USING XML AND SEMANTIC TECHNOLOGIES IN ASTROINFORMATICS TO MANAGE DATA

Guy Beech
School of Computing and Engineering, University of Huddersfield, HD1 3DH

5th April 2018

ABSTRACT

The aim of this doctoral research project is to develop contemporary data management solutions for the management and utilization of astronomy data. One key objective is to define an XMLSchema and ontologies catering for the whole scope of AstroInformatics data, which are accepted for use by the astronomical research and industry sectors.

This will involve discussions with various astronomical bodies to ascertain their opinions and needs and also the creation of new custom software within which to apply the XMLSchema and Ontologies, one outcome of which is to be the demonstration of intelligent automatic tasking of telescopes and other data collectors.

Another key objective is to develop a set of tools that make use of the schema and ontologies and which will support researchers in capturing, storing, exchanging and exploring available astronomy data.

Index Terms - XMLSchema, standard, data processing, ontology, astronomy.

INTRODUCTION

Free access to astronomy data is now widespread. This is a vast amount of data held in many different data stores and in many different structures and formats. When querying and retrieving such data it would be of great benefit to be able to format this retrieved data through a single, widely accepted data schema.

Standardising the format in which to save and retrieve astronomical data and also enabling computer systems to recognise the meaning of the existing data currently saved, without requiring human interpretation, would be a significant help in extracting greater information at lower cost. (Wall 2014)

This research proposes that standardisation is achievable using an XMLSchema for data transformation before or after storage. Ontologies designated to areas within this schema would help automated systems to understand the meaning of other systems data by using semantic metadata.

SOFTWARE SYSTEM OVERVIEW

The central application, as shown in Appendix 1 at the end of this paper, contains the user interface, user tools, references the XMLSchema and the Ontologies. It carries out the user request for information of an astronomy item or object, queries the databases and constructs an output of the type chosen by the user (eg: PDF, HTML, XML… etc) by transforming the data structure using XSLT and ontologies, the output being presented as an XMLDocument which is in accordance with the XMLSchema rules.

Therefore all output data produced, whatever the datasources, is combined into a single output document of a standard structure.

THE LOGICAL MODEL

Of course astronomy is a very wide ranging and complex subject area.

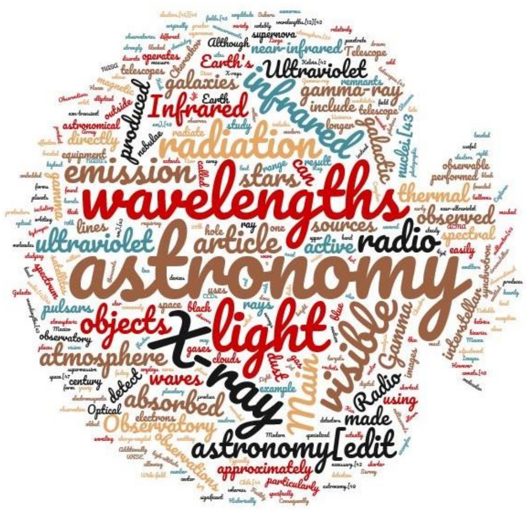


Diagram 1: A few Astronomy terms

For this reason the complexity needs to be broken down into less complex sub or child areas and each area needs to be able to understand the data held in other areas. This can be done by designing an ontology for any given XMLSchema area.

The XMLSchema itself needs to be designed in such a way that allows the schema to evolve over time without compromising its structure.

The skills required are not only in designing the XMLSchema but also to be able to decide on the scope of any given ontology within this XMLSchema.

A very simple example of the idea is shown below in Diagram 2:

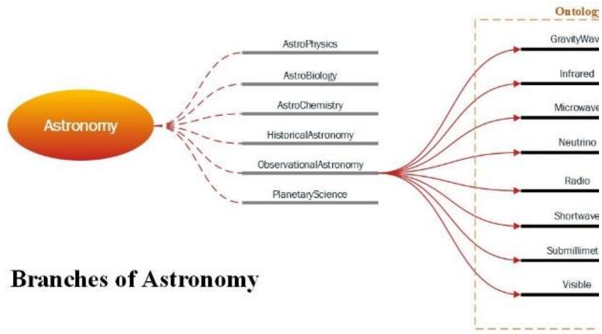


Diagram 2: Ontology Placement

Evidently, reality would be much more complicated than this diagam but in this instance all observational data is associated with a single ontology, leaving other areas to have different ontologies. Clearly here, although observational data is linked within a single ontology, there will need to be crossover of understanding between other ontologies to enable meaning to be carried to and from multiple ontologies in other areas of the XMLSchema.

IMPLEMENTATION OF THE PHYSICAL MODEL

During previous Masters' research a small area of XMLSchema was developed and software created to query astronomy repositories. The data retrieved from multiple data repositories was then stored in XMLDocuments in the format described by the XMLSchema before being transformed into different types of readable output. (Beech 2015) This of course was just a proof of concept, not a final structure in itself and is shown in Table1 and Figure1 below from that research.

Astronomy (root)
Astrophysics
Astrobiology
Astrochemistry
Historical Astronomy
Planetary Science
Observational Astronomy
Visible
Infrared
Radio
Microwave
Gravity wave
Shortwave
Neutrino
Submillimetre

Table 1: Astronomical Schema high level nodes

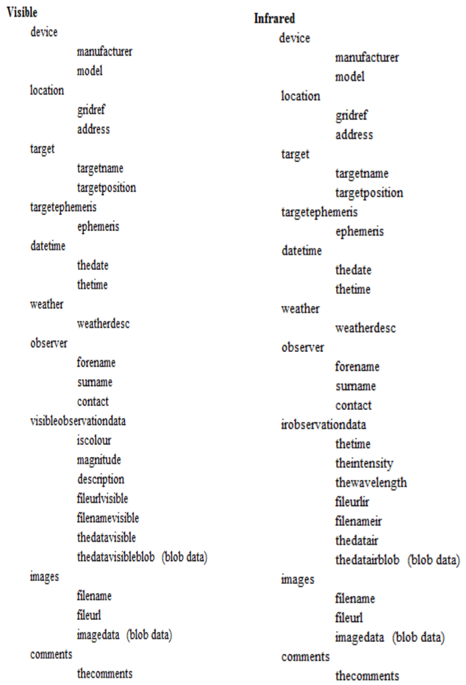


Fig 1: Detail of Visible and Infra-red nodes

It was only after the completion of the Masters research that the usefulness of semantic data was looked at and subsequently incorporated as part of this current doctoral research.

So, any ontologies are yet to be developed and then referenced by the XMLSchema in clearly described areas.

DEVELOPMENT OF DATA TOOLS

Once a common XMLSchema is under development and ontologies are being constructed then tools need to be developed. Examples of possible tools to query multiple data sources that are within an application guided by the XMLSchema are:

- Automated production of documents containing human readable data (PDF, Word etc) constructed from multiple data repositories
- Automated production of output documents that are machine to machine compatible
- Intelligent detection of inadequate or missing data resulting in the 'intelligent' auto tasking of telescopes to collect such data

There are sure to be many more tools that would be useful to astronomy professionals when considering data retrieval and interpretation.

FUTURE DEVELOPMENTS

Initial **practical and useful** development of the extensible schema and then the planning out of ontology areas are required. For this an initial schema area needs to be developed and ontologies need to be written. This can be extended and added to over time.

Software data tools require developing which the astronomy community would like to have!

CONCLUSIONS

There is of course much work to be done. Can I do this on my own? No, this is a major undertaking and it is too big a job for one person. I need the help and support of the astronomy community to bring to fruition this concept of automatic interpretation of open, linked data across the whole spectrum of Astronomy data storage.

I would be grateful for assistance from all and any who are interested in this approach to interpreting astronomy data, including:
The International Virtual Observatory Alliance (IVOA 2017).
The UK Solar System Data Centre (Space 2017).
W3c (W3cStandards 2017).
The Astrostatistics and Astroinformatics Portal (Feigelson 2017).

Let us work together to make this a reality.

CONTACT DETAILS

Address:
Guy Beech MScRes, BSc(Hons), FRAS
Collaborative & Research Systems Senior IT Officer
Corporate Information Systems
Computing and Library Services
University of Huddersfield Queensgate
Huddersfield
HD1 3DH

g.beech@hud.ac.uk
Tel: 01484 472032

REFERENCES

Beech, G. (2015). An Investigation of the Benefit of XML Technologies in Astronomical Data Interpretation. School of Computing and Engineering. Huddersfield University, West Yorkshire, UK

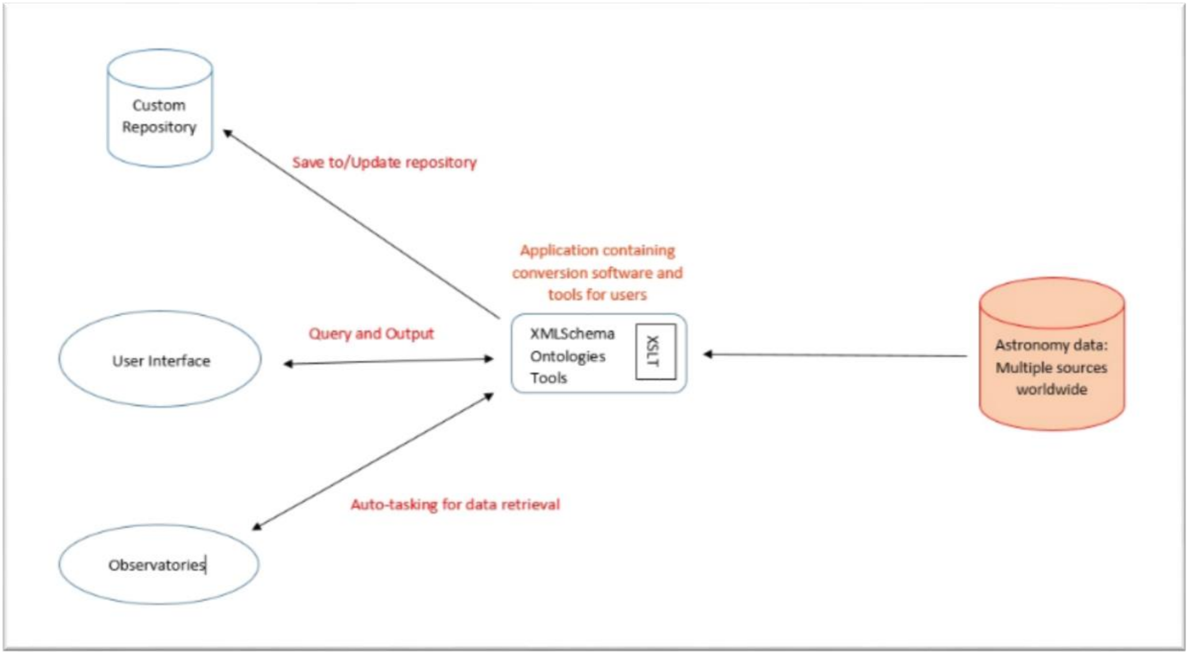
Feigelson, E. (2017). "Astrostatistics and Astroinformatics Portal." Retrieved 17/11/2017, 2017, from https://asaip.psu.edu/.

IVOA (2017). "International Virtual Observatory Alliance." Retrieved 01/12/2017, 2017, from http://ivoa.net/.

Space, R. (2017). "UK Solar System Data Centre (UKSSDC)." from https://www.ukssdc.ac.uk/.

W3cStandards (2017). All Standards and Drafts. W3c Standards.

Wall, M. (2014). Astronomy Overload. Space.com.



Appendix 1: Software System Overview