However, we found no evidence thatthis kind of relatively small imprecision in temporal sequencing significantly impacted planning or other science activities during the simulations.

One feature that the context maps lack is reference to a coordinate system. The field team requested that the RST deliver GPS waypoints that could be incorporated into specific activity plans required by the Mobile Agents system. A grid overlaying the maps showing latitude and longitude tick marks would have been very useful for this purpose.

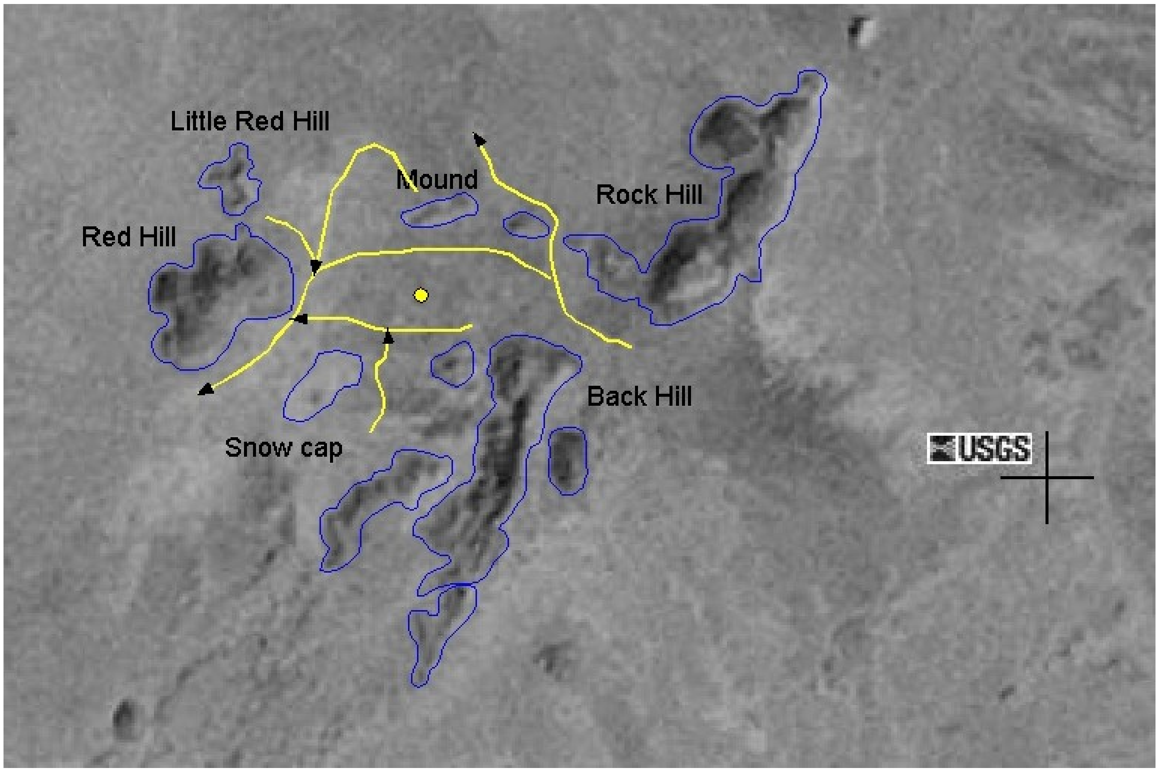

Figure 5. Annotations of composite USGS images of proposed geologic survey sites at MDRS drawn manually by a field geologist. Blue outlines indicate major land features, and yellow lines indicate likely fluvial traces.

4 Related Work

The use of "collaboratory" applications like ScienceOrganizer in which multiple actors share data and contextual knowledge is growing (Olson et al., 1998, Bebout, 2002, George Chin and Lansing, 2004), but is still far from widespread. This has limited the ability to study the nature and utility of various types of geospatial contextual information; certainly the optimal way to integrate geospatial, temporal and other types of information to trace moving objects that have arbitrary paths and relationships remains to be demonstrated. There is some knowledge as to the most efficient methods to conflate geospatial information from heterogenous sources in order to show locations of stationary objects (e.g., buildings (Michalowski et al., 2004)), and to locate moving objects with predefined paths (e.g., trains (Shahabi et al., 2001)), although the actual usefulness of this information in real-world settings that involve collaborative decision-making has yet to be shown.

Planetary surface visualization applications and frameworks that can provide sophisticated, three dimensional views of terrain with information overlays from other sources (e.g., Google Earth (earth.google.com) and NASA World Wind (worldwind.arc.nasa.gov)) could be used to provide the same type of geospatial context cues as the GIS. For the type of collaborative work involved in the Mobile Agents project, this approach presents some important issues to explore. Can the use of these sophisticated tools be incorporated into the workflow of teams like Mobile Agents as successfully as the GIS? What is the value of the type of dynamic geospatial views produced by these applications, compared with the static but persistent kind of context maps generated by the GIS?

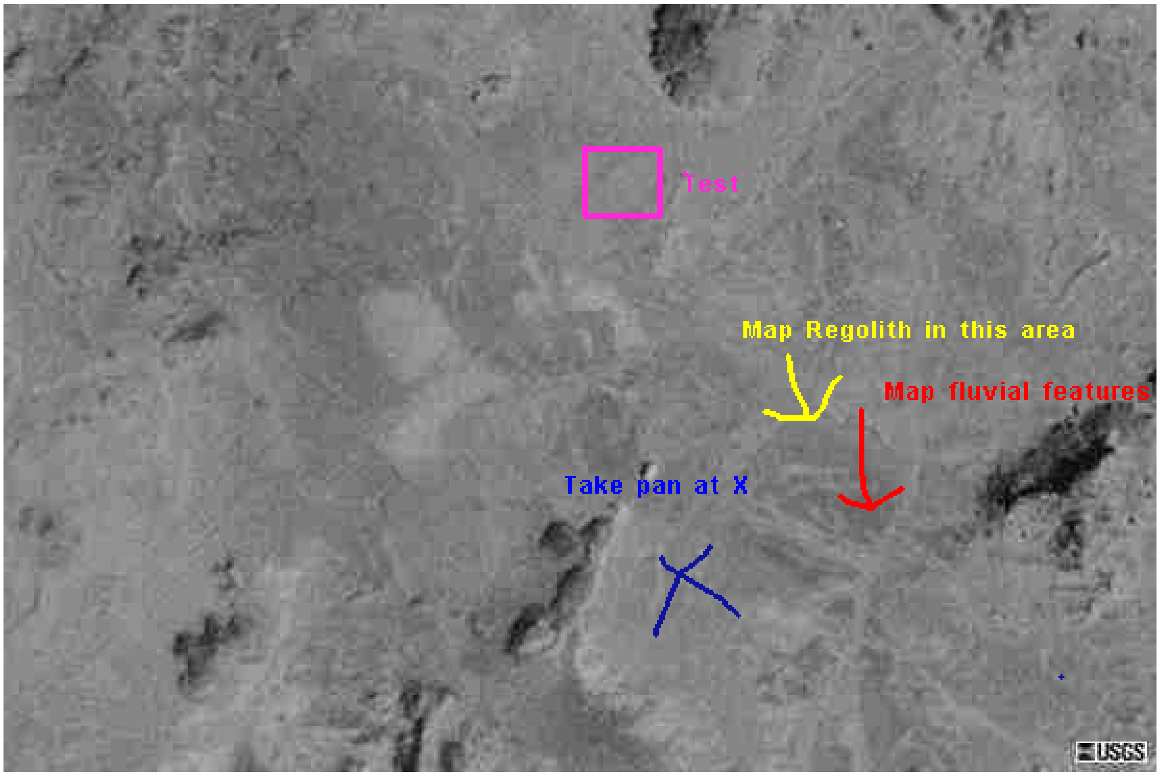

Figure 6. A human-annotated map image of the same area shown in Figures 2 and 3. After examining the GIS map shown in Figure 2, one of the remote science team members uploaded a cropped image (obtained through USGS) from her laptop to ScienceOrganizer, and then used the ScienceOrganizer Image-annotater to draw notional instructions for the field geologists directly on the image (the box at the top represents an initial test by the scientist of the Image-Annotator's capabilities). Instructions included 1) "Take pan (i.e., panoramic photograph) at (blue) X"; 2) "Map regolith in this area" (indicated by the yellow arrow); and 3) Map fluvial features" (in the area indicated by the red arrow).

5 Discussion

Coordinating medium- and large-scale scientific data-gathering campaigns presents many challenges, especially when some actively involved participants are not physically located at or near collection sites. We focused on supporting the collaboration between such participants and the rest of the team with computer-based tools that allow all participants to monitor campaign progress, view scientific data gathered with coordinated contextual geospatial, temporal, and geographic information, and jointly plan on-going campaign activities. We had the opportunity to test the utility of tools we developed as part of an existing software agent system to simulate the collaboration of *in situ* and remotely located geologists cataloguing sites of interest and collecting scientific samples. This kind of distributed collaboration is becoming the model for more and more scientific data gathering campaigns, and closely parallels the current (late-mission) collaboration model followed by participants in the on-going Mars Exploration Rover missions (Wick et al., 2005).

The remotely located science team made significant use of the technologies we developed to guide humans and robots during scientific field investigations. Our primary role during these experiments is to support such collaboration, but also includes attempting to identify obstacles to conducting such investigations efficiently. The participants in the collaboration were very receptive to trying new computer-based tools for communication and planning, and this must be kept in mind when comparing our observations to others in this field.

We observed scientists' use of context map images and image maps to discuss and guide sample collection in the field, arguably the most important function of remote scientific teams participating in such campaigns. During the 2005 field tests that we observed, the members of the remote science team found the maps, which depicted the precise sites of all samples collected, as well as locations where images and voice notes were recorded, very valuable for coordinating agreement on recommended plans for further field investigation. This contrasts with our observations in 2004 of similar field studies by the same team that did not have the type of maps produced by the GIS at their disposal. During the 2004 field tests, the scientists frequently spent a great deal of time trying to determine temporal sequence of collected data, to associate data with locations, and to grasp and communicate the geospatial relationships of these locations during planning meetings. The type of data product access and visualization provided by the context image maps during the 2005 field tests reduced the frequency of such "down time" activities, leaving the scientists more time to concentrate on discussing concerns relevant to domain scientific work.

Acknowledgements

We wish to acknowledge the significant contributions of Ron Van Hoof, Mike Scott, Ian Sturken, and Shawn Wolf. We also acknowledge our field geologists Abigail Semple and Brent Garry, and all members of the Mobile Agents Remote Science Team led by Shannon Rupert.

References

Bebout, B. M., S. P. Carpenter, D. J. Des Marais, M. Discipulo, T. Embaye, F. Garcia-Pichel, T. M. Hoehler, M. Hogan, L. L. Jahnke, R. M. Keller, S. R. Miller, L. E. Prufert-Bebout, C. Raleigh, M. Rothrock and K. Turk (2002). "Long term manipulations of intact microbial mat communities in a greenhouse collaboratory: Simulating Earth's present and past field environments", *Astrobiology,* 2**,** 383-402.

Berrios, D. C., Sierhuis, M. and Keller, R. M. (2004). "Organizational Memory for Interplanetary Collaborative Scientific Investigations" *7th Internataional Mars Society Conference*Chicago, IL.

Clancey, W. J., Lee, P. and Sierhuis, M. (2001). "Empirical requirements analysis for Mars surface operations using the flashline Mars Arctic Research Station" *Fourteenth International Florida Artificial Intelligence Research Society Conference*AAAI Press, Key West, FL, USA, pp. p 24-6.

Clancey, W. J., Sachs, P., Sierhuis, M. and Van Hoof, R. (1998). "Brahms: simulating practice for work systems design", *International Journal of Human-Computer Studies,* 49**,** 831-65.

Dey, A., Abowd, G. D. and Salber, D. (2001). "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications", *Human-Computer Interaction,* 16**,** 97-166.

George Chin, J. and Lansing, C. S. (2004). "Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory" *Proceedings of the 2004 ACM conference on Computer supported cooperative work*ACM Press, Chicago, Illinois, USA, pp. 409-418.

Keller, R. M., Berrios, D. C., Carvalho, R. E., Hall, D. R., Rich, S. J., Sturken, I. B., Swanson, K. J. and Wolfe, S. R. (2004). "SemanticOrganizer: A customizable semantic repository for distributed NASA project teams" *3rd International Semantic Web Conference (ISWC2004)*Springer-Verlag, Hiroshima, Japan.

Michalowski, M., Ambite, J., Thakkar, S., Tuchinda, R., Knoblock, C. and Minton, S. (2004). "Retrieving and semantically integrating heterogeneous data from the Web", *Intelligent Systems, IEEE [see also IEEE Intelligent Systems and Their Applications],* 19**,** 72-79.

Ogren, P., Fiorelli, E. and Leonard, N. E. (2004). "Cooperative control of mobile sensor networks:Adaptive gradient climbing in a distributed environment." *IEEE Transactions on Automatic Control,* 49**,** 1292-302.

Olson, G. M., Atkins, D. E., Clauer, R., Finholt, T. A., Jahanian, F., Killeen, T. L., Prakash, A. and Weymouth, T. (1998). "The Upper Atmospheric Research Collaboratory." *Interactions,* 5**,** 48-55.

Shahabi, C., Kolahdouzan, M. R., Thakkar, S., Ambite, J. L. and Knoblock, C. A. (2001) In *Proceedings of the 9th ACM international symposium on Advances in geographic information systems* ACM Press, Atlanta, Georgia, USA pp. 34-40

Sierhuis, M., Clancey, W. J. and Sims, M. H. (2002). "Multiagent modeling and simulation in human-robot mission operations work system design" *35th Annual Hawaii International Conference on System Sciences*(Ed, Sprague, R. H.) IEEE Comput. Soc, Big Island, HI, USA, pp. p 191-200.

Wick, J. V., Callas, J. L., Norris, J. S., Powell, M. W. and Vona, M. A., III (2005). "Distributed operations for the Mars Exploration Rover Mission with the science activity planner" *Aerospace, 2005 IEEE Conference,* pp. 4162-4173.