A Domain Ontology for CoRoT Data Analysis

Lucélia de Souza (UNICENTRO, UFPR)
Dra. Maria Salete Marcon Gomes Vaz (UEPG, UFPR)
Dr. Marcelo Emílio (UEPG)
Dr. José Carlos Ferreira da Rocha (UEPG)

24 de outubro de 2012

São Paulo Semantics Session IVOA Meeting

Summary

CoRoT Data Analysis - Exoplanet Field

In CoRoT data analysis process, some exoplanet search algorithms requires detrend algorithms as a pre-processing before running. Different detrend and transits detect algorithms can be applied to the same database.

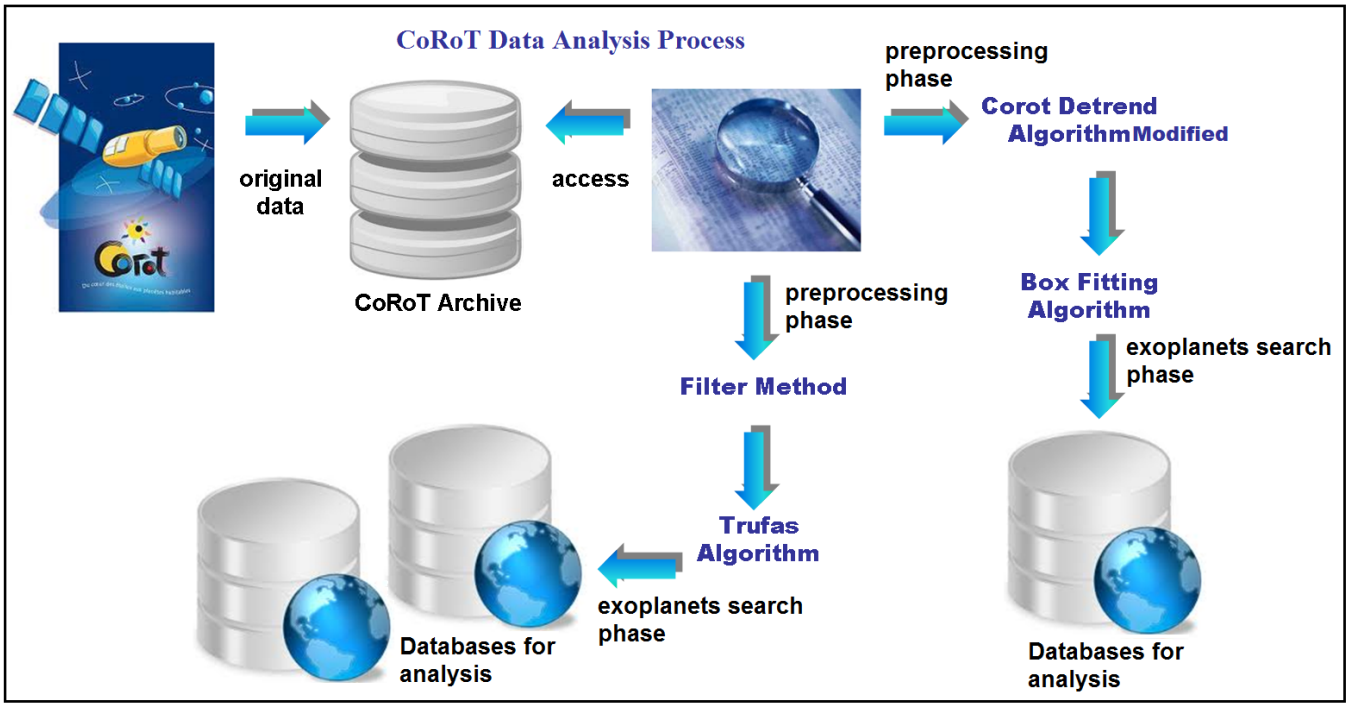

Fig. 1: Corot Data Analysis [The Authors].

Flexible Image Tranport System (FITS)

MANDATORY FITS KEYWORDS					
ALL HDUs	Array HDUs *	Conforming Extension	ASCII Table Extension	Binary Table Extension	Random Groups Records
DATE	BSCALE	EXT	TSCALn	TSCAKn	PTYPEn
ORIGIN	BZERO	EXTVER	TZEROn	TZEROn	PSCALn
BLOCKED **	BUNIT	EXTLEVEL	TNULLn	TNULLn	PZEROn
AUTHOR	BLANK		TTYPEn	TTYPEn	
REFERENC	CTYPEn		TUNITn	TUNITn	
COMMENT	CRPIXn			TDISPn	
HISTORY	CROTAn			TDIMn	
DATE-OBS	CRVALn			THEAP	
TELESCOP	CDELTn				
INSTRUME	DATAMAX				
OBSERVER	DATMIN				
OBJECT					
EQUINOX		**HISTORY keyword [ascii text char...]**			
EPOCH		(Hanisch et al, 2001)			
* Primary HDU					
** Deprecated					

Fig. 2: Mandatory FITS Keywords.

Metadata storing in images header or in the databases is insufficient, semantically, to generate provenance. This information is useful for local researchers, but not enough to share, reuse and reprocessing by scientific community.

Corot Data Analysis - Exoplanet Field

PreProcessing Phase

Data obtained from satellites need corrections, such as jumps and trends (detrend).

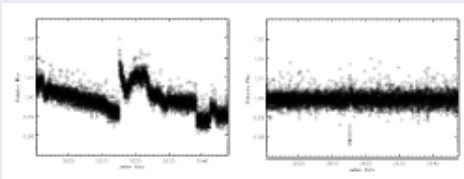

Fig. 3: Trends and jumps in light curves [Mislis et al 2010].

Analysis Phase

The analysis is dependent of context. In exoplanets search, is done the transits analysis.

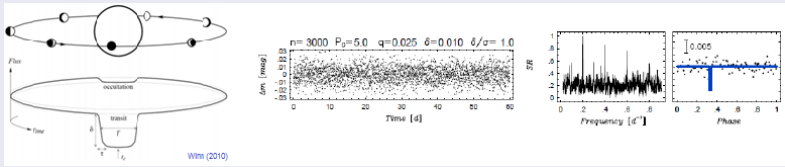

Fig. 4: Transits Analysis [Kovács et al 2002].

Objective

In CoRoT data analysis, tracking the origin of information and how the data were derived in each level in the data analysis process is essential to allowing sharing, reuse, reprocessing and further analysis.

The objective of this work is the development of a formalized and codified knowledge model by means of domain ontology as a way of ensuring semantic provenance information for retrieval later by humans or software components.

Ontologies and Semantic Provenance

An Ontology is a explicit specification of a conceptualization [Gruber 1993].

Borst (1997) redefines it for *Ontology is a formal specification of a shared conceptualization*.

An ontology is a relationship of four elements [Kiryakov 2006]:

$$O = \{\textbf{C}\text{lasses}, \textbf{R}\text{elations}, \textbf{I}\text{nstances and } \textbf{A}\text{xioms}\}$$

Provenance means origin or source. The semantic provenance [Sahoo et al 2008] relates with the Semantic Web [Berners-Lee 2001] and can be obtained by means of ontologies, which allow to represent the knowledge, structuring the informations of organized form and generate semantics in the data.

Development of a Domain Ontology

CoRoT Data Analysis Ontology

Methodological Approach

Methodology

- Ontology Development 101 [Noy and McGuinness 2001]

Tool

- Protégé 4.1 Tool [Knublauch 2004]

Reasoner

- Pellet 2.2 free Reasoner [Sirin et al 2007]

Plugin

- Ontograf Plugin [Falconer 2010]

Languages

- Ontology Web Language [Hitzler et al 2009] based on Description Logic [Baader 2003] (OWL-DL 2.0)
- SPARQL 1.1 Query Language [DuCharme 2011]
- Semantic Web Rule Language - SWRL (*DL-Safe Rules*) [Horrocks et al 2005]

Reuse

- Reuse of parts VSTO and SWEET Ontologies http://escience.rpi.edu/ontology/vsto/2/0/vsto.owl and http://sweet.jpl.nasa.gov/2.3/sweetAll.owl

Conventions of Nomenclature of Proposed Ontology

The nomenclature used is obtained from a review of literature as the concepts used envolving the context related. The mostly nomenclature conventions are based on [Cambrésy et al 2010] with some adaptations and extensions.

- Classes are defined begin with uppercase without spaces, e.g. *Algorithm*.
- Instantiation is defined using underscore when necessary, e.g. *trends_long_term*.
- Properties are defined begin with lowercase letter following by uppercase initial letters, e.g. *isDetrendOf*.
- Acronyms are used carefully, e.g. *CDA*.

Competence Questions

- *Which detrend algorithms are Polynomials?*
- *What are the statistical techniques (Linear, Polynomial, among others) used by detrend algorithms?*
- *The CoRoT Detrend Algorithm treats which systematic effect?*
- *What transit algorithm apply the Least-Squares method?*
- *Which shape is used by transit detection algorithms to model the periodic signal (Box, Fourier Series, Wavelet Transformations, among others)*
- *The light curves generated were developed from which detrend algorithm and by which statistical technique?*

Corot Data Analysis Ontology

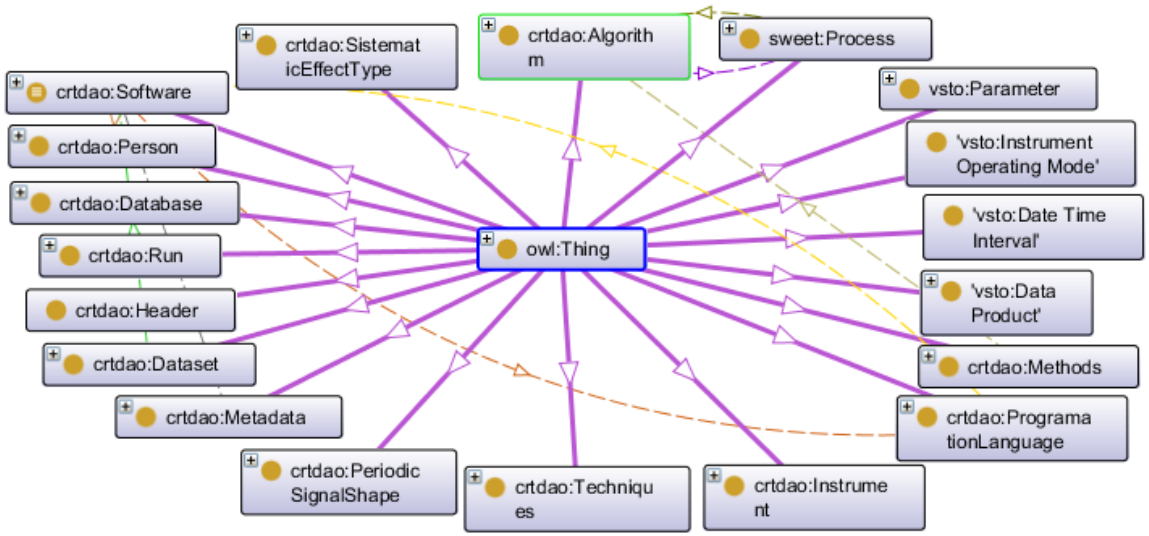

Fig. 5: Corot Data Analysis Ontology - Classes Diagram.

Detrend Algorithms

What are the names and acronyms of the detrend algorithms and their respective authors?

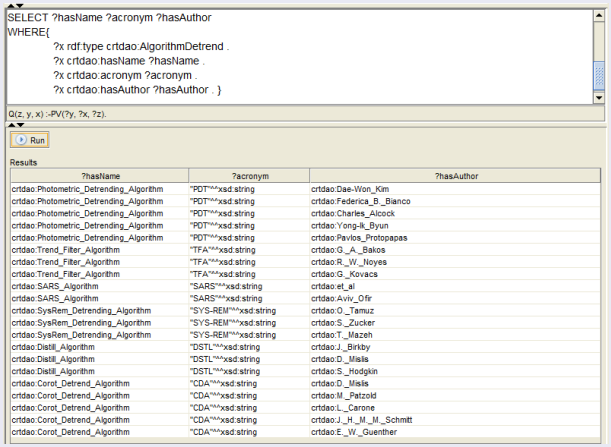

Fig. 6: Query on Detrend Algorithms

Transit Algorithms

What is the format of the periodic signal of transit algorithms?

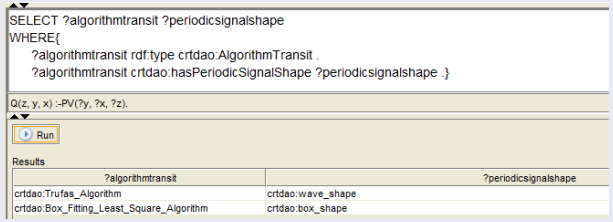

Fig. 7: Query on Transit Algorithms

SPARQL-DL Query Language - Cont.

Detrend Algorithms

What detrend algorithm is polynomial type?

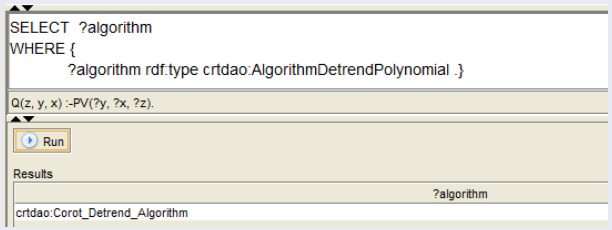

Fig. 8: Query on Detrend Algorithms type

Transit and Detrend Software

What detrend software was applied to transit software?

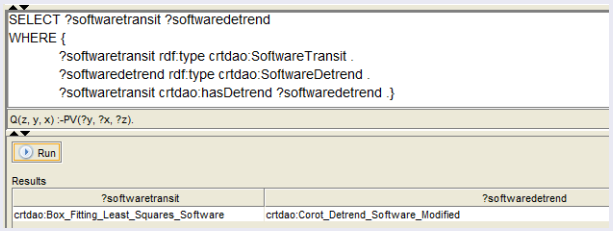

Fig. 9: Query on Transit and Detrend Software

Use of Rules

Rule for Detrend Algorithms

What detrend algorithm type use a polynomial function?

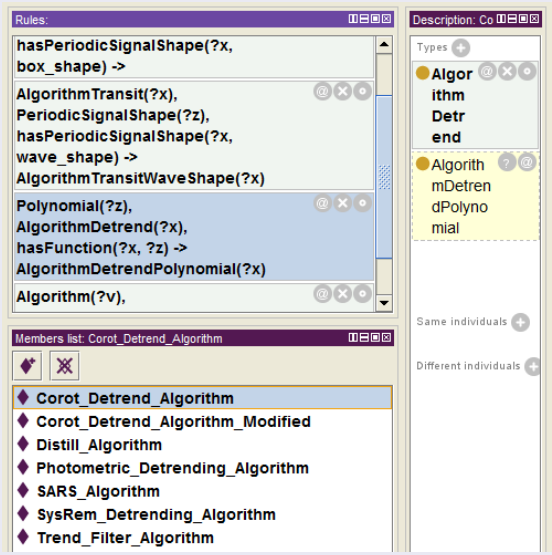

Fig. 10: Rule used for AlgorithmDetrendPolynomial

Rule for Transit Algorithms

What transit algorithm type have the periodic signal as box_shape?

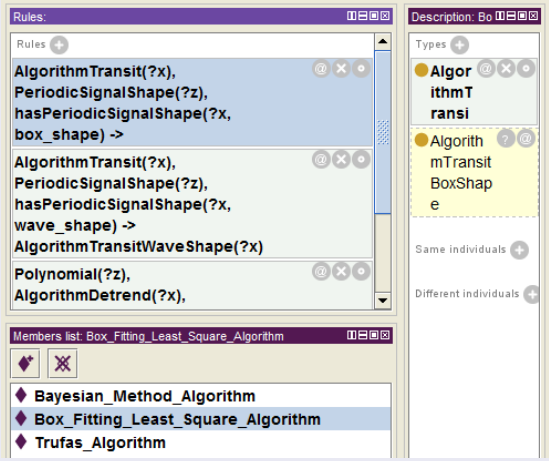

Fig. 11: Rule used for AlgorithmTransitBoxShape

PROOF MARKUP LANGUAGE - PML 2.0

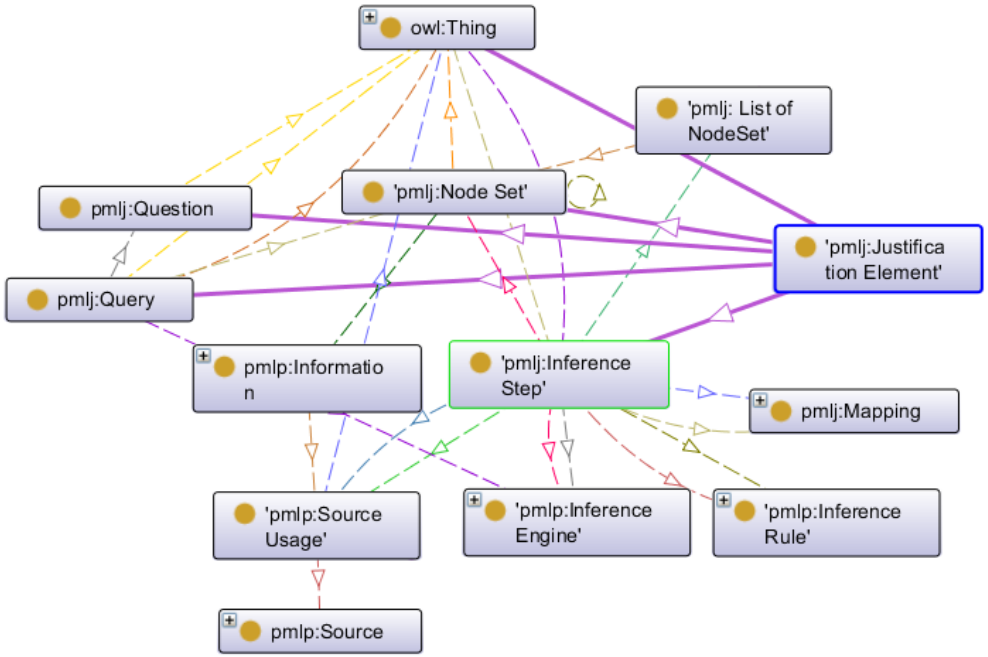

Fig. 12: Provenance and Justification Ontology [McGuinness 2007]

Domain Ontology and Model Provenance

Which the periodic signal shape of Box Fitting Least Square Algorithm?

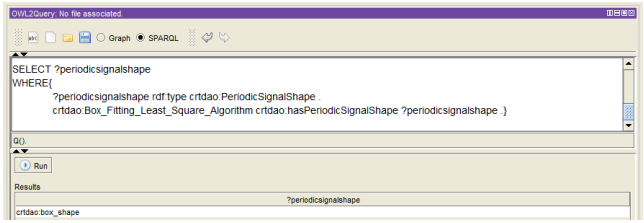

Fig. 13: Query on Domain Ontology

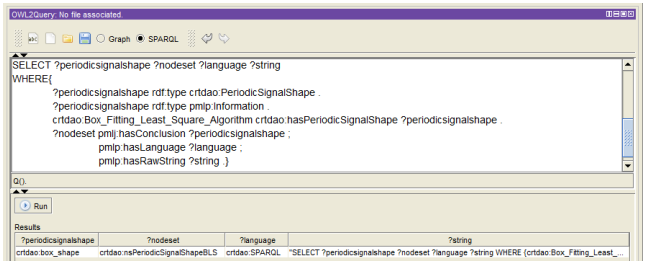

Fig. 14: Query on Domain and Provenance Ontologies

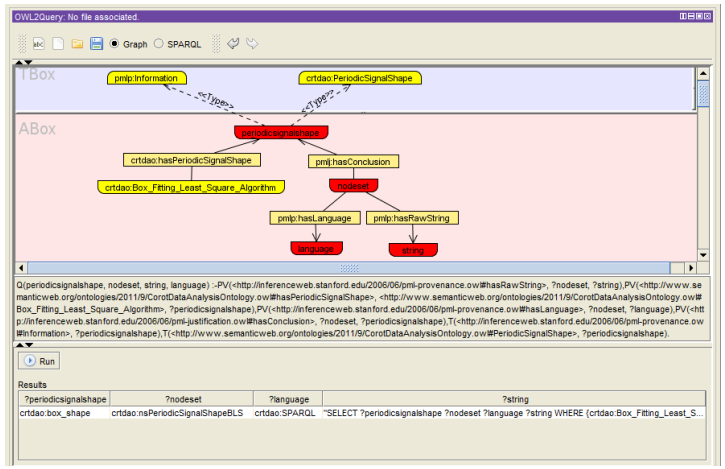

Fig. 15: Graph using the Domain and Provenance Ontologies

Conclusions and Future Works

- It is presented that is possible to generate semantic provenance in scientific images analysis.
- The semantic provenance information obtained will be integrated in a web framework for analysis.
- Despite the scope of this work, results can be expanded to fields of e-Science where the scientific images analysis requires preprocessing, adding semantic knowledge and allowing interoperability.
- How future works, the ontology will be expanded for modelling more algorithms related with the CoRoT data analysis (among others telescopes), being validated continuously.
- Examine the possibility of reuse of ontology for astronomical objects to verify which methods were used to detect exoplanets in stars of a certain type.

References

Baader, Franz and Calvanese, Diego and McGuinness, Deborah L. and Nardi, Daniele and Patel-Schneider, Peter F. The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press. New York, NY, USA. 2003.

Berners-Lee T., Hendler, J. and Lassila, Ora. The semantic web. Scientific American, 284(5):34–43, maio de 2001.

Borst, W. N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Doctoral Thesis. University of Tweenty, Centre for Telematica and Information Technology, Enschede, The Netherlands, 1997, (227 pp).

Cambrésy, L. and Derriere, S. and Padovani, P. and Preite Martinez A. and Richard, A. Ontology of Astronomical Object Types. IVOA Technical Note. Version 1.3, Jan de 2010.

De Souza, L. and Vaz, M. S. M. G. and Emílio, M. and da Rocha, J. C. F. and Boufleur, R. (2011). Data Analysis Provenance: Use Case for Exoplanet Search in CoRoT Database. Presented In: Proceedings of the Astronomical Data Analysis Software and Systems (ADASS XXI). (Vol. TBD). San Francisco: ASP Conference Series 2012.

DuCharme, B. Learning SPARQL. Oreilly and Associate Series. O'Reilly Media, 2011.

Falconer, Sean. Plugin Ontograf. 2010. URL http://protegewiki.stanford.edu/wiki/OntoGraf.

Gruber, Thomas R. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199–220, 1993.

Hanisch, R. J. and Farris, A. and Greisen, E. W. and Pence, W. D. and Schlesinger, B. M. and Teuben, P. J. and Thompson, R. W. and Warnock III, A. Definition of the Flexible Image Transport System (FITS). A & A. Vol. 376, 2001. (pp. 359–380).

Hitzler, P. and Krtzsch, M. and Rudolph, S. Semantic Web Modelling Languages (Part 2). Tutorial. IJCAI. 2009.

Horrocks, I. and Patel-Schneider, P. F. and Bechhofer, S. OWL Rules: A Proposal and Prototype Implementation. Journal of Web Semantics, V. 3, N. 1. 2005.

References - Cont.

Kiryakov, A. Semantic Web Technologies: trends and research in ontology-based systems. Ontologies for knowledge management. Davies, J. et al. (Eds), 2006. (pp 115-138).

Knublauch, H. and Fergerson, R. W. and Noy, N. F. and Musen, M. A. The Protégé OWL Plugin: An open development environment for semantic web applications. 3rd ISWC 2004, Hiroshima, Japan. (pp. 229–243).

Kovács, G., Zucker, S. and Mazeh, T. A Box-Fitting Algorithm in the Search for Periodic Transits. A&A 391:369-377, 2002.

McGuinness, D. L. and Ding, L. and da Silva, P. P. and Chang, C. (2007). PML 2: A Modular Explanation Interlingua. in Proceedings of the AAAI 2007 Workshop on Explanation-aware Computing (pp. 22–23).

Mislis, D. and Schmitt, J. H. M. M. and Carone, L. and Guenther, E. W. and Patzold, M. An Algorithm for Correcting CoRoT raw light curves. arXiv/1008.0300, 2010.

Noy, N. F. and McGuinness, D. L. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001- 0880. (25 pp.).

Boufleur, R. C. A Busca de Exoplanetas com as curvas de luz do CoRoT. Observatório Nacional. Departamento de Pós-Graduação. Curso de Pós-graduação em Astronomia. Dissertação de Mestrado. Rio de Janeiro, 2012.

Sahoo, S. S. and Sheth, A. and Henson, C. Semantic Provenance for eScience. IEEE Computer Society, 2008, (pp. 46-54).

Sirin, E. and Parsia, B. and Grau, B. C. and Kalyanpur, A. and Katz, Y. Pellet: A practical OWL-DL Reasoner. Web Semant. Volume 5. Number 2. Elsevier Science Publishers B. V. The Netherlands, June, 2007.

Questions?

Thanks by oportunity of participate of Semantics Session IVOA Meeting!