

Data Analysis Provenance: Use Case for Exoplanet Search in CoRoT Database

Lucélia de Souza^{1,2}, Maria Salete Marcon Gomes Vaz^{3,2}, Marcelo Emilio³, José Carlos Ferreira da Rocha³, Eduardo Janot Pacheco⁴ and Rodrigo Carlos Boufleur⁵

¹UNICENTRO, Rua S. Renna, 875, Guarapuava/PR BRAZIL 85.015-430

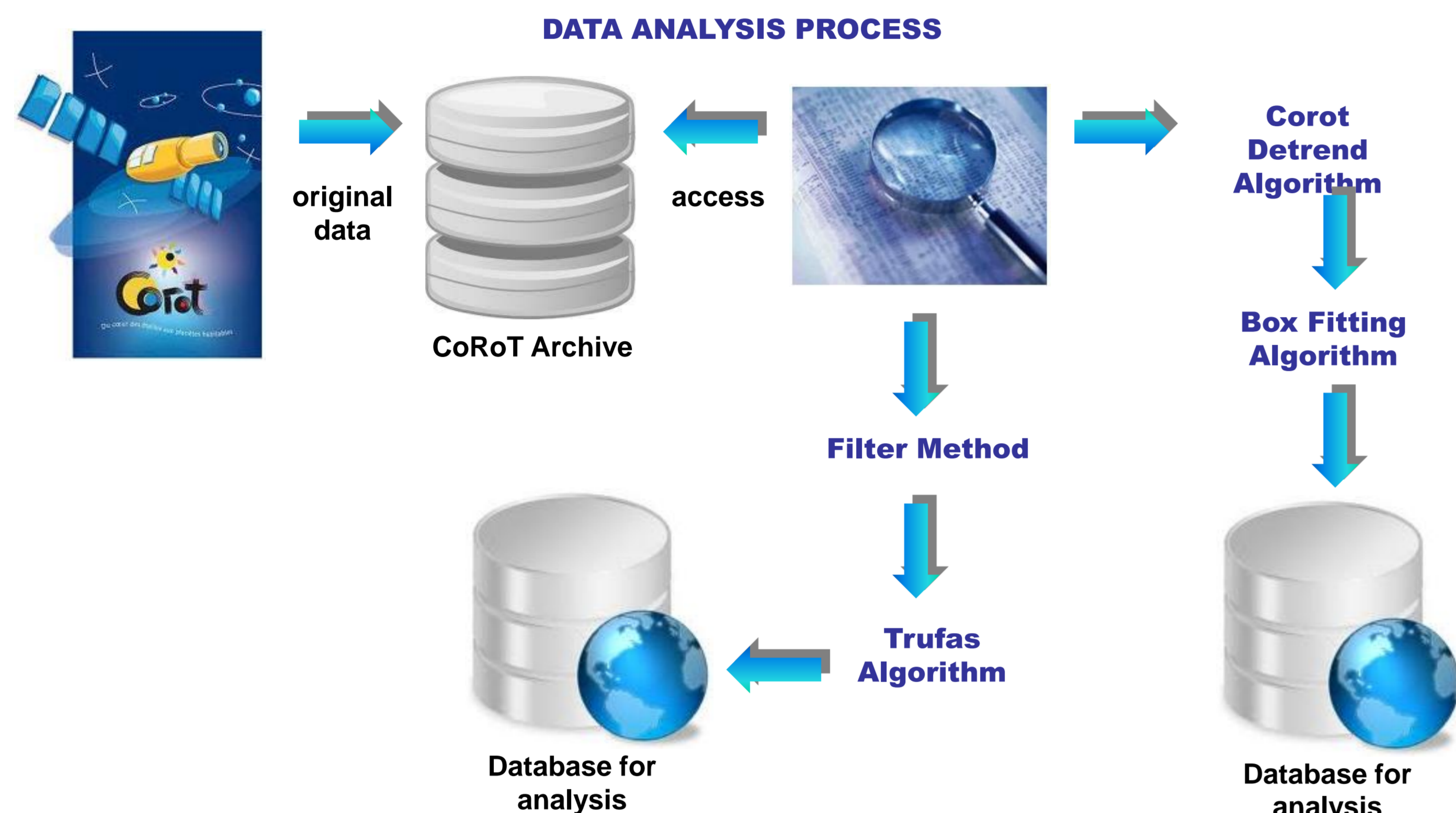
²UFPR, Rua Cel. F. H. dos Santos, 100, Curitiba/PR BRAZIL 81.531-980

³UEPG, Rua Carlos Cavalcanti, 4748, Ponta Grossa/PR BRAZIL 84.030-900

⁴IAG/USP, Rua do Matão, 1226, São Paulo/SP BRAZIL 05.508-090

⁵ON, Rua Gen. José Cristino, 77, Rio de Janeiro/RJ BRAZIL 20.921-400

INTRODUCTION



DATA ANALYSIS PROVENANCE

Provenance means origin or source.

- Levels of granularity
 - Coarse and Fine grain
- Data Provenance
 - Annotation or Non-Annotation
- Exoplanet Detecting Algorithms
 - provenance information records steps performed and generate knowledge in order to avoid reprocessing and contribute to sharing, reuse and analysis further

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					
EPOCH					

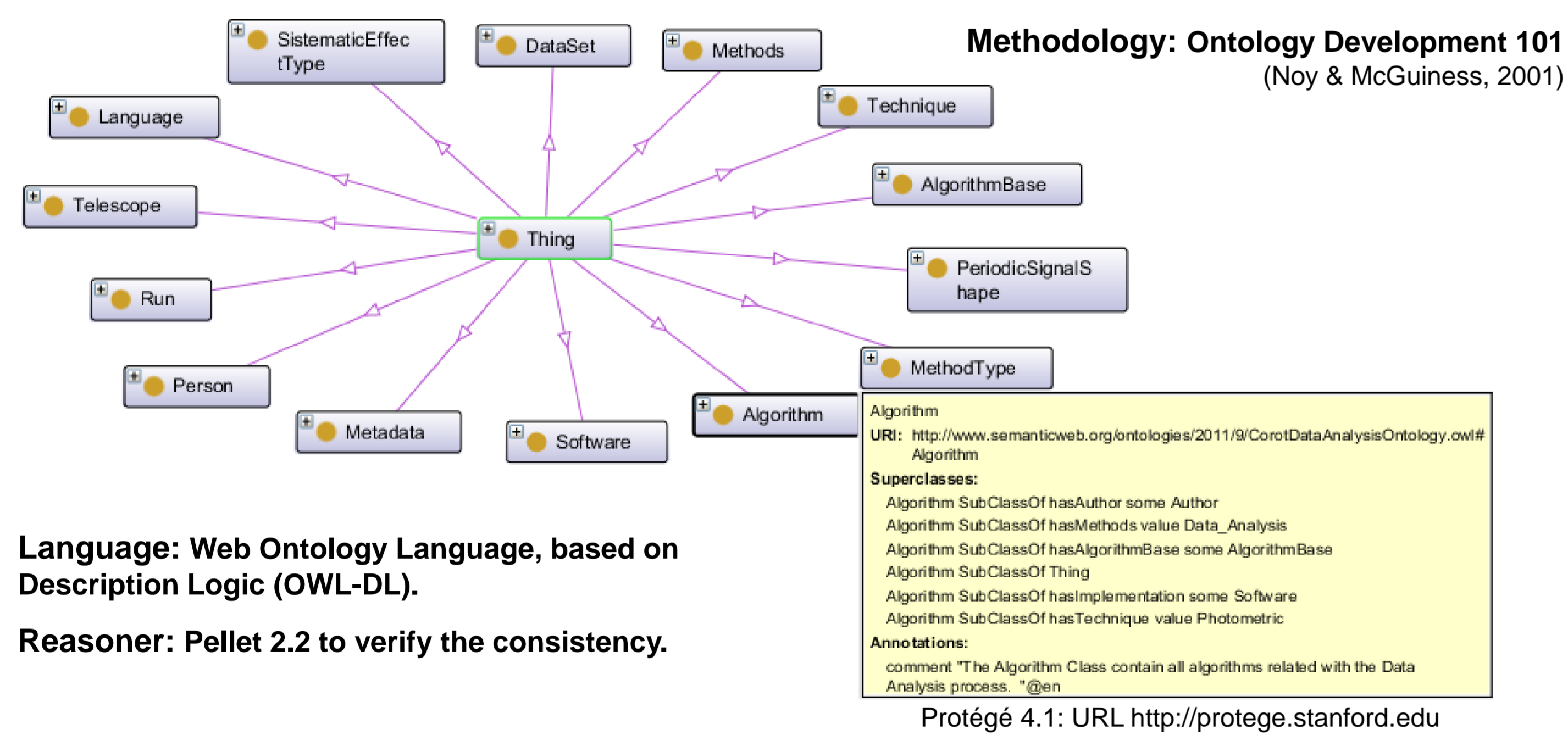
HISTORY keyword [ascii text char...]

* Primary HDU
** Deprecated

(Hanisch et al, 2001)

USE CASE FOR EXOPLANET SEARCH IN COROT DATABASE

DOMAIN ONTOLOGY



DESCRIPTION LOGIC QUERY IN THE DOMAIN ONTOLOGY

- Which detrending algorithm use the type of method Quadratic Programming?

Protégé 4.1: URL <http://protege.stanford.edu>

INFERENCE OF PELLET OWL REASONER

Pellet: <http://clarkparsia.com/pellet/>
Protégé 4.1: URL <http://protege.stanford.edu>

CONCLUSIONS AND FUTURE WORKS

- Provenance information add data semantics.
- Validation of the domain ontology by domain experts and ontologists.
- Integration by agents environment.
- Extension to fields of science in the data analysis process to facilitate reuse of data and enable its reproducibility.

Acknowledgments

- INCT Estudos do Espaço.
- CNPq Proc. 303873/2010-8.
- This work was conducted using the Protégé resource, which is supported by grant LM007885 from the United States National Library of Medicine.