**This document is the step-by-step set of instructions for the consulting recommendations.**

In response to our discussion regarding your data management practices, we have prepared the following document to outline the primary data management issues and a series of solutions (in phases) that will hopefully help you address these issues. We would be glad to work through implementing these solutions with you and/or your designated data manager.

In addition to these recommendations, we will also use the information collected in the interview as we develop policy and infrastructure recommendations for overall University research data management. Your contributions are most appreciated.

**Primary Data Management Issues** (From Excel Worksheet)

|  |  |  |
| --- | --- | --- |
|  | **Average** |  |
| File Formats Data Types | 2.666666667 | Satisfactory |
| Organization of Files | 2.5 | Satisfactory |
| Security Storage Backups | 1.333333333 | Fair |
| Copyright Privacy Confidentiality | 2 | Satisfactory |
| Data Documentation Metadata | 0.222222222 | Least Sustainable |
| **Cumulative** | 1.744444444 | Fair |

Phase 3

Phase 2

Phase 1

**Phase 1 (short-term)**

* **Develop a Data Management Plan (DMP):** a DMP is the basis of all data management, and is a critical tool in protecting the continuity of your research process. Once in place, it can continually be updated, provided to new members of the lab as guidelines, and easily be applied to future grant proposals. We will work with you to develop an appropriate plan for managing your data. This is a fundamental first step in improving process.
* Read Uva's "Laboratory Notebook and Recordkeeping" Policy: <https://policy.itc.virginia.edu/policy/policydisplay?id='RES-002>
* Properly document data sources used (Pull from FITS header)
* Properly document the context of the data collection (Pull from FITS header)
* Properly document all data collection methods (Pull from FITS header)
* Document all variable names and descriptions used in research data (May not apply)
* Understand and research appropriate metadata standards (Check Virtual Observatory standards)
* Document structure and organization of data files (You have good standards in place now they just need to be documented.)
* Document data validation and quality assurance processes (Log of reductions)
* Document and explain any codes and classification schemes used (N/A?)
* Document any file formats and software (including version) used (Version of IRAF/IDL software)
* Document information regarding confidentiality, access and use conditions (Check to see if proprietary period is included in the header)

**Phase 2 (long-term)**

* Use standard representation file formats such as ASCII or Unicode (N/A? as it is used in rare cases)
* Informal DMP has been improved to include these 8 categories: File Formats and Data Types, Organizing Files, Security/Storage/Backups, Funding Guidelines, Copyright & Privacy/Confidentiality, Data Documentation & Metadata, Archiving & Sharing Data, and Citing Data.
* Make the original document (DATA) "read only" then use a copy for analysis (You are already doing this!)
* Begin applying appropriate metadata standard to data and other research materials
* These are all related:
  + Use file version control software to assist with versioning issues!
  + Document the analysis process from raw data to "finished" analyzed data!
  + Document any algorithms used to transform data!

**Phase 3 (future)**

* Work with ITC to be sure that all servers and computers containing research data are being backed up to their specifications (with regular cycles and off-site storage)
* Have SciDaC review DMP
* Use log files to record every change made to a file no matter how small
* Use appropriate environmental conditions for storage media
* Store data storage media and servers in a physically secure environment
* Have data backups formally administered by ITC or some other system administrator (You already have a system administrator doing your backups.)
* Deposit your data in a subject repository (Possibly Virtual Observatory)
* Use ITC's Hierarchical Storage Management (HSM): <http://itc.virginia.edu/datastorage/hsm.html> (Concerns about this being a fixed cost, what to do when grant ($$$) ends?)
* Store research data in Cloud Storage.
* Have 3 copies of data: 1 local, 1 on local server and 1 on a remote server
* DMP is being followed by all members of the research team
* Fully apply the appropriate research community metadata standard to research data and materials.
* Additional suggestions:
  + Have data server reside on ITS’s “More Secure Network”
  + Have data server regularly scanned by ITS for security vulnerabilities