

AIRFrame: Integrating Diverse Digital Collections in Astrobiology

Rich Gazan

University of Hawaii Department of
Information and Computer Sciences
1680 East-West Road
Honolulu, HI 96822
(808) 956-7420
gazan@hawaii.edu

ABSTRACT

Astrobiology is an inherently interdisciplinary field concerned with questions of life in the universe. This paper describes the design and ongoing implementation of the Astrobiology Integrative Research Framework (AIRFrame), an open source, ontology-driven information system designed to ingest and analyze heterogeneous inputs of both published and unpublished data, and to identify and illustrate latent connections between research in astrobiology's diverse constituent fields.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – system issues, user issues.

General Terms

Management, Documentation, Design, Standardization.

Keywords

astrobiology, interdisciplinary science, collaboration

1. INTRODUCTION

In the domain of astrobiology, scientists from fields such as astronomy, geology, hydrology, molecular biology and many others work on the common question of life in the universe. To collaborate effectively, researchers must familiarize themselves with relevant information from disparate domains. The Astrobiology Integrative Research Framework (AIRFrame) [1] is a NASA Astrobiology Institute [4] project designed to help diverse researchers share information across disciplinary boundaries, identify canon and controversy in other domains, and to assess and refine their work against an appropriately broad disciplinary base.

2. FUNCTIONALITY

AIRFrame employs open source tools, while preserving the heterogeneity of each input source. Integrating diverse types of data requires that a core ontology—an underlying formal model for integrating source data—be created to enable the scalable assimilation of information from diverse sources [3]. AIRFrame exemplifies a core ontology, and is based on a theoretical framework of situated action [5], where the meaning of a document is not fixed upon creation, but can be changed with successive uses and associations. The information flowing to and

from AIRFrame includes published and unpublished astrobiology literature, both that collected by system developers and uploaded by researchers. Underlying AIRFrame is Textpresso [2], an open-source text mining and retrieval application for scientific literature. Initially designed for documents related to the Wormbase project [6], studying the biology and genome of *C. elegans*, the Textpresso architecture has proved sufficiently robust to have been applied to literatures across 17 different fields at this writing. Textpresso provides a full text search as well as concept and class categories to relate objects. For example, deuterium-hydrogen (D/H) ratio has an <observed-in> association with deep drill holes, chondritic meteorites and the interstellar medium, among others. Thus, literature in three traditionally separate areas of science can be linked not just through the common phrase D/H ratio, but in the shared sense of D/H ratio as an object of study.

3. CONCLUSION

AIRFrame provides astrobiology researchers a high-level framework for the productive exchange and discovery of information. Researchers who seek to answer the question of life in the universe need a way to understand how researchers in other disciplines have investigated the same tantalizing question.

4. ACKNOWLEDGMENTS

This material is based upon work supported by the National Aeronautics and Space Administration through the NASA Astrobiology Institute under Cooperative Agreement No. NNA08DA77A issued through the Office of Space Science.

5. REFERENCES

- [1] Astrobiology Integrative Research Framework. 2010. AIRFrame home page. <http://www.ifa.hawaii.edu/airframe/>.
- [2] California Institute of Technology. 2009. Textpresso home page. <http://www.textpresso.org/>.
- [3] Doerr, M., Hunter, J., and Lagoze, C. 2003. Towards a core ontology for information integration. *J. Dig. Info.* 4(1).
- [4] National Aeronautics and Space Administration. 2010. <http://astrobiology.nasa.gov/nai/about/>.
- [5] Suchman, L. A. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge, MA: Cambridge Press.
- [6] Wormbase. 2009. Wormbase: The Biology and Genome of *C. elegans*. <http://www.wormbase.org/>.