545061

Linking Science Analysis with Observation Planning: A Full Circle Data Lifecycle

Sandy Grosvenor[a], Jeremy Jones[b], Anuradha Koratkar[c], Connie Li[d], Jennifer Mackey[d], Ken Neher[a], and Karl Wolf[d]

[a]Booz-Allen Hamilton, 7404 Executive Place, Seabrook, MD 20706
[b]NASA Goddard Space Flight Center, Greenbelt, MD 20771
[c]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218
[d]Commerce One, 1100 West Street, Laurel, MD 20707

ABSTRACT

A clear goal of the Virtual Observatory (VO) is to enable new science through analysis of integrated astronomical archives. An additional and powerful possibility of the VO is to link and integrate these new analyses with planning of new observations. By providing tools that can be used for observation planning in the VO, the VO will allow the data lifecycle to come full circle: from theory to observations to data and back around to new theories and new observations.

The Scientist's Expert Assistant (SEA) Simulation Facility (SSF) is working to combine the ability to access existing archives with the ability to model and visualize new observations. Integrating the two will allow astronomers to better use the integrated archives of the VO to plan and predict the success of potential new observations more efficiently. The full circle lifecycle enabled by SEA can allow astronomers to make substantial leaps in the quality of data and science returns on new observations.

Our paper examines the exciting potential of integrating archival analysis with new observation planning, such as performing data calibration analysis on archival images and using that analysis to predict the success of new observations, or performing dynamic signal-to-noise analysis combining historical results with modeling of new instruments or targets.

We will also describe how the development of the SSF is progressing and what have been its successes and challenges.

KEYWORDS: data lifecycle, virtual observatory, visual tools, observation planning

1. INTRODUCTION

The SEA team was created to investigate the role of software in reducing science operations costs for the Next Generation Space Telescope (NGST). The team prototyped proposal development tools for the Hubble Space Telescope (HST) because end-user support is a major portion of the HST's science operations cost. The SEA team is currently attempting to prototype an SEA Simulation Facility (SSF) that consists of a visual framework that integrates observation modeling with the ability to easily "plug-in" different target models and/or instrument and observatory models. The goal is to allow astronomers to visualize and compare different observation simulations.

The research we have done of prototyping new proposal development tools and the feedback we have received from practicing astronomers has reinforced our belief that there is a critical need for visual, intuitive, and user-friendly tools in the field. This will be even more essential to the success of the VO, where the software tools will be the "instruments", providing the equivalent data capture functions performed by instrument hardware in traditional observatories. One important piece of maximizing the effectiveness of the VO is receiving relatively little discussion. It is a link that will turn the science data flow from a predominantly linear process into an integrated full circle where new VO-enabled analyses will, in turn, spawn new ideas and new observations.

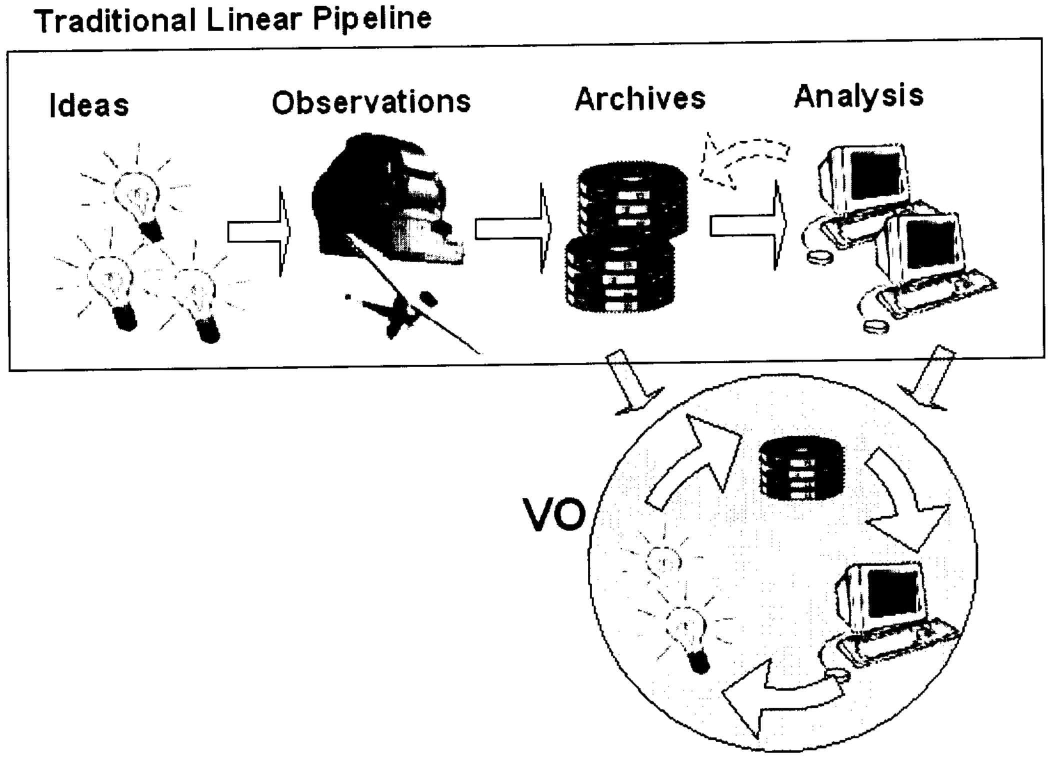

Figure 1: The Virtual Observatory adds a valuable new step in the traditional linear pipeline.

2. THE NEED FOR FULL CIRCLE

Current data pipelines are primarily linear (see Figure 1): they start with an observing proposal, generate observations and archival data and then end with scientific analysis on that archived data. Some of the analytical results (represented by the dotted arrow) also feed back into the archives. The VO will provide a new and exciting ability to make new scientific discoveries using archives. It will dramatically improve the value of archive programs by integrating access to archives from different observatories with advanced data mining and data analysis tools. This effectively adds a new phase to the end of the traditional observing pipeline (shaded circle in Figure 1).

We suggest that to realize the true capabilities of the VO, it is critical to integrate this new "end" of the pipeline back into the start of the pipeline. A *full circle data pipeline* will link the current "end" of the current data pipeline – data archives and analyses – with the beginning of the pipeline – the design and planning of new observing proposals. We believe it is critical to have tools developed for VO that will be able to integrate VO's discoveries back into a new iteration of observing proposals since archival analysis and VO studies will trigger new ideas for new observations. These tools can easily integrate back into the formulation of ideas into new observations (Figure 2).

The VO should be thought of as much more than just a repository of multi-wavelength data located across the Internet. As the name suggests it is an observatory whose assets (databases and software tools) will help in the discovery of new science that is traditionally not possible. *A fundamental service of an observatory is to support its users in the investigative process so that the user can move from one facet of science data life cycle to another (and back!) as intuitively as possible.* Hence, when we discuss the design and development of the VO we need to consider a paradigm in which the complete astronomical investigative process is not compartmental or linear, but is a seamless and intuitive flow from one aspect to another through the science data life cycle. Astronomers do this intuitively today, but they must

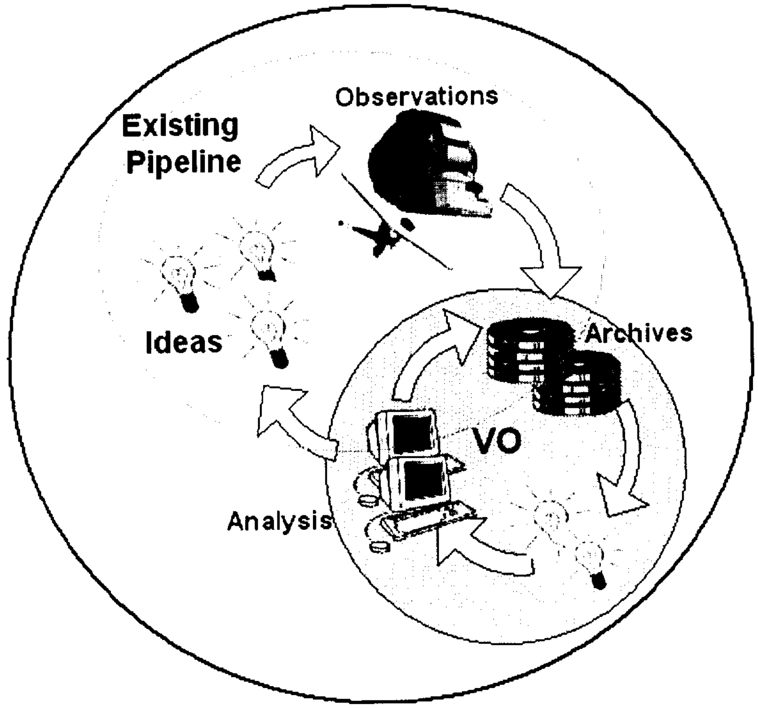

Figure 2: A more effective full circle model for VO pipeline.

use a variety of tools that do not seamlessly communicate with each other and have significant learning curves, and are not applicable across different observatories.

Further, in the era of the VO, the archives that make up the VO must be populated by high quality, well-defined, well-characterized data to allow new scientific explorations and investigations. Routinely obtaining quality data is important to any observatory, but this task becomes critical if we are to realize the potential of the VO. *Obtaining high-quality scientific data from either space-based or ground-based observatories requires efficient, informed, and complete observation planning.* Yet, in the present discussions of the VO we rarely discuss observation-planning tools, which can integrate the VO's observing capabilities with other "traditional" observatories. These planning tools do not just encompass the defining of science goals (phase 1) or laying out proposed observations (phase 2), but also developing and refining ideas and observing hypotheses (phase 0), i.e., not just tools to help prepare an observing proposal, but tools that will help simulate observing conditions and data and finally develop an observing proposal.

Another key to the success of the VO is the ability of archives to be able to provide users with the "pedigree" of the data in the archives. Planning tools that help simulate data are extremely useful not only to design an observation, but also to characterize and determine the quality of any archival observation. This characteristic of planning tools is extremely important if astronomers who are not the original observers are to use the data effectively. Thus we see that planning tools as conceived above are not only essential for observation planning but are also just as important for archival image analysis. These tools can *provide a better understanding of the limitations of the archive data, and improve our ability to quantify scientific results and conclusions from the data.*

3. WHAT IS THE SEA SIMULATION FACILITY

The SEA Simulation Facility is an extension of our original SEA work. It is a prototype effort designed to explore new visual approaches to allow astronomers to simulate the quality of proposed observations based on known observing parameters, such as target properties (e.g. brightness, spectral energy distribution), instrument setup (e.g., choice of detector gain, dithering, orientation of field-of-view, level of calibration), and observatory conditions (e.g. sky

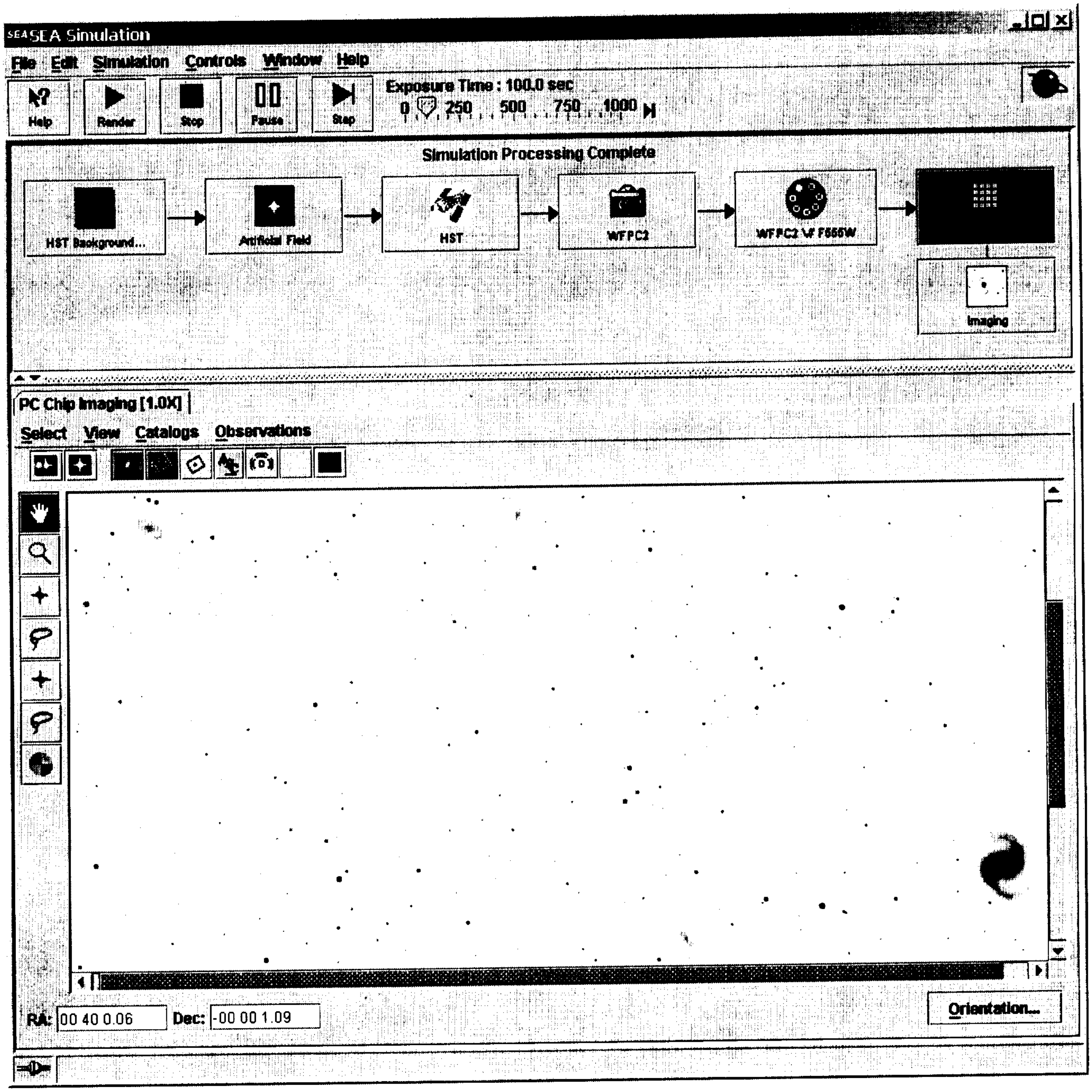

Figure 3: SEA Simulation Facility.

conditions). With the tools in the facility, observers can make better and more informed decisions about scientific trade-offs in real-time and they will be able to easily recognize and prevent problems.

The idea behind the SSF is to provide a simple but useful and effective facility to model an exposure. The SSF consists of a graphical "pipeline" and one or more visualizations (Figure 3). The pipeline represents the light path from source to detector. The pipeline consists of a sequence of models (called EnergyModels) that model some aspects of the light path. The pipeline is arranged in a left-to-right fashion where the source models are on the left and the right most model is the detector. The arrows indicate the direction of flow of the photons in the light path.

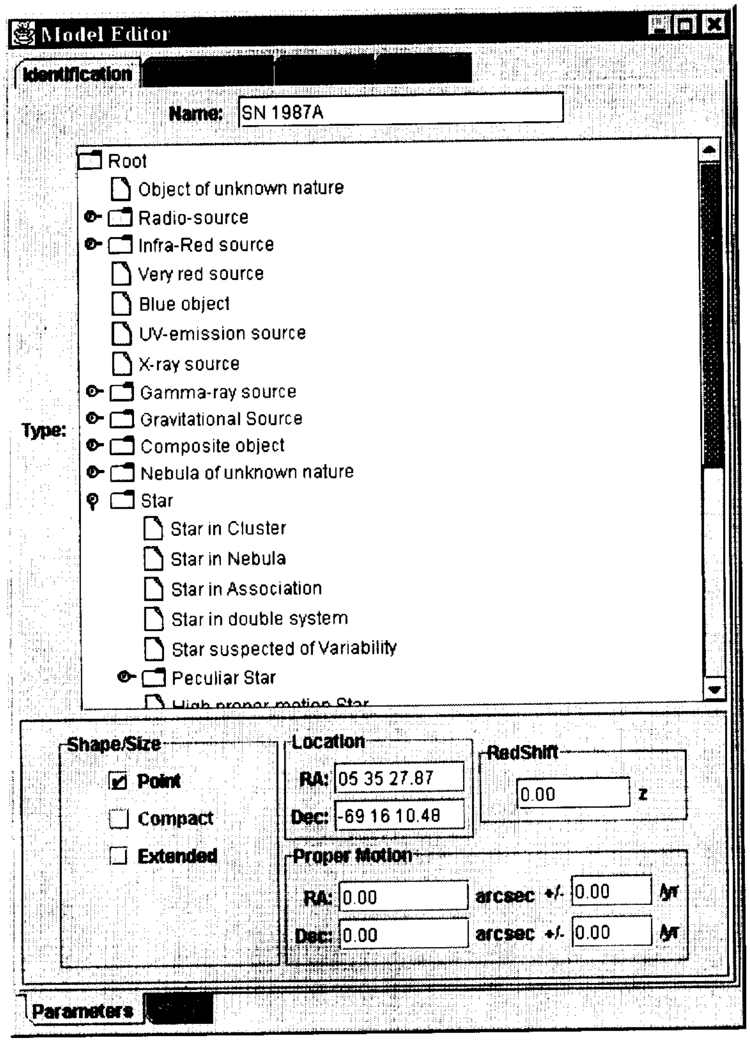

Figure 4: Sample Point Source Editor.

Each EnergyModel models some specific behavior of the light. For example, an EnergyModel can model a Point Source object such as a star. In addition to modeling sources of energy, EnergyModels can model distortions in the light path such as the scattering characteristics of interstellar dust. EnergyModels also model the characteristics of the Observatory, all the way down to an individual detector.

At any point along the pipeline, the user can attach one or more visualizations. These visualizations graphically show the results of the pipeline as it might appear at that point in the pipeline. Figure 3 above shows a visualization at the end of the pipeline therefore it is displaying the results of what the detector would see after rendering the data. Currently, we only have a single imaging visualization. Our short-term plans include additional visualizations for spectroscopy.

Users are free to add, remove, and rearrange EnergyModels to suit their simulation needs. Visualizations can be attached anywhere along the pipeline. Typically, the user tells the pipeline to render, and then views the result. The user can then tweak the EnergyModel parameters and then render again. Comparing the results is often very informative and can help a user understand some of the nuances of the light path and of the particular simulated instrument.

4. HOW DOES THE SEA SIMULATION FACILITY APPLY TO THE VO?

A key to the success of the VO will be its effectiveness at integrating and applying its ability to access and analyze multiple and vast archives. The ability to link the archives back into the observation planning process should have a major impact on the value of the VO. The integration that SEA provides with its visual access to archival data is a key component of the full-circle data flow that we believe is essential for VO success.

Another key to the success of the VO is the ability of archives to be able to provide users with the "pedigree" of the data in the archives. The ability to characterize and determine the quality of any archival observation is extremely important if the image is to be used by other astronomers who are not the original observers. The prototype simulation pipeline for planning new observations is a key component that will help determine the data pedigree by comparing simulations with archival data. Aberrations showing up in the archival data could be modeled in the simulation pipeline. Once the aberrations are understood, the pedigree of the data can be used to determine the significance of scientific results from the data.

The SSF is working to combine the ability to access existing archives, with the ability to model and visualize observations. Integrating the two will allow astronomers to better use emerging integrated archives of the VO to plan and predict the success of potential new observations.

The SSF provides benefits to a variety of potential users.

Observers can use a simulation to:

- Effectively determine how various parameters affect their data and scientific objectives
- Act as a "Phase 0" tool for the initial "framing" of the observations
- Validate proposed observations ahead of time
- Support new complex instruments which drive the need for newer visualization tools

Observatory Staff can use simulation to:

- Characterize their telescope, instruments, and detectors
- Calibrate instruments with fewer observations

Archive Users can use simulation to:

- Understand the quality and limitations of an archival image

The prototype SSF is intended to be more of an integrating framework and a comparative tool than a highly realistic imaging simulator. Our focus as software engineers is on the framework and the visualizations. While we have developed some simple target and instrument models to provide a testing basis, we hope that this framework will allow instrument scientists and engineers to easily integrate more detailed, scientifically complex, models of different components of an observation's light path. We are working on defining the interface parameters to enable distributed models to plug into the SSF framework without having to fully re-write the models. *Our goals are to link models, not re-write them. As with the initial SEA prototype, we are striving for user interfaces that let astronomers focus on science rather than on technical programming issues.*

5. HOW DOES THIS HELP SOLVE OUR FULL CIRCLE DILEMMA?

SSF allows users to use existing archival data and research as the basis of exploring new ideas for observations, and then help refine the idea and establish its feasibility. Once the feasibility is established, SSF, as a component of SEA, can help turn the parameters of a simulation into a detailed proposal for submission to the appropriate observatory. Thus, we have tools that allow a scientist to move from ideas to observation to archival data and back intuitively and seamlessly.

For example, suppose the results of a VO data-mining study of galaxies provides ratios of galaxy types and their distribution in size, then this information (numerical values) can be easily plugged into the SSF to generate a region of the sky that closely reflects observations by using the SSF's Field Editor. The user can edit the parameters in the Field

Editor to further refine the spectral characteristics, surface brightnesses, redshifts, etc., of the objects in the field. The user can also add more objects in the field. In short, the user can interactively generate a target field. This target field can then be used to determine observational parameters, such as detector parameters, exposure times, etc., that would help a scientist achieve his/her scientific objectives. If the astronomer is using the SSF for an existing telescope, the simulated image can be compared with archived images. If the astronomer is planning to develop a new telescope then the SSF's pluggable design allows a user to put in specifications of the new telescope/detector to simulate the data. The SSF's Signal-to-Noise Analyzer can be used to study the predicted SNR at various parts of the image. If the observer is not satisfied with the results of the simulation, the various parameters can be interactively adjusted and data simulated once again. The "pièce de resistance" of the tools is that the SSF is part of the exposure toolset within SEA. Hence, the user's simulation can easily be turned into a set of exposure parameters for a new observing proposal.

We note that in the above example we used the results of a VO analysis to develop an observing proposal. However, the SSF can also be used in the other direction, i.e. an image of an hypothetical target can be simulated and this can then be used to search for archival images that have the same characteristics as the simulated image. This strategy is yet to be fully implemented in the SSF.

These are just scenarios where the SSF provides a link between new observation planning and VO's ability to data mine existing images.

6. SSF STATUS

The SSF is still a work in progress and many features have yet to be developed. The primary effort thus far has been in the design and implementation of the core framework, which includes the simulation pipeline and EnergyModel design. A basic simulation model framework has been completed which allows arbitrary simulation modules to be utilized. This framework includes models for point sources, extended sources, and fields of sources. We have developed initial classes for HST's WFPC2 instrument, and VLT's FORS1 instrument. These classes define relatively simple models of the instruments, their filters, and detectors. These models are suitable for demonstration purposes but will need to be refined before they are accurate enough for science planning. We are working with teams at STScI and at ESO to enable this. We are also actively seeking additional outside resources for simulation models, including other instruments, or other general-purpose models such as source models.

The user interface for the SSF is well developed and the framework is essentially complete. This includes a user-interface (UI) representation of the simulation pipeline that can be manipulated, plus a UI framework for managing visualizations. Currently only a single visualization type exists which displays a 2D image of the energy data set. A 3D visualization of spectroscopic data is currently being designed.

7. SSF ROAD MAP

The primary objective for future development of the SSF is to improve the fidelity of the simulation models and to incorporate additional models. At the detector level, a cosmic ray model has been developed but additional artifact models, including diffraction spikes and saturation, are planned. Additionally, improvements in the source models are planned, including support for absorption and redshift. We also plan to develop geometric source models to augment the usefulness of the SSF in instrument calibration.

As mentioned previously, a 3D visualization of spectroscopic data is currently under investigation. This visualization should provide the same level of interactivity and presentation for spectroscopic data that the Visual Target Tuner enables for 2D image data. We also plan to provide other visualizations, including a table summary of data statistics and a tool that produces false color images from 2D slices through the data set.

The above tasks will culminate in a release at the end of the fiscal year, September 2001. Assuming continued funding, plans for the following year include the ability to branch the pipeline into different output configurations, the ability to aggregate models together, and the ability to define models through XML descriptions. The goal of the latter would be to encourage domain experts to be able to easily define models without knowledge of the SSF framework or having to write Java code. Finally, we intend to publicize the SSF to other observatories by writing a "cookbook" which will clearly explain how to add new models and modify the pipeline for other observatories/instruments. Both this document and the SSF source code will be freely available to any organization that would like to use them.

8. CONCLUSIONS

Astronomers already intuitively think in full circles, ideas to observations to analysis to more ideas. In an era where the VO will unlock immense treasures by linking our astronomical archives together, we will have a critical need for a common set of visual and easy-to-use tools. These tools should support and enable the intuitive workings of astronomer's minds, not turn astronomers into computer or instrument scientists. Tools such as SSF bring the power to link the research, discoveries, and analyses that the VO makes possible with the formulation, refinement and validation of new observing ideas. Additionally, these tools must be common across computer hardware types and across observatories and wavelength specialties. Multi-use tools like SSF will increase the efficiency of all of our observatories as well as magnify the value of all existing archives and of the VO.

ACKNOWLEDGEMENTS

This work was initially supported by the Next Generation Space Telescope and is now funded by the Applied Information Systems Research grant NAG-9512 and SOMO Technology Development Program.