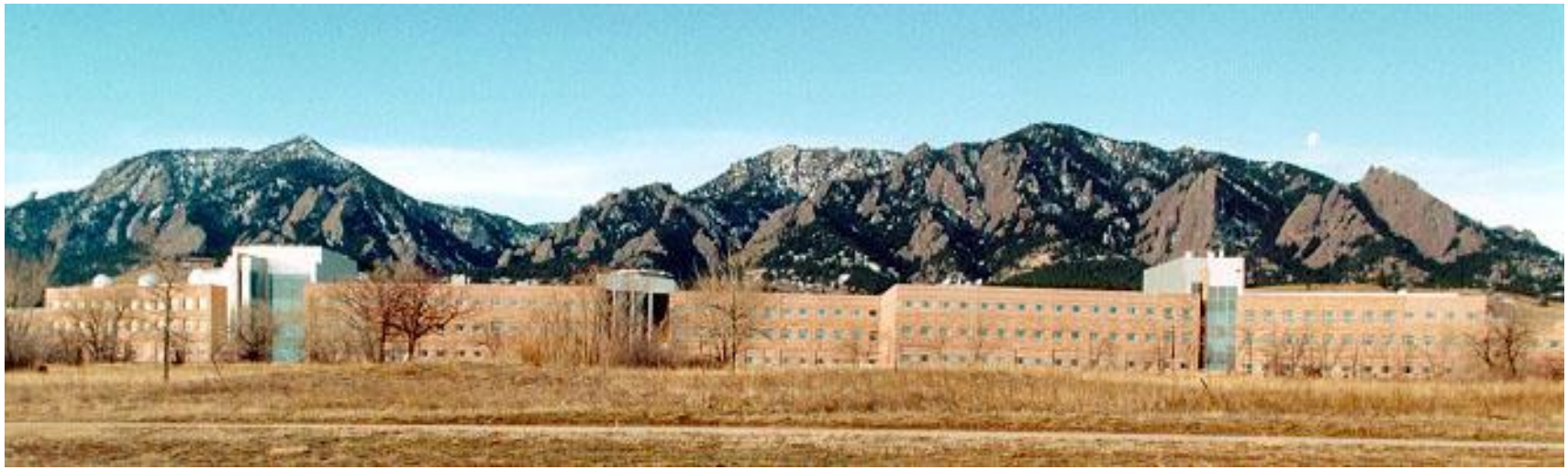


SWx Data Stewardship in NOAA

30 April 2010

Bill Denig

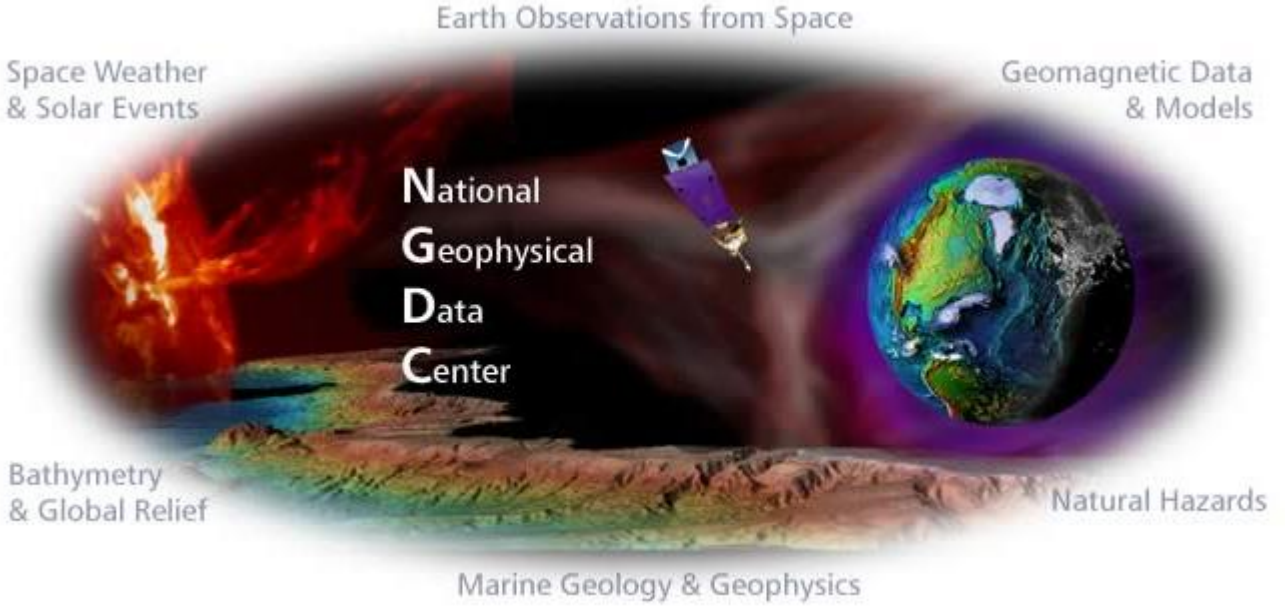
Solar & Terrestrial Physics Division

World Data Center for Solar-Terrestrial Physics

303 497-6323

William.Denig@noaa.gov

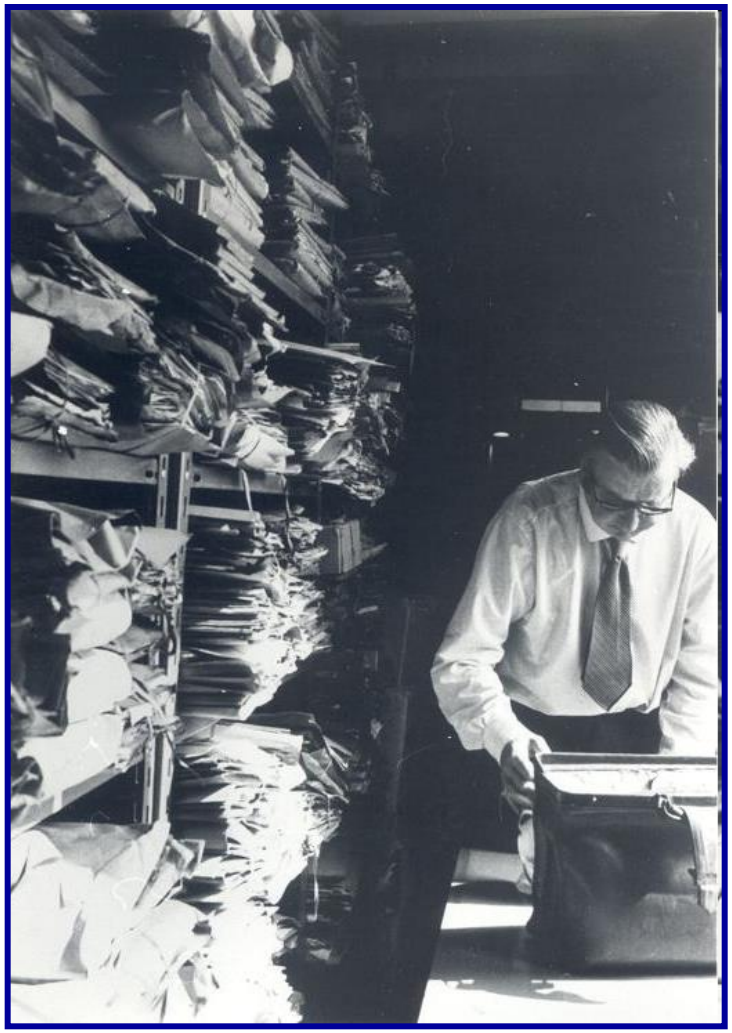
Scientific Data Stewardship
National Geophysical Data Center



NGDC's Mission is *to provide long-term scientific data stewardship for the Nation's geophysical data, ensuring quality, integrity, and accessibility.*

NGDC's Vision is *to be the world's leading provider of geophysical and environmental data, information, and products.*

Scientific Data Stewardship
Challenges



***Dr.* Rob Redmon – circa 2040**

- **Managing the Nation's operational SWx datasets**
  - NOAA's satellite space environmental data & models
  - DoD's space weather data – ground & space
  - Other duties, as assigned – World Data Center
- **Safeguarding historical datasets**
  - Geomagnetic / solar indices & records
  - Solar synoptic drawings and photographs
  - NOAA Climate Data Modernization Program
- **Documenting relevant datasets – metadata[1]**
  - Global Change Master Directory – ISO 19115
  - Space Physics Archive Search and Extract (SPASE)
- **Archiving data for long-term preservation**
  - Comprehensive Large Array-data Stewardship System (CLASS)
  - Open Archival Information System (OAIS)
- **Developing data discovery / access tools**
  - Space Physics Interactive Data Resource
  - New STP website available for comment

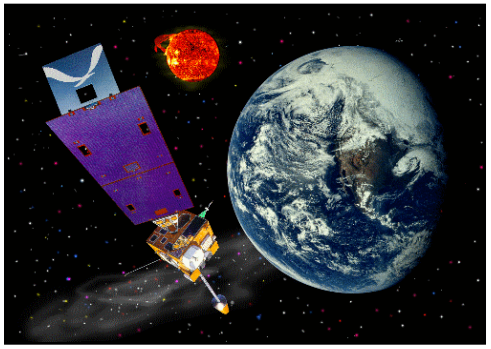
[1]Not to be discussed

Operational SWx Datasets

NOAA's Current SWx Datasets - Satellites

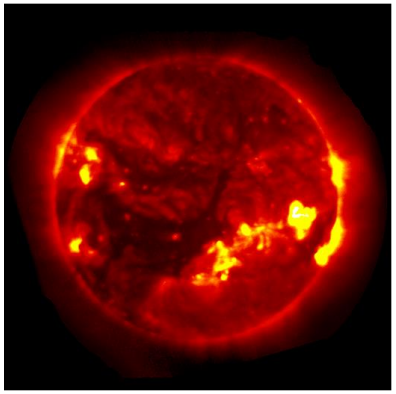
GOES Space Environment Monitor

- Geosynchronous Orbit, since 1974
- Elements: In Situ Magnetic Fields
  Whole Sun X-ray Flux
  Energetic Particles
- All Data are Online



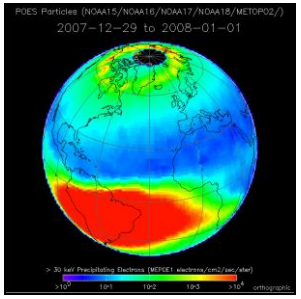
GOES Solar X-ray Imager – GOES 12-15

- Geosynchronous Orbit, since 2003
- X-ray Images taken every minute
- All Data Are Online (once operational)



POES/MetOp Energetic Particle Detector

- Polar Low Earth Orbit
- Energetic Particles Archived Since 1979
- All Data Are Online

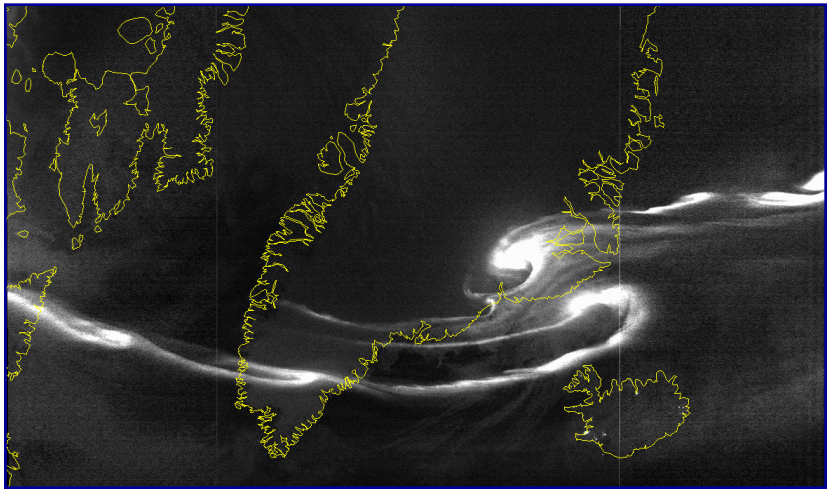


Operational SWx Datasets

USAF SWx Datasets – Satellite & Ground

Solar Electro-Optical Network (SEON)

- Solar Optical Observing Network (SOON) – daily sunspot drawings from 4+ SEON sites
- Radio Solar Telescope Network (RSTN) – Solar radio noise backgrounds at selected frequencies (245 MHz – 15.4 GHz)

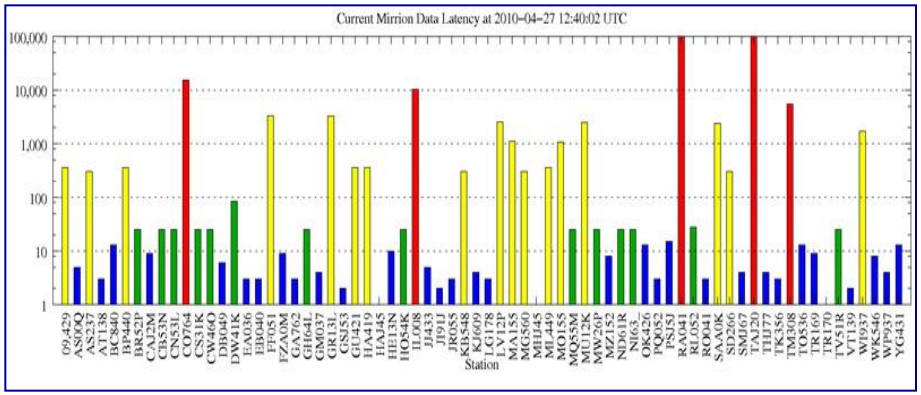


MIRRION Ionosonde Data System

- Digital Ionospheric Sounding System (DISS)
- Vertical Incidence Pulsed Ionospheric Radar (VIPER)
- International ionosonde datasets

Defense Meteorological Satellite Program (DMSP)

- Operational Linescan System auroral imagery
- Space Weather Sensors – SSJ / SSIES / SSM
- SSUSI / SSULI *tbd*



Operational SWx Datasets
Other Duties, as assigned

NGDC serves as both a National Data Center and a World Data Center[1]

NRC The National Research Council

Government data centers are repositories for the nation's environmental data. Methods of data archiving and stewardship are complemented by strategies for ingesting large volumes of raw data. In addition, data centers perform a valuable service to the scientific community through data quality control, integration, and value-added activities, such as processing data and developing tools for data analysis and presentation. In many cases, they have been successful in developing a laudable level of customer service and satisfaction.

NRC, 2003

There is a need for global federations of professional state of the art data management institutions, working together and exchanging practices. Such federations can provide quality assurance and promote data publishing, providing the backbone for the development of a global virtual library for scientific data. They can also complement and assist the multitude of very worthy voluntary initiatives that flourish alongside them by helping to develop and disseminate good practices and standards.

ICSU, 2007

[1]The World Data Center (WDC) structure was officially disbanded in 2008. WDCs will be incorporated into the new World Data System

Safeguarding Historical Datasets
Geomagnetic & Solar Indices

GEOMAGNETIC INDICES BULLETIN

NATIONAL GEOPHYSICAL DATA CENTER
Telephone 303-497-6135
Solar-Terrestrial Physics Division (E/GC2)
325 Broadway, Boulder, Colorado 80303 USA

MONTHLY SUMMARY OF GEOMAGNETIC ACTIVITY

THE GEOMAGNETIC FIELD. The intensity of the Earth's magnetic field at any point in space and time arises from the MAIN field internal to the planet; from the electrical currents flowing in the ionized upper atmosphere; and from the currents induced within the Earth's crust. The main field component varies slowly in time and can be grossly described as a bar magnet with north and south poles that extend well out into space.

The main field creates a cavity in interplanetary space called the magnetosphere, where the Earth's magnetic field dominates any field carried by the charged particles of the solar wind. The magnetosphere shape resembles a comet--a shape owing to the interaction with the solar wind; it is compressed on the side toward the sun and tail-like on the side away from the sun. The magnetosphere also directs the flow of the particles about the Earth.

Particles flowing in the magnetosphere and ionosphere generate currents, which in turn cause variations in the intensity of the Earth's magnetic field. These EXTERNAL currents generate additional currents in the Earth's upper atmospheric layers, which vary on much shorter time scales than the main field, and they create magnetic changes as large as 10% of the main field.

Certain current systems derive their energy from the regular changing solar radiation throughout the day and year. Other irregular current systems obtain their energy from the interaction of the solar wind with the magnetosphere, from the magnetosphere itself, from the interaction between the magnetosphere and the ionosphere, and from the ionosphere itself. Magnetic activity indices, including those reported below, are designed to describe variations in the Earth's main field caused by these irregular currents.

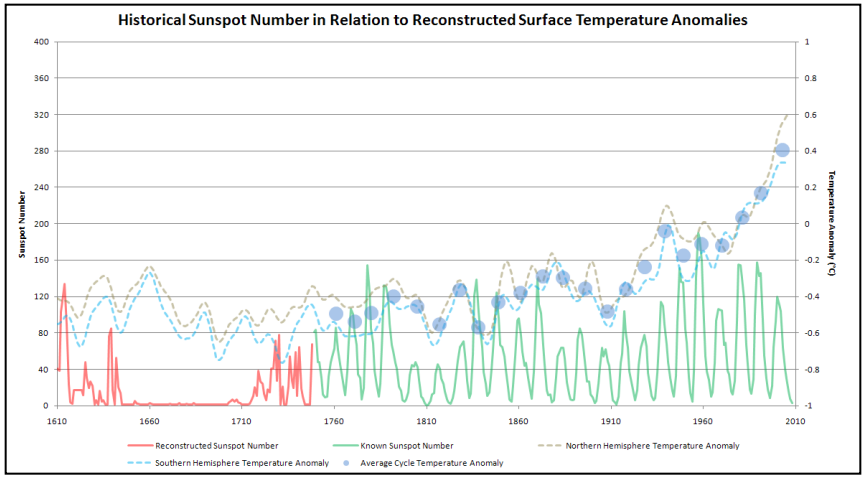
MARCH 1989

Day Cal	Day Jul	Day Bart	Rank Q/D	Kp Three-Hour Indices 1	2	3	4	5	6	7	8	Kp Sum	Ap	Cp	sc (UT)	AFR	An	As	Am	aa Provisional N	S	M1	M2
1	60	18	Q2A	3+	3	2	3-	3+	3-	2	2+	21+	12	0.7		11	23	24	24	24	30	24	30
2	61	19		3+	6-	4-	3+	4	3+	3	3+	30-	25	1.2	0247	19	43	45	44	44	51	53	42
3	62	20		3	5	6-	5-	5-	5	4	3	35	37	1.4		27	50	53	52	42	55	48	49
4	63	21	Q3A	3+	2+	2	3+	4+	3	1	1-	20	13	0.8		9	21	22	22	25	23	24	25
5	64	22		3-	4	5	5+	5-	3+	4-	3+	32	30	1.3		28	48	38	43	54	42	54	41
6	65	23		4	3	3+	4-	5-	4	4	3+	30	24	1.2		22	38	36	37	42	28	31	39
7	66	24	Q9A	4	4+	3+	3	3-	3	2	3	25+	18	1.0		12	28	25	27	40	21	35	26
8	67	25		2+	2+	3-	0+	1-	4+	5+	6-	24-	24	1.1	1755	20	40	36	38	51	35	13	73
9	68	26		5+	5	3+	3+	4-	4	4-	5-	33	31	1.3		21	48	49	48	52	51	46	58
10	69	27	Q10A	4+	4	4-	3-	3-	3	3	3+	27-	19	1.0		17	31	32	32	31	28	31	28
11	70	1	Q8A	3-	3	3	4-	3	3	3-	4	25	17	0.9		12	27	33	30	32	29	33	28
12	71	2		5	3	3	3-	3	5-	4	3	28+	23	1.1		16	35	39	37	40	36	37	39
13	72	3	D1	6	8-	9-	8+	8+	8+	9-	9	65	246	2.2	0127	248	397	390	393	357	340	244	452
14	73	4	D2	9	8-	8-	6-	5	5+	8-	7+	55+	158	2.0		125	230	228	229	214	201	307	108
15	74	5		7-	6	5-	5	4+	5-	4	3	38+	49	1.6		28	70	67	69	71	78	103	45
16	75	6	D5	2	5	5+	7-	5+	5	5	4-	38	50	1.6	0532	38	79	72	76	69	71	78	63
17	76	7		4+	5+	5+	4+	4+	4+	4-	3-	34+	34	1.4		21	51	53	52	48	47	58	38
18	77	8	Q4A	1	2	2	2	4	5+	1+	2	20-	15	0.8		13	23	25	24	17	33	11	39
19	78	9		1	4	6+	6+	7-	6	3+	3-	36+	55	1.6	0423	30	72	77	75	61	88	78	72
20	79	10	Q5A	2+	4-	3+	2+	2+	2+	2-	4	22	14	0.8		12	23	18	21	27	16	19	24
21	80	11		4	4	4+	4-	4-	3-	3	4-	29	22	1.1		20	36	32	34	35	28	37	26
22	81	12		3	3	5-	4	4+	5-	6	6-	35+	39	1.4		26	50	49	50	69	42	34	77
23	82	13		4+	3	3	4+	5	6-	5-	5-	35-	36	1.4		19	56	56	56	48	58	33	74
24	83	14	Q7A	5	4+	3+	2+	2-	2-	1	1	20+	16	0.9		14	27	23	25	24	25	37	12
25	84	15	Q1A	1+	1	1+	2+	1+	3-	4	3	17	10	0.6		8	19	18	19	20	20	12	28
26	85	16	Q6A	3-	3	2-	3	4-	2+	2	4	22+	14	0.8	2249	12	27	22	24	30	24	20	34
27	86	17		3+	3-	4+	4-	6-	6+	5	5+	36+	44	1.5	1342	37	69	55	62	74	60	40	94
28	87	18		4+	3+	4+	6-	6-	5	4	4-	36	39	1.4		33	64	54	59	52	57	46	63
29	88	19	D3	6+	6	5+	4+	5-	5+	7	6	45	71	1.8		38	81	84	82	85	72	74	84
30	89	20		4+	5+	5+	4-	3	3+	6-	7-	37+	47	1.5		35	62	67	64	56	67	58	65
31	90	21	D4	5+	5+	4+	4+	5	6-	6-	5+	41	52	1.6		38	64	77	71	59	72	54	77
MEAN												41	1.26				62	61	62	61	59		60

See back side for definitions of column heading.

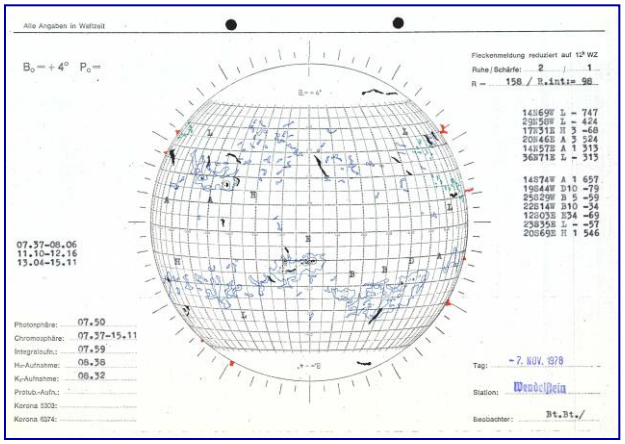
NGDC acquires and archives a variety of solar and geomagnetic indices which are made available in tabular format or as time-series data. Archived indices include:

- Daily/monthly/yearly sunspot numbers
- Solar radio flux, $F_{10.7}$, USAF data
- Geomagnetic K_p, A_p, Dst, AA, AA*
- AMIE derived indices (SWx Climatology)
- Cosmic ray datasets

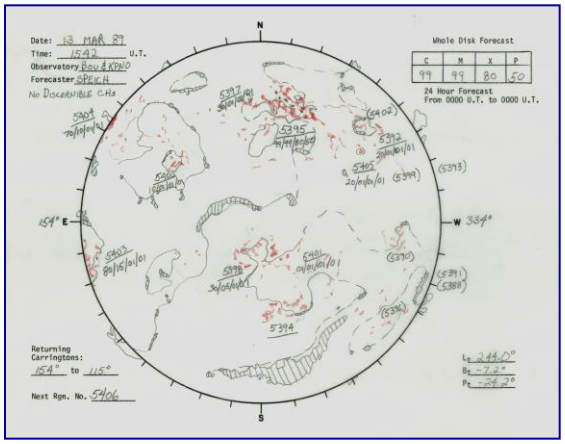


Safeguarding Historical Datasets
Solar Synoptic Drawings

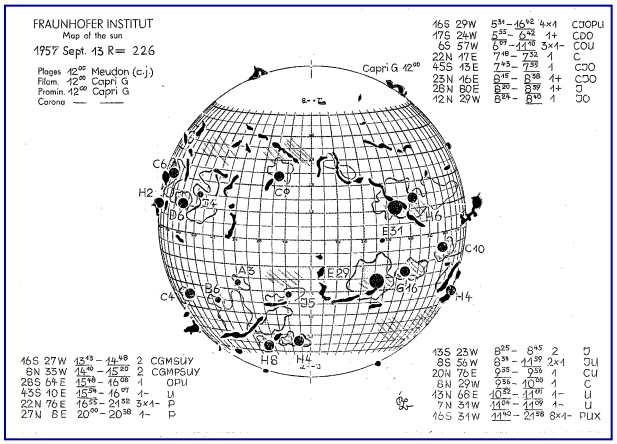
Archives include a variety of solar synoptic drawings, sunspot images and photographs derived from mostly ground observatories. Digitization done through the Climate Database Moderization Program.



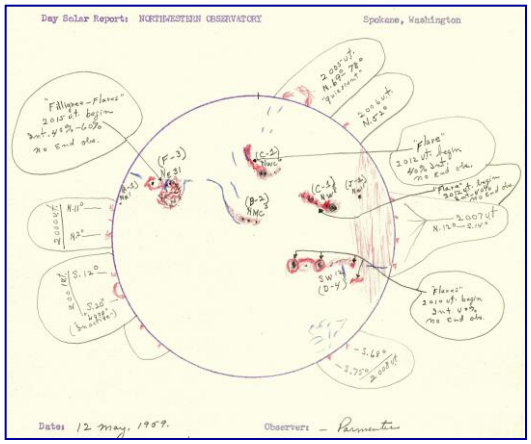
**Wendelstein Observatory*
(1947-1987)**



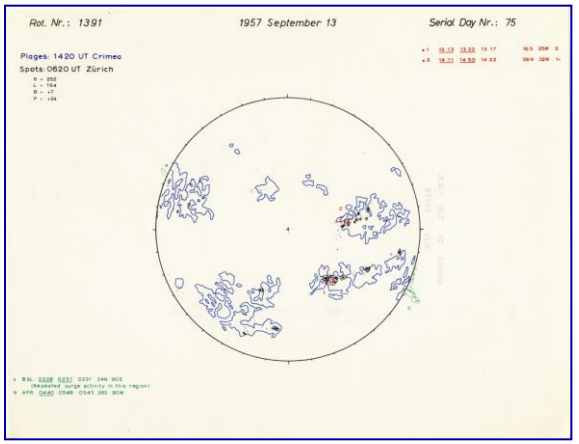
**Boulder Composite Drawings*
(1972-present)**



**Fraunhoffer Institute
(1956 - 1973)**



**Northwestern Observatory*
(1958 -1970)**



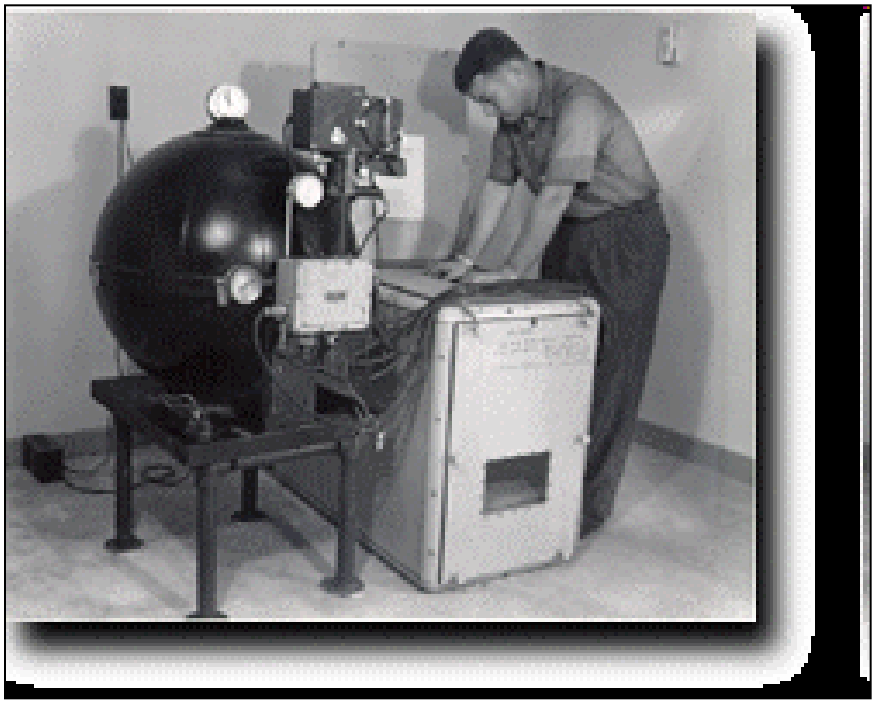
**Drawings from the IGY
(1957 - 1958)**

*Recently added to the archives (2009-2010)

Safeguarding Historical Datasets
Climate Database Moderization Program

Utilizing a National Resource

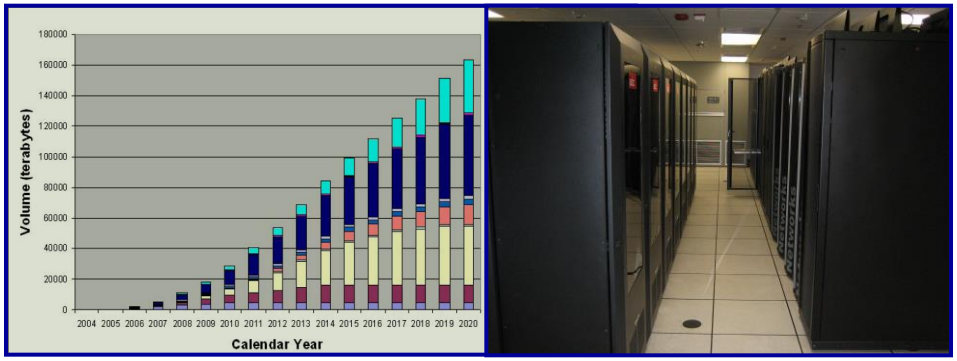
The NOAA CDMP program provides the resources to "rescue" historical datasets by conversion to digital form. NGDC has extensively used this program to facilitate public access to these data, including solar images and composite drawings (previous slide), film ionograms, and geomagnetic data. Continuing programs include digitizing the original records of Scott Forbush (1936 – 1960).



Archive Preservation
CLASS / SNAAP

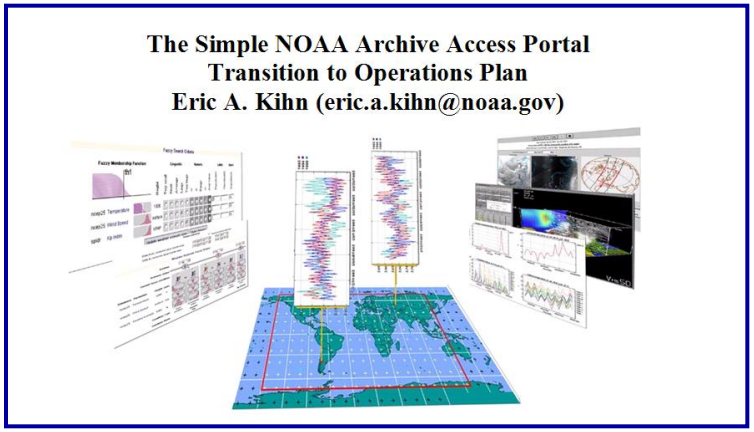
Comprehensive Large Array-data Stewardship System

- CLASS is NOAA's enterprise level data archive and distribution system for NOAA's environmental datasets.
- CLASS is sized to accommodate data from new environmental satellite systems such as GOES-R and NPOESS (now JPSS) plus model datasets and other data sources
- **NOAA's data centers retain responsibility for scientific data stewardship of CLASS datasets**



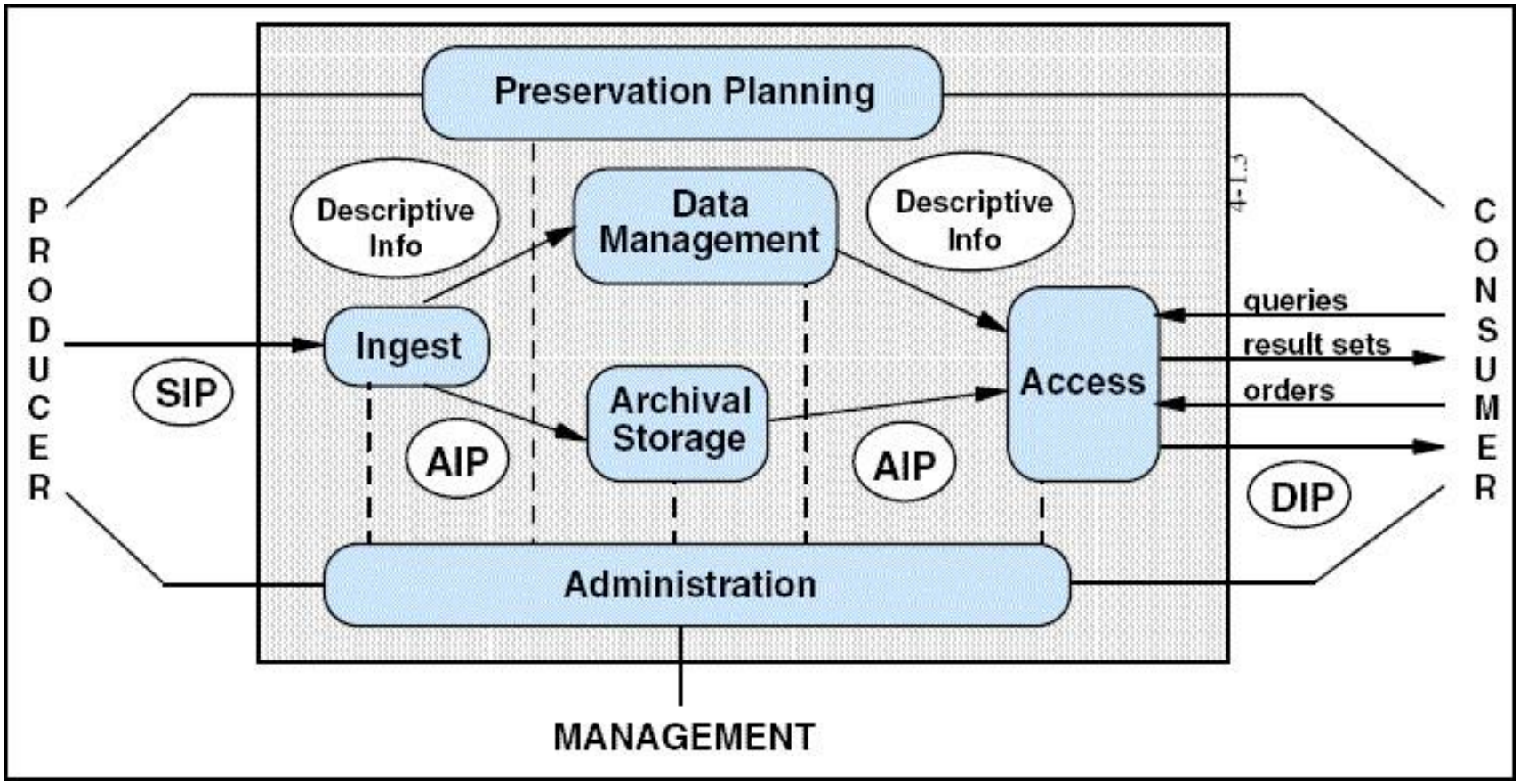
NOAA's Enterprise Archive & Access Tool

NEAAT is an Application Programming Interface (API) that facilitates user access to distributed NOAA environmental datasets including those available in CLASS. NEAAT provides a capability to integrate diverse data systems via a standards-based interface having well defined protocols.



Archive Preservation
Open Archival Information System (OAIS)

STP & CLASS are implementing the OAIS structure for stewarding SWx environmental databases. Changes to the end-to-end management of data include a more rigorous adherence to standards including policies for accepting new datasets, preparing submission agreements, and maintaining proper metadata (see also NAO 212-15).



Data Discovery & Access Tools
Space Physics Interactive Data Resource

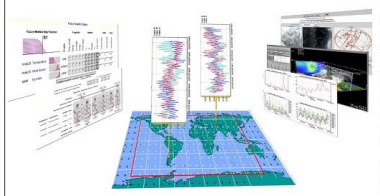
The Space Physics Interactive Data Resource - ReST Web Services

Eric Kihn, , Mikhail Zhizhin, Peter Elespuru , Rob Redmon, NATIONAL GEOPHYSICAL DATA CENTER

Abstract
(Integration, Convenience, Utility)

The Solar Terrestrial Physics division of the NOAA National Geophysical Data Center (NGDC) in Boulder, CO is focused on the dissemination of high quality space climate and space weather data sets and services. The Space Physics Interactive Data Resource (SPIDR) system is the primary web distribution method for NGDC's STP data and metadata holdings. This includes key NOAA space data sets such as GOES, POES, DMSP and also ground array data collected through the World Data Center. Through outreach to the user community it has been recognized that the SPIDR web-application style of data distribution does not meet the entire communities' needs or preference for data access, for example those who wish to develop applications which access the data directly, or those who prefer to integrate data through their own tools. In response we have developed a ReST style interface to all of the SPIDR data and metadata holdings. This poster presents the details, capabilities and potential use cases for the SPIDR ReST services. In addition we present a new API and IDL client built on the ReST interface for rapid access to all of NGDC's SPIDR accessible data sets.

Motivation – To make NGDC's data holdings available, easily accessible and simple to integrate with existing applications. To make sure that the archive service can be assumed in client infrastructure.



The National Geophysical Data Center supports science through the provision of "science quality" data and information pertaining to space. The development of a web services layer on SPIDR is meant to support users who wish to integrate the power of a national archive into their architectures. SPIDR supports both SOAP and ReST style access to key data holdings with the vision that all NGDC stewarded data will be made available seamlessly to client developed applications. This is accomplished through the provision of high quality data and metadata in a robust but simple access mechanism. It is hoped that the NGDC archive data will be presented to many user communities, through many interfaces tailored specifically to that designated user community. This is aligned with the NASA Virtual Observatory (VO) philosophy and in fact many NGDC data sets are being served through relevant VOs.

In addition to its space weather holdings NGDC also serves satellite, ocean, and modeled terrestrial weather data, much of which can be accessed though the same system.

Clients and Applications

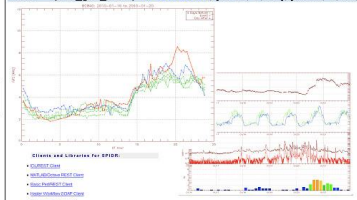
The addition of Rest and SOAP style web services into SPIDR allows for the incorporation of space weather resources directly into many types of clients. For example NGDC has developed an IDL based client with simple functionality:

```
spidr_get_data( parameter, date range )
spidr_get_metadata( parameter )
```

This returns an analysis ready IDL structure containing both data and metadata. This simple client example is available at: https://sourceforge.net/projects/spidr-idl/

Example output from the client is shown below. In addition NGDC has developed "crib sheets" which help new users develop and extend their own client to use the data.

```
iono = spidr_get_data( 'foF2.BC840', [2010, 1, 16], [2010, 1, 21])
```



Above is a list of the current demonstration clients developed by NGDC for accessing SPIDR data. All are open source and freely modifiable

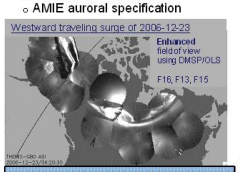
Accessing NGDC data on the ever popular iPhone app.

Data Assets and Services

NGDC's Auroral Assets

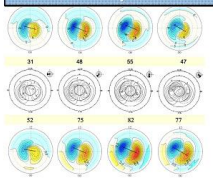
- DMSP/OLS - visible auroral imagery
- DMSP J4 – precipitating particles
- DMSP IDM – ion drift
- AMIE auroral specification

Westward traveling surge of 2006-12-23



SPIDR Data Holdings

Example Model Data: Assimilative Mapping of Ionospheric Electrodynamics



SPIDR Data Dashboard

SPIDR provides a "data dashboard" which provides update information for various station based data sets.

SPIDR ReST Web Services

Real Time Ionosonde (MIRRION)

REST-based API Specifics The ReST interface provides the means to obtain data and metadata, but it also has the ability to provide high quality image plots (PNG). ReST calls are very straight forward conceptually, and amount to constructing a URL with the appropriate fields Data and images are available from the GetData call, and designated through a 'format' parameter. Metadata is obtained from the GetMetadata call. There are details for all three, data, plots, and metadata, below. Creating a ReST call for SPIDR consists of formulating your base URL, and then adding the additional pieces you need to accomplish your desired goal. The base URL will typically be http://spidr.ngdc.noaa.gov/spidr/servlet/. What you tack on to that will vary based on whether you want data, metadata, or images, and for which parameters and when. The nice thing about ReSTful web services is that examples can easily be provided through a web browser. Here are several examples that illustrate some of the capabilities of the REST web services.

The full documentation is available at:
http://docs.google.com/gview?url=http://spidr.ngdc.noaa.gov/spidr/docs/SPIDR.REST.WSGuide.en.pdf

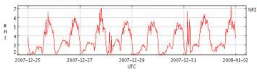
Example SPIDR ReST Calls and Results

A call for .kp index data, in CSV format
- format=csv
- param=index_kp
- dateFrom=20090924
- dateTo=20091001

URL/call
http://spidr.ngdc.noaa.gov/spidr/servlet/GetData?format=csv&param=index_kp&dateFrom=20071225&dateTo=20080101

A line plot of the foF2 ionospheric parameter for the Boulder ionosonde, with the following call parameters
- format=image
- param=foF2.BC840
- dateFrom=20071225
- dateTo=20080101
- marks=none
- height=200

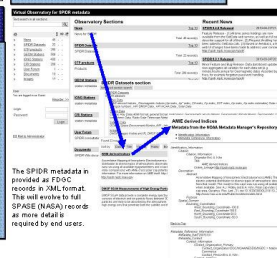
URL/call
http://spidr.ngdc.noaa.gov/spidr/servlet/GetData?format=image&param=foF2.BC840&dateFrom=20071225&dateTo=20080101&marks=none&height=200



Metadata Services

SPIDR currently provides metadata (documentation about data) via a tool called VxOWare. This metadata is available both through the SPIDR web-application and the ReST style web service so that client developers can serve this crucial component of data delivery to end users. ReSt style access is via a describe call to the GetMetadata service. e.g **http://spidr.ngdc.noaa.gov/spidr/servlet/GetMetadata?describe**
Just like with the GetData call, you can provide a parameter to the describe call to drill down and obtain more specific information about specific stations. e.g.
http://spidr.ngdc.noaa.gov/spidr/servlet/GetMetadata?describe&param=cri
Unlike data, the GetMetadata call doesn't have a third tier of metadata for a theme + platform, as it doesn't pertain to metadata. Only data have an associated time range, which is what the third tier supports. The GetMetadata call will accept such requests, but suggests that you probably wanted the associated GetData call instead. e.g.
http://spidr.ngdc.noaa.gov/spidr/servlet/GetMetadata?describe&param=cri.BKSN

SPIDR Metadata Tools



The VxOWare used to manage and access SPIDR metadata is an open source product available at: http://sourceforge.net/projects/vxoware

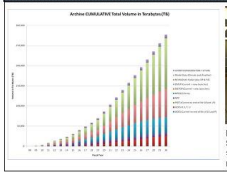
The SPIDR metadata in provided as FDGC records in XML format. This will evolve to full SPASE (NASA) records as more detail is required by end users.

Partners (incomplete list)

HMO, WDC, DIAS, UMASS/LOWELL, U.S. AIR FORCE, Australian Government IPS Radio and Space Services, Science & Technology Facilities Council Rutherford Appleton Laboratory

Storage and Dissemination Capabilities

NGDC Data Holdings

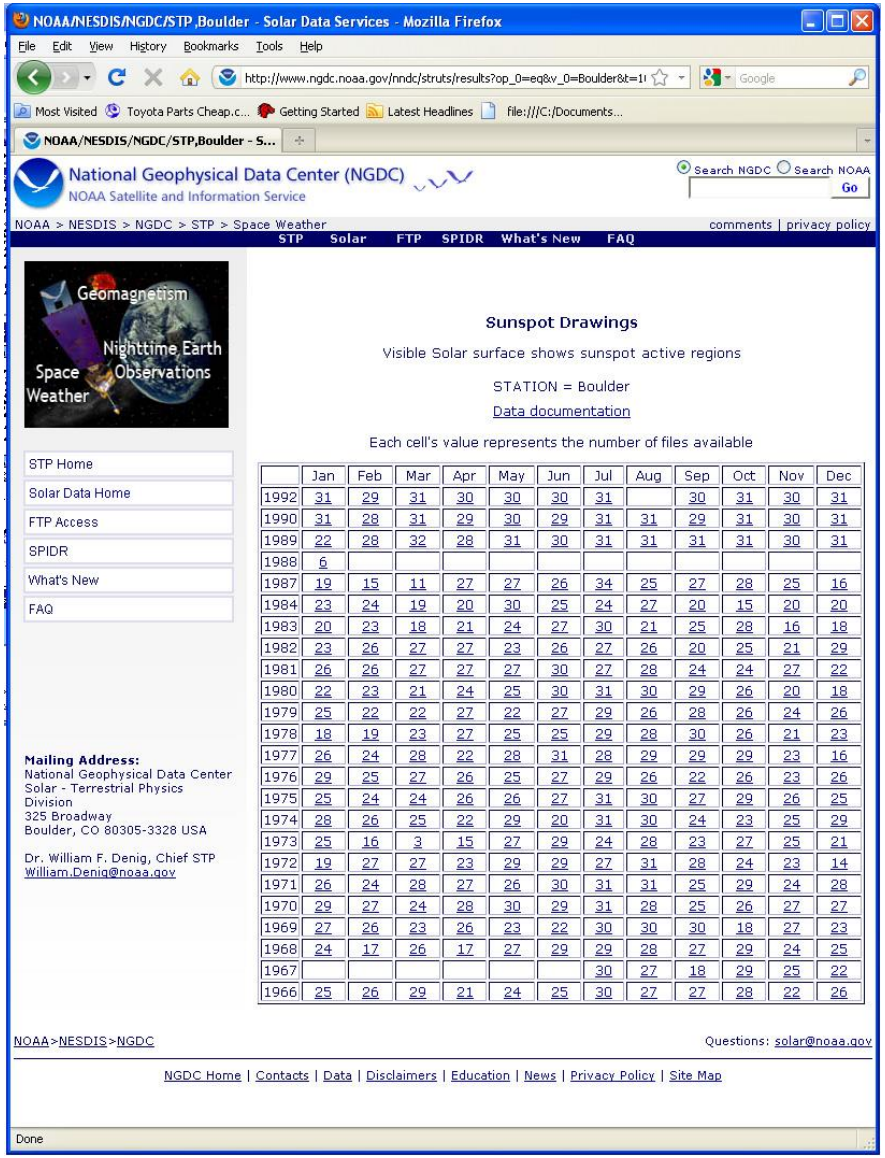


NGDC resources include the Comprehensive Large Array Stewardship System (CLASS). CLASS is a fully redundant tape and spinning disk system capable of handling 100's or petabytes of data with a planned network capacity of 8 Tb/day ingest and 24Tb of distribution.

Data Discovery & Access Tools
New STP Website

On April 26th STP released a new website to review holdings and download datafiles. Data are available through the NGDC homepage or directly from:

http://www.ngdc.noaa.gov/stp/stp.html



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1992	31	29	31	30	30	30	31		30	31	30	31
1990	31	28	31	29	30	29	31	31	29	31	30	31
1989	22	28	32	28	31	30	31	31	31	31	30	31
1988	6											
1987	19	15	11	27	27	26	34	25	27	28	25	16
1984	23	24	19	20	30	25	24	27	20	15	20	20
1983	20	23	18	21	24	27	30	21	25	28	16	18
1982	23	26	27	27	23	26	27	26	20	25	21	29
1981	26	26	27	27	27	30	27	28	24	24	27	22
1980	22	23	21	24	25	30	31	30	29	26	20	18
1979	25	22	22	27	22	27	29	26	28	26	24	26
1978	18	19	23	27	25	25	29	28	30	26	21	23
1977	26	24	28	22	28	31	28	29	29	29	23	16
1976	29	25	27	26	25	27	29	26	22	26	23	26
1975	25	24	24	26	26	27	31	30	27	29	26	25
1974	28	26	25	22	29	20	31	30	24	23	25	29
1973	25	16	3	15	27	29	24	28	23	27	25	21
1972	19	27	27	23	29	29	27	31	28	24	23	14
1971	26	24	28	27	26	30	31	31	25	29	24	28
1970	29	27	24	28	30	29	31	28	25	26	27	27
1969	27	26	23	26	23	22	30	30	30	18	27	23
1968	24	17	26	17	27	29	29	28	27	29	24	25
1967							30	27	18	29	25	22
1966	25	26	29	21	24	25	30	27	27	28	22	26

Scientific Data Stewardship
Concluding Remarks

- **NGDC/STP is responsible for the historical preservation of NOAA's and DoD's operational SWx data plus other mission related datasets**
- **STP datasets include a variety of current and historical SWx datasets dating from the 1957 International Geophysical Year (IGY) and earlier**
- **SWx datasets must have associated content and format descriptions to ensure future utility of the data (metadata – not discussed herein)**
- **NOAA capabilities and policies are used for the long-term historical preservation of environmental datasets (see NAO 212-15)**
- **Data access tools facilitate user search and discovery of SWx datasets within and available through STP**

NAO 212-15 – Management of Environmental and Geospatial Data and Information (12/2/2008)