

# Documentating Solar Data

## aka. Minimum Information for Solar Observations

For years the Virtual Solar Observatory has received questions from data producers about how they should prepare their FITS files ... and we replied with something like *'just do whatever makes sense for you, and we'll figure out how to deal with it'*. We've since come to realize that this doesn't actually help data producers in the long term.

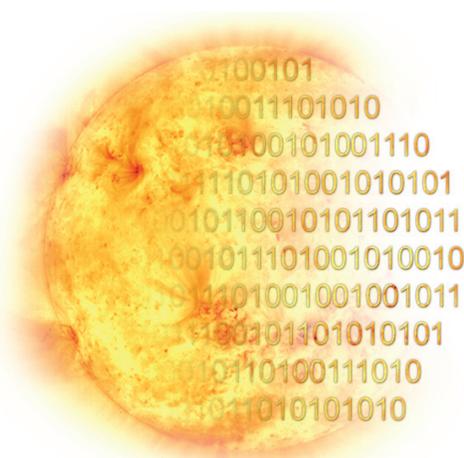
Because so many experiments are highly specialized, producers of data may not understand how non-discipline scientists may wish to use their data. As such, they don't know what metadata may be useful for other scientists to find and use the data they are providing, and may not include it in their FITS files or supporting documentation.

After our review of metadata for the 2011 AGU meeting, we were asked to create a template showing best practices. We presented some recommendations at the 2012 AAS/SPD meeting, but it really didn't have the flexibility that might be needed for highly specialized experiments. As such, we're attempting to take a step back and look at what questions your data or its supporting documentation should be able to answer.

### Our goal is to:

- Ensure that current and future **researchers can use your data**.
- Ensure that researchers will properly **acknowledge your data**.
- **Reduce** the amount of **time needed to support researchers**.
- **Reduce** the likelihood of **data or metadata being misunderstood**.
- **Reduce** the chance of **improper use of the data**.
- **Reduce** the amount of **effort needed to use the data**.

We believe that both solar physicists and non-discipline scientists should be able to easily understand what is in a data file from an instrument that they have never dealt with before, and quickly determine if it is useful for their purposes.



Virtual Solar Observatory

***Well thought out documentation, organization, file naming and metadata (FITS headers) will make a difference.***

Our current draft (v3.1) of the 'Minimum Information' recommendations is posted at:  
<http://docs.virtualsolar.org/wiki/MinimumInformation>

We would like to know if there are any objections or issues with items that we have included in our list. If there are items that may not be appropriate for your type of experiment, or if there are items that may be missing for the type of analysis that you need for your research.

*Please see the URL for the complete list. What follows is only a subset:*

### **The Overall Collection (High Level)**

- What is the name of experiment?
- Who ran the experiment?  
(organization/institution, PIs)
- If a researcher has questions:
  - Where can they get documentation?  
(website; published papers)
  - How can they get help or report possible problems? (website w/ contact info or a generic email like 'instrument@...')
- How should the experiment be acknowledged in published research?
- What was the goal of the experiment?
- What instruments were used to perform the experiment? (names, acronyms/abbreviations)
- Where were they?  
(spacecraft or observatory name, general location (eg, 'near L1', 'near earth', 'Tenerife, Canary Islands'))
- When did the experiment run?
- What type of observations were collected?
- What type of derived products are available?  
(eg, 'white light coronograms', 'EUV images', 'x-ray spectroscopy', 'daily plots', 'carrington maps')
- Are there caveats or other warnings for potential users of the data? For example:
  - issues with the collection process that might make the data unsuitable for specific uses;
  - known environmental conditions that introduce error during certain periods?
  - Known biases introduced in the calibration or other processing?
  - Known misleading metadata (eg, clock drift)?
  - Any other potential sources of error?
- Is there software in SolarSoft to use the data?
  - If so, where can we get documentation for it?
- Is there any other recommended software to use the data?

### **Dataset Details (Mid Level)**

- What different datasets are in the overall collection?
  - ... different sensors / detectors / cameras
  - ... different observing modes (filters, polarization, cadences, exposure times)
  - ... different processed forms

#### **For each specific dataset:**

- Is there a name or title to distinguish it from the other available datasets?
- What type of data is it?  
(eg, intensity, magnetic field, temperature)
- What are the defining characteristics of the dataset?  
(eg, level of processing, detector used, calibration version, observing mode (filters, exposure time, cadence, etc.))
- What is the purpose / intended use of this specific dataset? (ie, why was the dataset created?)  
...

### **File & Observation Specific Details (Low Level)**

- Is the file in a self-describing scientific format?
- Does it mention how to uniquely reference this file or observation to report problems or check for an updated calibration?
- Does it mention which specific dataset it's a member of?
- Does it provide a URL or other reference to the documentation?
- Have you provided:
  - The time of the observation?
  - The duration (exposure) of the observation?
  - The location of the detector?
  - The pointing of the detector (if appropriate)?
  - Any other details of the observing mode that may vary between observations?
- A checksum to verify file integrity?

### **FITS Specific issues**

- Does the file clearly state that it's a FITS file?
- Is there a reference to the FITS standard, or a link to the FITS website?
- Do the headers include the assigned filename?
- Do the headers include units in the comments?
- Do the headers spell out abbreviations or other coded values?



This and related documentation are available at:  
<http://virtualsolar.org/checklists>