Apr 6th, 12:30 PM

A Case Study in Astronomy Data Management

Megan Potterbusch
Simmons College, School of Library & Information Science, megan.potter@simmons.edu

A Case Study in Astronomy Data Management

Megan Potterbusch pottem@simmons.edu Simmons College

Introduction

By using the data practices of a particular astronomy research team using the radial velocity method to discover exoplanets as a guiding case example, this poster demonstrates data management practices for multi-team, research collaborations in the field of astronomy.

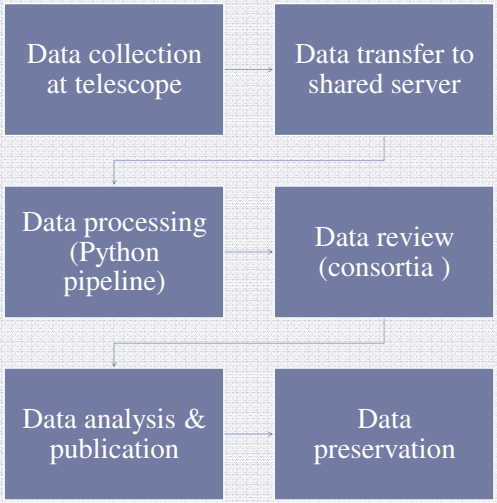
Irénée du Pont Telescope at Las Campanas Observatory
Credit: Krysztof Ulaczylk CC-By-SA

Analysis

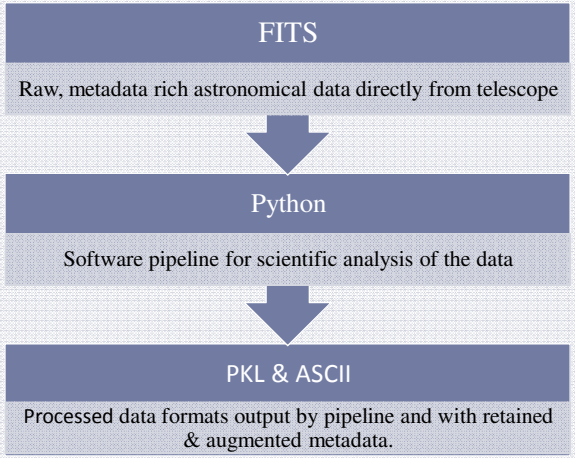
- Telescopes provide metadata maintained throughout analysis
- No data dictionary just conventions within field (units noted)
- Secured data storage in multiple locations
- Version control system for software
  - Versions associated with final datasets may not be noted
- Copyright important for software
- Data will become Open Access upon project completion
- No definite plan for long term data access
- Processed data may be submitted to publishers

Results

Partial Data Lifecycle



File Format Progression

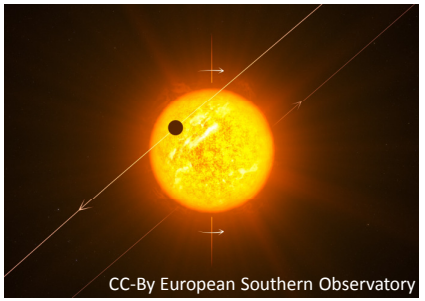


Astronomy Data Management

Research Products	Format	Metadata	Storage, Backup, Security
Observational	Raw: FITS format	Standardized. Generated by telescope. Human and machine readable.	Multiple locations. Institutionally managed systems with external back-ups.
Spectroscopic & Photometric	Processed: .PKL or CSV	Standardized. Generated by Software pipeline.	Multiple locations. Professionally managed systems. Public repositories linked to by publishers.
Code/software	Python	README file (stored with code) is required and comments in code are suggested.	Version control system. Releases created and preserved for each cleaned resultant dataset.
Visualizations	Dynamically generated from code or PDF for publication	Caption included with publication.	Stored within publications.

Methods

- 60 minute interview with local primary investigator
- Transcription of interview analyzed for strengths and areas for improvement.
- Tools used:
  - Class discussion
  - Lectures
  - Readings
  - Toolkits provided by instructors and guests



CC-By European Southern Observatory

Conclusion

In many ways, the astronomy field is exceptional in terms of data and metadata management; however, challenges still arise when dealing with newer technology. Best practices for management and preservation of programmed algorithms, such as the Python pipeline, continue to develop. The perception of infinite digital storage capacity can lead to poor data curation practices. Overall, the specificity of the astronomy discipline benefits from well-established domain-based practices.

Acknowledgements

Mercedes Lopez-Morales, Elaine R. Martin, Regina Fisher Raboin, Julie Goldman
Simmons LIS 532G, Fall 2015