Abstract:
Recent missions such as STEREO and Hinode have dramatically increased the volume of solar physics data available. It may thus be onerous or impractical to inspect every available image, spectrum or other data; instead, solar physicists may come to rely on higher level event and feature catalogs to identify periods of interest for further analysis.

Earlier work from the Virtual Solar Observatory (VSO) has allowed searching for data using the LASCO CME, GOES X-Ray, RHESSI Flare, NOAA Active Regions, KPVT Coronal Holes and the GONG Duty Cycle catalogs, but we have been unable to adapt to the many varied catalogs that have been produced in the field.

There are multiple catalogs for each type of event and feature, generated using data from different instruments and processed by different people or different computer algorithms. These catalogs may contain different records, categorize the events differently, or simply contain different information about the events that they contain.

We present on working being done to ingest and display other existing catalogs for use in searching for data, as well as plans on integrating with the Heliophysics Event Knowledgebase (HEK) being developed for the Solar Dynamics Observatory (SDO).

Catalogs vs. Events
Although the surface there is some duplication of effort with the Heliophysics Events Knowledgebase (HEK) being developed for SDO, there are some subtle differences:

Catalogs tend to be the results of an analysis over a given period of time, rather than just a sampling of random times. As such, they can imply that a given item did not exist during a certain period of time, rather than just a sampling of random times. As such, they can imply that a given item did not exist during a certain period of time.

Catalogs thus require additional information to describe not just the events and the requirements for the events inclusion in the list, but more detailed information about the catalog as a whole.

Catalogs may contain information about more than one type of feature or event per entry, and they may deal with reoccurrences by tracking an event or feature through multiple records.

Catalogs need not describe physical events and features, but can include campaign catalogs, data catalogs or just simple time-series data.

Describing Catalogs
Although we can display catalogs in their native format, we can offer the ability for users to filter, sort and otherwise manipulate the catalogs. To do this, we need to know information about what the fields within the catalog are. Catalogs are often distributed in HTML, MS Excel or as ASCII text, all of which provide great flexibility, but may make it more difficult for us to identify data fields, as the catalog may place different types of information in the same location within a record:

For each of the fields identified, we classify the type of information in it to allow it to properly display, sort and filter the field:

- Time
  - UTC, Carrington Rotation
- Time Range
  - May be a point, area or volume
- Spatial
  - Requires coordinate system and reference frame
- Boolean
  - May be integer or float, bounded or infinite
  - May have a precision or be infinitely variable
- Numeric Ranges
  - May only be orderable (greatest -> least)
  - May be unordered (list of instruments)
- Free Text
- Primary Key
  - eg. NOAA Active Region #, RHESSI flare #
- Foreign Keys
  - May provide a reference to other catalogs of events or features, or to data objects.
- Enumerations
  - May be orderable (greatest -> least)
- Flags
  - Typically reflect a series of boolean values, but may be a way of encoding non-mutually exclusive enumerations.
  - The allowed codings may be extended or deprecated over the life of the catalog.
- URL
  - May be more detailed information about the object being recorded, or to browse images, plots, or movies.

Often, we need to know what the significance of a field’s absence is; there may be a coding to explain the reason for the absence.

The fields may have a useful summary representation:

- minimum
- maximum
- count
- average
- extent / coverage
- histogram
- ratio
- occurrances vs. count
- existence

Semantics
Even if two catalogs have fields that are of the same basic type and the same units, we need to understand what the concepts are that the fields represent to be able to determine the appropriate linkages between catalogs.

In addition, each of the fields may have a reference frame:

- The ‘time of observation’ may be the spacecraft time, or be adjusted to the time at Earth.
- X/Y Coordinates require knowing the location of the observer, and require translation for STEREO/SECCHI.

We hope that in the future we will be able to sufficiently describe the catalogs so that we can derive other fields for improved filtering and matching, such as filtering on event duration when the catalog only contains start and end times.

Catalog Encoding
The values in the catalog may be proxy representations.

- For each of the fields identified, we classify the type of information in it to allow it to properly display, sort and filter the field:
- Time
  - UTC, Carrington Rotation
- Time Range
  - May be a point, area or volume
- Spatial
  - Requires coordinate system and reference frame
- Boolean
  - May be integer or float, bounded or infinite
  - May have a precision or be infinitely variable
- Numeric Ranges
  - May only be orderable (greatest -> least)
  - May be unordered (list of instruments)
- Free Text
- Primary Key
  - eg. NOAA Active Region #, RHESSI flare #
- Foreign Keys
  - May provide a reference to other catalogs of events or features, or to data objects.
- Enumerations
  - May be orderable (greatest -> least)
- Flags
  - Typically reflect a series of boolean values, but may be a way of encoding non-mutually exclusive enumerations.
  - The allowed codings may be extended or deprecated over the life of the catalog.
- URL
  - May be more detailed information about the object being recorded, or to browse images, plots, or movies.

Often, we need to know what the significance of a field’s absence is; there may be a coding to explain the reason for the absence.

The fields may have a useful summary representation:

- minimum
- maximum
- count
- average
- extent / coverage
- histogram
- ratio
- occurrances vs. count
- existence

Semantics
Even if two catalogs have fields that are of the same basic type and the same units, we need to understand what the concepts are that the fields represent to be able to determine the appropriate linkages between catalogs.

In addition, each of the fields may have a reference frame:

- The ‘time of observation’ may be the spacecraft time, or be adjusted to the time at Earth.
- X/Y Coordinates require knowing the location of the observer, and require translation for STEREO/SECCHI.

We hope that in the future we will be able to sufficiently describe the catalogs so that we can derive other fields for improved filtering and matching, such as filtering on event duration when the catalog only contains start and end times.

References
[Helioviewer] http://helioviewer.org/
[VOTable] http://www.ivoa.net/cgi-bin/wiki/view/IVOA/votable