2010 ANNUAL STATISTICS AND HIGHLIGHTS REPORT FOR THE

NATIONAL SPACE SCIENCE DATA CENTER

Edwin J. Grayzeck

National Space Science Data Center

Greenbelt, Maryland 20771

2011-Mar-01

Table of Contents

PREFACE

1. INTRODUCTION

2. HIGHLIGHTS

3. DATA MANAGED AT NSSDC AND 2010 INFLOW AND OUTFLOW

- 3.1. Data Inflow
- **3.2.** Data Outflow

4. ADDITIONAL NSSDC SERVICES

- 4.1. NSSDC Information Management System (NIMS)
- 4.2. SPASE and Virtual Observatories (VOs)
- 4.3. Consultative Committee for Space Science Data Systems (CCSDS)
- 4.4. Sun-Earth Day (SED)

Glossary

PREFACE

The National Space Science Data Center (NSSDC) serves as the permanent archive for NASA's Space Science community. A major component of its mission is to ensure future data accessibility and usability. NSSDC also provides current data access, complementary to the efforts of other NASA "active archives," in support of the NASA and international astrophysics and space physics research enterprises. Finally, NSSDC is a conduit for the general public and education community to acquire NASA space science data that may interest them.

Herein we report on the activities of the NSSDC for the calendar year 2010. As much as possible, we report the same statistics as in previous years to enable interested parties who wish to compare accomplishments year-to-year. Nevertheless, as NSSDC evolves, some statistical tables have been updated to better reflect current operations. Note specifically that many statistics in this report are only comparable to those from the 2005 and later, since our central NIMS database was revamped in 2005.

NSSDC is pleased to issue this 2010 Annual Report describing the 2010 growth and evolution of NSSDC's data archives, access pathways, and other tools and services, as well as the access to those data and services by NSSDC's customer communities. This report has been made WWW-accessible in the hope that readers will avail themselves of the opportunity to link to the services reported herein.

I welcome suggestions from users for improvements to this Annual Report and to NSSDC services.

Edwin J. Grayzeck

Head, National Space Science Data Center

1. INTRODUCTION

This report characterizes NSSDC's data holdings, metadata holdings, access pathways, and value-added data products, tools, and services at the end of 2010, with a focus on the 2010 activities leading to that end-of-year state. In addition this report characterizes the nature and access to NSSDC's data and services by its many users from various communities. It is assumed the reader will have a general familiarity with NSSDC and its mission. For more information see the top NSSDC web page at http://nssdc.gsfc.nasa.gov/.

In 2005 we assembled an external user group, the NSSDC User Group (NUG), which meets periodically to provide findings regarding NSSDC goals. The group's reports, most recently from March 2009, are on-line and can be linked to from the NSSDC homepage. The NUG will meet again in early 2011.

2. HIGHLIGHTS

The center of this report is the 13 Tables which summarize NSSDC activities in 2010. In most cases these numbers speak for themselves, though it is irresistible to address a few highlights.

The most important result of NSSDC's 2010 continuing activities is the preservation of growing space science data volumes, ensuring their continuing and future accessibility to the space science, education and general public communities. NSSDC's archive has now grown to 108 TB of space science data and an additional 3.3 TB of Earth science data. During 2010, 19 TB of data were added to the NSSDC; the largest components of that were in our holdings for HEASARC and FERMI.

In 2001 NSSDC began using a reengineered data management approach, which stored data as Archive Information Packages (AIPs, i.e. bundles of data and related descriptive information) written to DLTs. In 2010 the staff made several enhancements and extensions to the NSSDC production software, the package called "xware," and procedures. The xman software that produces manifests of data submissions was modified to differentiate between various versions of submission information packages from the Planetary Data System (PDS). The xchunk software that generates AIPs was enhanced to be able to process PDS data volumes built to older versions of the PDS volume standard. An ingest database was developed to support NSSDC's ingest operation. The database provides information needed by data providers to generate data submission manifests and manages the ingest process from data receipt to final archival of data to permanent media. Finally, NSSDC changed from SuperDLT to LTO-4 for its archival storage media.

3. DATA MANAGED AT NSSDC AND 2010 INFLOW AND OUTFLOW

There are several ways to characterize the multi-disciplinary NSSDC archive and we use most of them herein to give a true sense of the NSSDC, i.e. byte counts, numbers of distinct datasets, and numbers of media volumes managed, as well as the diversity of datasets and of media types. For the remainder of this section we will present this variety of statistics in tables, similar in format

to prior years' reports, though recognizing that the content of some tables will not be comparable to those for years earlier than 2005, when our databases were revamped.

Table 1. Counts of NSSDC Datasets on December 31, 2010						
Discipline	Digital	Non-Digital	Total			
Astrophysics	228	76	304			
Space/Solar Physics	1,207	666	1,873			
Planetary Science	1,116	761	1,877			
Earth Science	109	130	239			
Other	117	440	557			
Total	2,777	2,073	4,850			

By the end of 2010 NSSDC was managing 4,850 distinct datasets and accompanying documentation packages. Table 1 indicates the disciplines from which these datasets come and whether the datasets are digital or non-digital. Space Physics had been the dominant discipline by dataset counts, accounting for nearly half of NSSDC's holdings. This reflected that in its early years NASA launched a preponderance of space physics missions and also that space physics spacecraft typically carry more independent experiments than do astrophysics missions. The total number of datasets increased in 2010 by 425; the vast majority of those (338) were digital datasets from Planetary Data System (PDS), which began delivering a backlog of its data to NSSDC in 2010. Now the dataset counts are nearly equivalent for Space Physics and Planetary.

NSSDC manages almost as many non-digital (e.g. film, microfilm and microfiche) datasets as digital datasets, though in recent years newly arriving data has been all digital. NSSDC also has generated digital versions for some of its film archive, often in response to requests.

Table 2 is a different characterization of the NSSDC archive, showing byte counts for the entire digital archive. Some of the byte counts, particularly for older media, are estimates involving assumptions about the mean numbers of bytes on various media types for some datasets. We see the expected shift to planetary emphasis began this year with the arrival of large planetary datasets, showing an increase of our planetary holdings by 3.5 TB, compared to 2009.

Table 2. State of the NSSDC Archive December 31, 2010				
	All Digital Data (TB)			
Astrophysics	53.39			
Space Physics	27.78			
Planetary	9.48			
Earth Science	3.27			
Other	16.85			
Total	110.77			

Table 2. State of the NSS	DC Archive I	December 31	2010
		December 31,	2010

Data are also being moved from NSSDC's traditional offline archive to a near line archive based on DLT, SDLT, and LTO-4 jukeboxes attached to unix and linux servers, respectively. Data are newly archived in Archive Information Packages (AIPs), which hold data files and companion

attribute files and are media-independent and platform-independent. These are defined as per the AIP concept of the ISO/CCSDS Open Archival Information System reference model. Table 3 shows the volumes of data ingested to this portion of the archive for 2008-10; the total of data stored as AIPs has reached 11.6 TB.

	2008		2009		2010	
	AIPs	GB	AIPs	GB	AIPs	GB
ALOUETTE	17208	9.42	16281	9.43		
APOLLO					2	0.02
GP-B					532	82.41
ISIS	114775	40.40	22168	10.47		
LEGACY						
DATA	871	6.40	1073	14.97	1282	50.51
PDS DATA	88	288.95	171	247.66	2180	2492.14
RHESSI	8216	587.03	7233	577.52	6178	424.74
SPDF					2	0.06
WIND	346	0.29	367	0.30	275	0.22
Totals	141,504	932.50	47,293	860.35	10,451	3050.09

Table 3. Data	Indested to	Nearline	Permanent	Archive
I able J. Dala	ingesteu to	Near IIIIe	I CIMANCIII	

In Table 3 "Legacy Data" refers to data currently archived on aging magnetic tapes, generally 9-track tapes or 3480 cartridges. These data are read and ingested as AIPs.

About half of the data stored in AIPs are made network-accessible on nssdcftp for the convenience of some of the user community. Table 4 (below) lists NSSDC's network-accessible Space Physics data as of 31 December 2010 for projects with > 1GB. Spacecraft with less data - including Aeros, Alouette, ARCAD, Cassini, DMR, Galileo, Genesis, Hinotori, Mariner 10, Prognoz 6,7, and 9, San Marco, SNOE and miscellaneous others – are summed as "Other" at the end of list. It must be noted that nssdcftp has a continually evolving directory structure, so the table represents only a snapshot.

Table 5 (below) characterizes the digital media types managed at NSSDC, not including backup copies. It should be noted that most volumes are replicable and have one backup volume. The Earth Science tapes are continuing to be transferred to NASA's Earth Science archive, thus the 2010 summary shows a further reduction of over 1100 mostly 3480 cartridges in the Earth Science discipline.

Top Level Dir	GB
ACE	51.021
AE	1.279
AMPTE	2.348
APOLLO	13.566
CRRES	34.061
DE	186.642
DIRBE	19.963
EXPLORER	4.401
GEOTAIL	2.945
HELIOS	2.538
IBEX	9.067
IMAGE	260.995
IMP	114.601
ISEE	36.060
ISIS	206.721
MAGSAT	1.869
OGO	2.397
OMNI	29.359
PIONEER	4.961
SAMPEX	54.070
SOLARMAX	20.892
ULYSSES	297.337
VOYAGER	30.461
WIND	31.446
Others*	9.098
TOTAL	1428.099
* total for spacecraft each, including Ae	

Table 4. Space Science Data FTP Accessible from NSSDC on December 31, 2010

* total for spacecraft with <1Gb data each, including Aeros, Alouette, ARCAD, Cassini, DMR, Galileo, Genesis, Hinotori, Mariner 10, Prognoz 6,7, and 9, San Marco, SNOE, Viking and miscellaneous additional spacecraft.

	Astro- physics	Space Physics	Planetary Science	Earth Science	TOTAL
4-mm Tape	717	92	3	0	812
8-mm Tape	189	503	74	0	766
9-Track Tape	531	4743	1646	10137	17057
3480 Cartridges	466	1937	1125	1556	5084
DLT	65	31	2	0	98
LTO-4	65	0	0	0	65
CD	854	23169	5346	54	29423
DVD	1015	1333	255	0	2603
12 Worm	0	4	0	0	4
M-O Disk	274	0	0	0	274
Total	4176	31812	8451	11747	56186

Table 5. Counts of Volumes* at NSSDC Archive on Dec 31, 2010

* Backup volumes and those not attributable to these 4 disciplines are not included.

Discipline	Microfilm (reels)	Microfiche (cards)	Film (ft)	Film (frames)	Movie (reels)	Maps	Slides
Astrophysics	2357	5963	100	11975	0	0	62
Earth Science	55	773	13067	64610	4	0	0
Planetary Science	2188	13492	409081	335053	181	1773	768
Space Physics	34589	30507	580	6173	2	0	36893
Other	3224	3624	3785	6206	20	0	2112
Totals	42,413	54,359	426,613	424,017	207	1773	39,835

Table 6. Analog Data Products at NSSDC by Discipline

Table 6 lists NSSDC's analog archive holdings by disciplines and by form factor. The film and movie counts are based on the inventory of the photo materials started in 2007. Movie reels had not been called out separately in previous reports, now they are. Film (feet) are reels of uncut film containing sequences of still photos. We removed from the inventory any backup copies of images, which had notably increased previously reported numbers. In 2008 we completed an inventory of the microfilm and microfiche portions of the analog archive, increasing both. Apparently the earlier estimates were conservative. This table is unchanged from 2008 and not comparable to versions from earlier years. The continuing inventories of other categories -- e.g. sorting through duplicate sets of media – have been suspended due to budgetary considerations. We expect no changes in this summary table in the foreseeable future.

3.1 Data Inflow

	Astrophysics	Space Physics	Planetary	Total
4-mm Tapes	130	0	0	130
CDs	1	12	0	13
DVDs	29	123	0	152
LTO-4	11	-6	0	5
Totals	171	129	0	300

Table 7. Media Arriving at NSSDC During 2010*

* Ephemeris and Other data not included.

Table 8. Data Arriving at NSSDC During 2010

	GB	Sub-Totals by Discipline
Astrophysics		
FERMI	9600.01	
FUSE	101.80	
GALEX	2049.85	
HEASARC	3908.14	
		15659.80
Planetary		
PDS	2492.14	
		2492.14
Space Physics		
ACE	2.56	
Apollo	0.02	
RHESSI	424.74	
SPDF	0.06	
Twins	311.47	
Ulysses	1.01	
Wind	0.22	
		740.07
Miscellaneous		
Gravity Probe-B	82.41	
Earth Science	-3.44	
Legacy Data	50.51	
		129.48
TOTAL		19021.49

Table 7 (above) characterizes the in-flowing media types by discipline. The numbers are net, i.e. some providers redeliver data on other media and we return/subtract their originals, so as not to double count, which that also explains any negative numbers. In recent years, CDs were the dominate input media type received; in 2010 it was DVDs. However, with their larger capacity, the much smaller number of LTO-4 tapes delivered a larger volume of data.

Table 8 shows by project the data volumes that NSSDC received in 2010, approximately 19 TB of new data via a combination of electronic deliveries and on media. Dominating the statistics are HEASARC and FERMI in Astrophysics plus PDS in Planetary. The actual HEASARC delivery was many terabytes larger, but by the terms of the MOU data sets delivered earlier are removed from NSSDC and archived on the newly delivered media. We expected an increase of data deliveries from PDS and saw the first wave of that in 2010.

3.2 Data Outflow

NSSDC provides access to its data holdings with network-accessible data for chosen datasets and, in

addition, through a user support infrastructure for the mailing of offline digital and non-digital data volumes. Most electronic interfaces are accessible through NSSDC's WWW home page

and include special WWW-based interfaces to specific datasets or groups thereof and ftp pathways to a range of data files maintained permanently on NSSDC disks at nssdcftp.

Much of the data outflow discussed in NSSDC Annual Reports before 2003 was activity within SPDF, which maintains the Active Archive for NASA Space Physics missions; the activities of CDAWeb, etc, now are covered in SPDF reports elsewhere. Because NSSDC and SPDF have been and are still co-located at NASA's Goddard Space Flight Center, nssdcftp remained a shared resource for now and that is reflected in the reported statistics.

	2006	2007	2008	2009	2010
Photo Gallery	936,039	720,213	481,089	193,577	152,533
Spacecraft Data	802,438	689,961	1,856,362	1,564,930	2,314,095
All others on nssdcftp	998,388	930,342	634,070	615,242	789,413
TOTAL	2,736,865	2,340,516	2,971,521	2,373,749	3,256,041

Table 9. Number of Files Downloaded via FTP

Table 9 gives the annual counts of files downloaded from nssdcftp and singles out those from the Photo Gallery and Spacecraft Data subdirectories, those with high activity. The Photo Gallery continued to decline, possibly because of the proliferation of websites with space photos. Researchers downloading data files via ftp from the Spacecraft Data subdirectory had a notable increase from recent years, showing the high interest in and great value of these services provided by NSSDC and SPDF on this shared resource.

NSSDC responded to 57 distinct requests for ftp data and for data products. Table 10 below characterizes the user community of requestors. To a very large extent it is the U.S. and international general public, the education enterprise, publishers, etc. and their desire for NASA imagery.

Table 10.	NSSDC User	Community	(Offline Red	quests Only)	for CY 2010
-----------	------------	-----------	--------------	--------------	-------------

Affiliation Category	Total Requests	Percent
No Affiliation [General Public]	9	16%
Non_US	6	11%
US Academic Institutions	8	14%
US Private Industry	1	2%
NASA/GSFC	28	49%
NASA Centers, Excluding GSFC	3	5%
Other Government Agencies	2	4%
Total	57	100%

Table 11 (below) gives the counts of requests for offline datasets from various disciplines in 2010 and as integrated over NSSDC's history. Note particularly the dominance of planetary data over both time scales. This is largely associated with lunar and planetary image data that are widely requested by the general public. The number of requests is different than in Table 10 because some requests are for data/items related to more than one discipline, so are double counted in Table 10; some requests are for documents or other form of information that are not counted as a datasets, thus not included in Table 11.

DISCIPLINE	Data Set Requests 1968 - 2010	Data Set Requests 2010
Astrophysics	11466	5
Earth Science	7156	0
Planetary Science	47616	22
Space & Solar Physics	9199	15
Ephemeris	97	0
Other	48	3
TOTAL	75582	45

In 2010 data distribution via ftp was preferred due to its immediacy and to NSSDC budget limitations, though some media were distributed. Table 12a below shows the distribution of data served within these categories for 2010 and the previous two years; Table 12b the distribution of the items by discipline for the same years. Both Tables 12a and 12b show ftp as the dominant mode of distribution, even though these statistics include only ftp data that were newly posted in response to a request. Table 12 does not include all data previously available via ftp.

Tables 12a,b. NSSDC Offline Data Dissemination Statistics 2008-2010

	Table 12a				Table 12	2b	
ITEMS	2008	2009	2010	DISCIPLINE	2008	2009	2010
DISCs	186	196	209	Astrophysics	95	10737	27
PRINTED	109	36	21	Planetary	20979	740	932
OTHER	4	11	2	Space Physics	102	9470	6319
FTP	20877	20719	7059	Other	0	15	13
TOTAL	21176	20962	7291	TOTAL	21176	20962	7291

NOTES:

DISCs include CDs & DVDs

PRINTED materials include Photos, Posters, Maps, Documents

OTHER media include Microfilm, Microfiche, Tapes, Videos, Reels

FTP include Data, Documents, & Photos posted for FTP download, not files already posted

4. ADDITIONAL NSSDC SERVICES

In addition to its archive of scientific data and the variety of data interfaces characterized in the preceding sections, NSSDC offers a number of additional services, which are described below.

4.1 NSSDC Information Management System (NIMS)

The NSSDC Information Management System (NIMS) encompasses most of the separate databases that NSSDC has used to track data and information through the years. The NSSDC has a long term goal of incorporating its off-line data inventory system into NIMS and a major effort for this is underway.

Subpartition	Number of Records as of 12/31/2010	Number Added in 2010
Spacecraft Experiment Dataset	6673 5413 5849	116 15 469
Totals	17,935	600

Table 13. NIMS/JEDS Database Statistics for CY 2010

Number of spacecraft with experiment records - 1,083 Number of experiments with datasets at NSSDC - 1,610 Additional datasets associated only with spacecraft, not experiments - 656 Additional datasets that are not associated with spacecraft/experiment - 430

Table 13 details the pertinent NIMS database statistics for 2010. NIMS identifies virtually all launched spacecraft, the experiments carried by many of these spacecraft, and datasets from these spacecraft primarily as archived at NSSDC. This portion of the database is the source of information for many of NSSDC's WWW information pages. The NSSDC Master Catalog (NMC) dynamically generates WWW pages so that the latest information is presented to the user. A number of discipline and project pages are based on information derived from NIMS or utilize the NMC to generate such information.

4.2 SPASE and Virtual Observatories (VOs)

With over 30 years experience in managing and preserving digital information comprising thousands of datasets, NSSDC is acutely aware of the need to acquire and preserve data and adequate documentation to ensure they are independently understandable and usable for current and future researchers. In this era of increasingly seamless access to archived data the NSSDC mission also includes support for the Virtual Observatories, particularly as part of the development of the Space Physics Archive Search & Exchange (SPASE), the dictionary which will be the common language among space physics archives.

Version (2.2.0) of the SPASE Data Model was released on 11 November 2010. NSSDC continues to actively participate in the evolution of SPASE and will be updating the registry interface to accommodate changes in the recent Data Model. More information can be found at http://www.spase-group.org/ which also has a link from NSSDC through its VO Portal.

4.3 Consultative Committee for Space Science Data Systems (CCSDS)

The NSSDC continues to lead the Data Archiving Group within the Consultative Committee for Space Data Systems (CCSDS). This Working Group continues to campaign for the widespread adoption of Archiving Standards such as the *Reference Model for an Open Archival Information System (OAIS)*. The reader is referred to <u>http://www.ccsds.org/</u> for specifics.

Within the last year the updated OAIS Reference Model CCSDS Standard was reviewed within CCSDS and is now being reviewed as a new version of the ISO Standard. The first follow-on archiving standard -- the *Producer-Archive Ingest Methodology Abstract Standard (PAIMAS)*, essentially a checklist of activities for ingest to archives -- was reconfirmed unchanged as a CCSDS Standard and an ISO Standard. In addition, a new standard - the *Producer-Archive Ingest Specification*, an XML-based modeling and transfer of data objects transferred from a Producer to Archives – is nearing completion of its technical content. An update to the legacy standard for describing binary data down to the bit level, *EAST Specification Language*, was approved as an updated CCSDS and is also currently undergoing ISO review.

NSSDC is also part of the leadership for the Repository Audit and Certification (RAC) Working Group within CCSDS. The RAC Working Group is defining two new standards - *Audit and Certification of Trustworthy Digital Repositories*, which provides best practices to maintain information over the long term, and *Guidelines for Auditors of Trustworthy Digital Repositories*, which provides procedures for ISO auditors to certify Archives. Both of these standards have passed CCSDS review and are currently under international review to become new ISO Standards.

CCSDS also hosts the Information Packaging and Registries Working Group, the Information Architecture Working Group and the Space Assigned Numbers and Authority Working Group, all of which have significant involvement from NSSDC personnel. From these groups, *the XML Formatted Data Unit* was published as an ISO Standard.

4.4 Sun-Earth Day

In 2010 the NASA Sun-Earth Day team, with major NSSDC participation, prepared for and orchestrated Sun-Earth Day 2010 with the theme "Magnetic Storms" for the main event on March 20. Many thousands of packets of information were sent to teachers, scientists and others for Sun-Earth Day programs, reaching hundreds of thousands of people with live webcasts and podcasts. Throughout the year the team also sponsored a number of workshops and teacher professional development events reaching thousands of teachers, amateur astronomers, and the general public in partnership with Heliophysics missions, museums, science centers, and planetariums, as well as science and educational professional societies. The same team of scientists and educators are working to support the next Sun-Earth Day on Mar 19, 2011 (see http://sunearth.nasa.gov/); the theme "Ancient Mysteries, Future Discoveries" will be highlighted.

Glossary

ACE	Advanced Composition Explorer
ACL	Atmospheric Explorer
AIP	Archive Information Package Arc Aurorale et Densite
ARCAD	
CCSDS	Consultative Committee for Space Data Systems
CDAWeb	Coordinated Data Analysis Web
CRRES	Chemical Release and Radiation Effects Satellite
DE	Dynamics Explorer
DIRBE	Diffuse Infrared Background Experiment
DLT	Digital Linear Tape
DMR	Differential Microwave Radiometers
DVD	Digital Versatile Disk (originally, V = video)
FTP	File Transfer Protocol
FUSE	Far ultraviolet Spectroscopic Explorer
GALEX	Galaxy Evolution Explorer
GB	Gigabyte
GP-B	Gravity Probe-B
GSFC	Goddard Space Flight Center
HEASARC	High Energy Astrophysics Science Archive Research Center
IBEX	Interstellar Boundary Explorer
IMAGE	Imager for Magnetopause-to-Aurora Global Exploration
IMP	Interplanetary Monitoring Platform
ISEE	International Sun-Earth Explorer
ISIS	International Satellite for Ionosphere Studies
ISO	International Organization for Standardization
JEDS	Java Experiments, Datasets, Spacecraft
LTO-4	Linear Tape-Open [i.e. open standard], 4 th generation
MAGSAT	MAGnetic field SATellite
M-O	Magneto-optic
NIMS	NSSDC Information Management System
NMC	NSSDC Master Catalog
NSSDC	National Space Science Data Center
NUG	NSSDC User Group
OAIS	Open Archival Information System
OGO	Orbiting Geophysical Observatories
OMNI	Interplanetary Medium Data (not an acronym)
PAIMAS	Producer-Archive Ingest Methodology Abstract Standard
PDS	Planetary Data System
RHESSI	Reuven Ramaty High Energy Solar Spectroscopic Imager
SAMPEX	Solar Anomalous and Magnetospheric Particle Explorer
SDLT	Super DLT (see above)
SNOE	Student Nitrogen Oxide Explorer
SPASE	Space Physics Archive Search & Exchange
SPDF	Space Physics Data Facility
TB	Terabyte
TWINS	Two Wide-angle Imaging Neutral-atom Spectrometers
VO	Virtual Observatory
WORM	Write-Once, Read-Many