Twelve Experiments, One Goal: Archive Usable Science Data for the Cassini Mission to Saturn

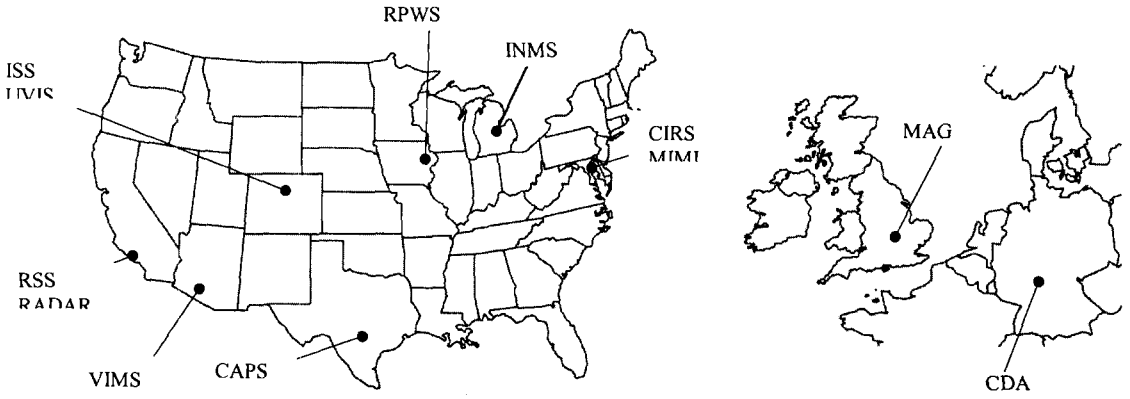
D. L. Conner[1], R. F. Beebe[2],

[1]Jet Propulsion Laboratory, Pasadena, CA 91109, USA.

[2] *New Mexico State University, Las Cruces, NM 88003, USA.*

The Cassini Mission to Saturn will expand our understanding of the Saturn system and our solar system. The Cassini spacecraft with the Huygens Probe was launched on 15 October 1997[1,2,3]. Saturn Orbit Insertion (SOI) occurs on 1 July 2004. The probe will be released from the orbiter on 24 December 2004 and on 14 January 2005, the ESA Huygens probe will descend through the atmosphere of Titan and relay data to the orbiter for up to 2.5 hours. Cassini will tour the Saturn system for 4 years and then the baseline mission ends mid 2008, for a total mission duration of 10.7 years. The importance of this unique tour of the Saturn system will be revealed in the wealth of scientific findings obtained from the data acquired. Archiving this data in a usable form is critical for future scientific studies of this remarkable system. With NASA's current two decadal schedule of recurring visits to the individual outer planets, it is necessary to ensure access to the data for future users. A reliable Cassini archive will support a data analysis program, allow comparisons of the Saturnian system with the Jovian system, (the Galileo and Jupiter Icy Moons Orbiter (JIMO) missions), and provide a basis for future mission planning.

The task of archiving Cassini science data requires significant resources, development of critical timelines, and management of twelve instrument teams who were individually selected and are distributed around the world. Developing a data archive that is of consistent quality and detail has been a significant challenge in the Cassini distributed operations environment (fig. 1). Downlinked files are cleaned and transmitted to individual data centers at each instrument home site where they process the data into usable formats. The individual teams have a one-year validation period to organize the data and solve calibration problems before submitting data to the Planetary Data System (PDS) for release to the science community and archiving (fig. 2). It is assumed that data processing will be carried out in tandem with early science analysis of selected components of the data. Incorporating consistent standards into the various systems requires a great deal of communication and teamwork. It is also important for distributed teams to report archiving progress to the project to facilitate upward status reporting.



CAPS	Cassini Plasma Spectrometer, *Southwest Research Institute, San Antonio, USA*
CDA	Cosmic Dust Analyzer, *Max Plank Institute für Kemphysi, Heidelberg, Germany*
CIRS	Composite Infrared Spectrometer, *Goddard Space Flight Center, Greenbelt, USA*
ISS	Imaging Science Subsystem, *Space Science Institute, Boulder, USA*
INMS	Ion And Neutral Mass Spectrometer, *University of Michigan, Ann Arbor, USA*
MAG	Dual Technique Magnetometer, *Imperial College, London, UK*
MIMI	Magnetospheric Imaging Instrument, *Johns Hopkins University, Baltimore, USA*
RADAR	RADAR, *Jet Propulsion Laboratory, Pasadena, USA*
RPWS	Radio And Plasma Wave Spectrometer, *University of Iowa, USA*
RSS	Radio Science Subsystem, *Jet Propulsion Laboratory, Pasadena, USA*
UVIS	Ultraviolet Imaging Spectrograph, *University of Colorado, Boulder, USA*
VIMS	Visible And Infrared Mapping Spectrometer, *University of Arizona, Tucson, USA*

Figure 1. Cassini Distributed Operations Environment

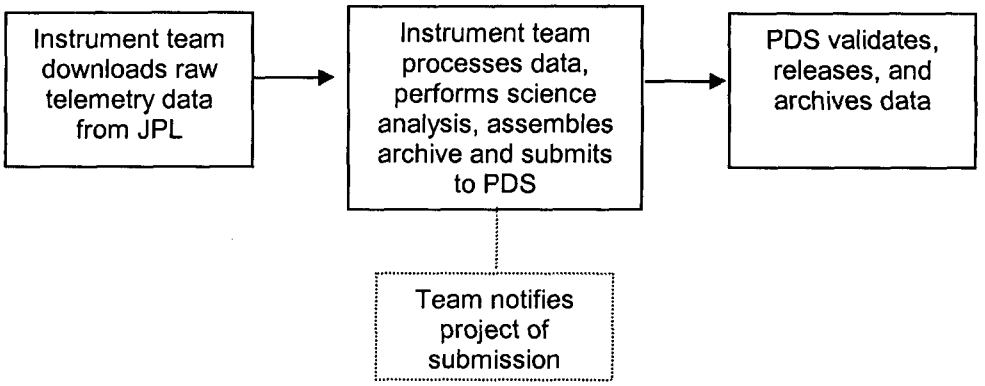


Figure 2. Distributed Operations Data Flow

Communication has been the key to successfully developing and operating in a distributed environment. The timeline in figure 3 illustrates how the archive development has progressed through the mission and the various working groups that were formed to advance development. The Cassini tour will consist of a series of passes through the Saturnian system leading to a complex observing sequence that will place huge demands on the mission staff; therefore it was necessary to define data products and pipelines prior to the Saturn Approach Phase of the mission.

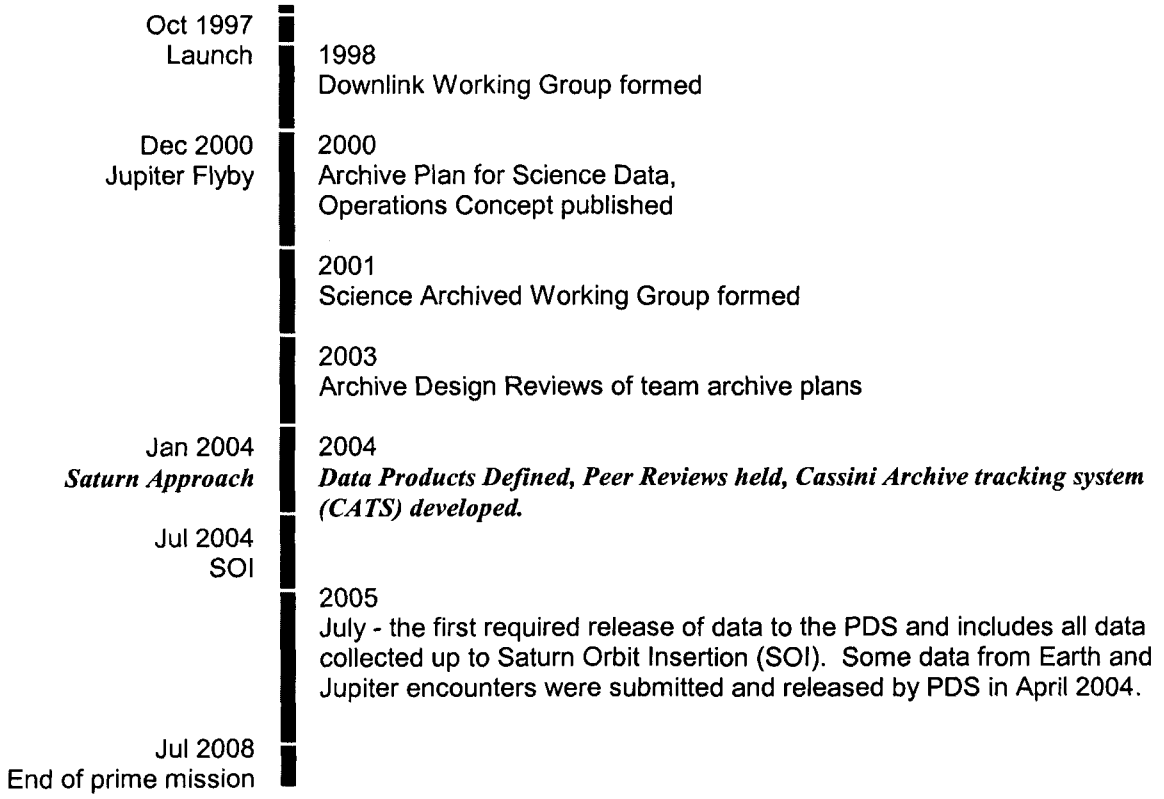


Figure 3. Archive Development Timeline

In 1998 the Cassini Downlink Working Group (DWG) was formed with members from the Cassini operations system team, instrument teams and a PDS representative. The group began preliminary discussions to define an Archive Generation and Validation Operations Concept, and finalize requirements that were formalized in 2000 and incorporated in the Cassini Functional Design Document [4]. The operations concept defined a direct interface between the instrument teams and PDS to develop, produce, and deliver archive products. Teams are required to archive raw and calibrated data, or raw data and calibration files and algorithms that can be used to produce calibrated data. The DWG also used sessions to introduce instrument teams to PDS standards and requirements. Doing this early ensured that the operational cost of archiving data with the PDS was understood and could be budgeted. It was equally important that PDS understood the volume and types of data returned by instruments so they could plan for the operations of archiving, preserving, and distributing this data to the science community.

The DWG also developed the Cassini/Huygens Archive Plan for Science Data (APSD) [5], which was signed by the Project, PDS, and NASA HQ. The APSD defines roles and responsibilities, describes the datasets teams plan to archive and provides data volume estimates. The work done in the DWG laid the foundation for a more concentrated development effort that began in January 2001 with the formation of the Science Archive Working Group (SAWG). The SAWG was chartered to complete the development of the archive system and included representatives from the program, each instrument team and appropriate PDS discipline nodes. The working group concentrated on providing instrument archiving teams with PDS standards [6] training, defining Cassini standards for representing parameters such as time and coordinate systems, and developing the master index keyword parameters. Archiving teams worked with the PDS nodes that specialize in their science discipline to define data products and archive volume layouts in interface specification documents. Over 75 types of data products have been defined and an estimated 10 Terabytes of data will be archived.

By January 2003 preliminary archive design reviews began, and each instrument team's data product definitions, data formats, label files, documentation, and validation plans were scrutinized by a review board with membership from PDS and the science user community. The reviews resulted in nearly 100 action items to be addressed by the project, instrument teams, and the PDS. These detailed archive design reviews, proved to be an effective means for communicating archive designs. As a result, data formats were clarified, documentation plans improved, and requirements better understood. Teams gained the knowledge to focus on the areas, (outlined in the below table), that required the most attention and would have the greatest influence on the usability of data.

DOCUMENTATION	of the mission, spacecraft, instruments and datasets; a detailed history of the mission and full description of the capabilities.
STANDARDIZATION	of data formats - to ensure long-term access, data formats should be as simple as possible.
PRESERVATION	of navigation and calibration - retrieval of pointing information and calibration need to be planned early and accomplished in a monitored incremental process.
VALIDATION	of data products – validation for science content is best accomplished by distributing products to the science team for science analysis in their archive form.
OPTIMIZATION	for correlative studies - coordination among experiments to ensure correlative studies are supported by defining standards for data products, indexing and documentation.

DOCUMENTATION

Providing appropriate context for the data is critical to understanding how that data can and cannot be used. Several types of documentation are important for understanding archived data. These include: the mission plan, description of spacecraft subsystems, instrument design and operations description, and characteristics of the dataset. Maintaining the mission plan with details of the mission as it progresses is crucial to understanding the possible contents of datasets. The mission plan should be periodically updated to reflect changes in the mission and science objectives, tour geometry, and spacecraft and instrument operations. Spacecraft subsystems, individual instrument assemblies and hardware components, instrument orientations and interfaces with the spacecraft as well as operational capabilities and instrument modes should be thoroughly described. Detailing instrument component characteristics is important because specific hardware technology may come into play when analyzing data and initial documentation may be lost for outdated instrumentation. Performance specifications of instrument components and individual detectors should be accompanied by a history of operational capabilities.

Two important aspects of dataset documentation involve data format specifications and calibration algorithms. Datasets must be described at the bit level in interface specifications and human-readable calibration/reduction algorithms must be provided to ensure that data access and algorithms can be implemented in future programming languages and computing environments. The amount of detail that is needed in dataset documentation is usually underestimated. Or, there may be insufficient resources to document how the data were collected, processed, and used by the science team.

In addition to data formats and calibration algorithms, specific efforts should include the maintenance and preservation of an observer's notebook. The observer's notebook documents science goals associated with specific observations, unplanned data outages, processing anomalies, instrument anomalies and their effect on the data, and other ongoing errata information.

Cassini uses a centralized system to capture high level science goals associated with observations during the science planning process so that science goals are associated with resource usage. These data will be archived as a set of searchable Event SPICE [7,8,9] files that could be combined with the observer's notebook to relate desired goals to data.

Including detailed documentation that describes the mission plan, spacecraft subsystems, instrument design and operations, and characteristics of the dataset are imperative to understanding the dataset because they provide the context and instructions for using the data.

STANDARDIZATION

Complicated data structures that require specialized software for access are a problem for long-term archives. If data are formatted for use in an environment that requires specific software they tend to be proprietary in nature and their accessibility is inherently short-lived. Because individual teams tend to rely on familiar formats, (and they have strong opinions of how databases should be designed), it is important to emphasize standardization of formats among instruments and across missions, if possible, long before

data processing systems are conceived. For most non-imaging instruments, tables or time series ASCII formats provide the most flexibility. However, many modern instruments involve various modes of spatially-resolved, hyper-spectral imaging, yielding data that are best stored in arrays with multiple back planes (images and cubes). For higher-level spatially resolved products it may be convenient to store reduced geometry parameters related to each pixel in additional back planes, further complicating the data, but greatly enhancing its ease of use. Consideration of the possible complexity of these datasets illustrates the challenge in establishing standards. However, by using formats that meet PDS standards, developers can ensure their data will be accessible using basic PDS maintained tools.

In a mission like Cassini that involves remote sensing, (CIRS, ISS, RADAR, RSS, UVIS, and VIMS), and in situ measurements, (CAPS, CDA, INMS, MAG, MIMI, RPWS), (fig. 3), correlative keyword parameters are not common to all instruments. Time of observation is the most fundamentally shared parameter and should be consistently expressed among datasets.

An example of data format standardization that has worked very well is the PDS Navigation Ancillary Information Facility (NAIF) SPICE System[7,8,9]. The SPICE system defines data file formats for storing spacecraft and planetary ephemeris, spacecraft pointing, planetary constants, and instrument field-of-view and mounting offsets. All planetary missions are now using SPICE files to store and distribute these ancillary data. NAIF also maintains a toolkit in many languages to access these files in most computing environments. Usage of the NAIF SPICE system ensures common geometric parameters can be incorporated into the dataset.

It is necessary to define standard formats that are adaptable enough to accommodate different types of data and to ensure that keyword parameters are consistent among instruments so that datasets can be associated based upon common temporal and spatial coverage.

PRESERVATION

The process for creating and distributing ancillary data such as navigation and spacecraft pointing was planned early in the mission so that the schedule of availability could be factored into data processing schedules. Operational designs such as the order and frequency of downlinked data and length and position of individual orbits within the Saturnian system affect the production schedules for creating ancillary data and must be considered when establishing requirements on the data archiving schedule. Spacecraft navigation, pointing reconstruction, and planetary orbital elements need to be updated periodically as the mission progresses so they can be used as inputs to the data processing pipeline and be archived with science data. One of the most difficult problems for a user of data after it has been archived is to determine the observing geometry for a specific data file. This emphasizes the critical need to ensure ancillary geometric data is included with datasets as parameters within data files for in situ experiments and within keyword parameters for remote sensing experiments. However, improvements of these geometric ancillary data through optical navigation and ephemeris refinements may not be available to factor in at the time of science data archiving. The preservation of these ancillary data and inclusion of

documentation and subroutines for producing typical parameters will serve the science community for the long term.

Calibration is an area that continues to be worked as the mission progresses. The importance of preserving early pre-flight calibration and associated documentation cannot be stressed enough since these can be used a baselines for future calibrations. As calibration techniques progress and become consolidated throughout the mission, the algorithms used to calibrate data and any calibration files or models will be completed. It is important to preserve a calibration history that describes temporal changes in instruments and identifies which calibration scenarios apply. Without this critical calibration information, data may not be usable.

Project and instrument team vigilance is required to maintain ancillary data and calibration information. Without the proper archiving of these data, datasets will be short-lived and very limited in their use.

VALIDATION

The data processing pipeline should be constructed so that team members use the data files destined for archive to perform preliminary science analysis is the most important form of validation. Science team members can provide feedback to the archivist to ensure the data content is optimized for science analysis. Examples of feedback include: data confidence statements, requests for format changes, additions to the dataset, and calibration changes. Software can be developed to verify that data products conform to defined formats; however, the development of consistent documentation and methods of reliable calibration requires a strong link between the science team and the archivist. An early review of a sample volume and the functional interface between the instrument team archivist and the PDS partner are extremely important for developing a properly structured and complete archive. The exercise of producing a sample volume, which can be an iterative process, provides the science team archivist experience in assembling a PDS compliant archive and the PDS partner experience with accessing the data files and understanding the archive volume layout. The sample volume has proven to be an invaluable tool for validating the archive design and functional interface. Once the archivist is ready to produce a complete volume for external review, PDS coordinates a peer review of the dataset. PDS peer reviews include members of the science community that are internal and external to the team and function as the final check that a dataset can be used to perform science analysis. Periodic reviews of the documentation should be held during the systematic production of datasets to verify temporal changes such as instrument performance, data acquisition schedules, observation plans, and mission extensions are reflected accurately. A well-maintained observers notebook can be used to check that the documentation represents the content of the dataset and verify that the team has acquired the data planned.

Pre-production validation in the form of sample volumes and peer reviews are important to validate the design of the production system. The most important validation is obtained by active use of the dataset to perform science analysis during the mission, which ensures its usability once the data are archived.

OPTIMIZATION

The ultimate goal is to integrate each dataset from a Cassini instrument into the PDS archive. Integration requires adequate long-term standardization that allows users to carry out correlative studies utilizing multiple instruments, possibly from multiple missions, to address a variety of scientific questions. PDS has developed long-term standards that define keyword parameters that are needed to support cross correlative studies in disciplines such as atmospheres, surface morphology, and magnetic field and particle interaction. These keyword parameters map to attributes in the master index. The master index is a database of parameters that enables the future user to search for and access only those datasets that satisfy the search criteria. Early in the planning phase, the PDS partner must work with the instrument archivist to ensure that keywords, formats, coordinate systems [10], time representation and map projections, if applicable, are well understood and consistent with PDS standards. As the mission progresses, the archiving team should populate the master index with parameters that define temporal, spectral, and spatial resolution and coverage as well as illumination and viewing geometries.

The early definition of keywords that will be used to populate the master index and integrate the datasets with spacecraft pointing information, allows discriminating searches, limits down loading of extraneous data and simplifies data use.

In summary, developing a quality archive of the Cassini data has been challenging in a distributed operations environment for a mission with a broad range of experiments. Communication through regular working group meetings has been the key to success for developing a cohesive system and identifying solutions to problems. Science teams now have mature designs for producing PDS compliant archives. Although, some development to implement those plans into operational systems still remains, the design foundation has been laid. Some teams have already produced archive volumes that have undergone PDS peer review, which demonstrates that their system implementation is ready for pipeline production. There have been many lessons learned with respect to proper documentation, standardization of data formats, the preservation of ancillary data, and optimization of datasets, which we have detailed in this paper. And, although the Cassini tour is composed of a series of unique orbits, each utilizing the instruments differently, the archiving challenge is not unique. Constructing an archive that provides prompt and standardized data access is a task faced by all NASA and ESA missions. The length of time between a mission conception and final completion of the dataset ensures that there will be staff turnover, new budget constraints and various instrument and spacecraft anomalies. In addition, many of the current missions include new instrument groups who are participating in a mission for the first time, requiring procedures be in place for efficient archive training. The Cassini mission presents many archiving challenges. The lessons learned by Cassini will be used to improve PDS standards and the archive planning process for future planetary missions.

References:

[1] C. Russell, Ed. SSR, Vol 104, Nos 1-4 (2002)

[2] D. Seal, Cassini Mission Plan, JPL D-5564, Revision O (October 2003)

[3] R. Mitchell, The Cassini/Huygens Mission, 54th International Astronautical Congress, The World Space Congress (2003)

[4] Cassini Functional Functional Design Document, JPL 699-500-3-GS/D, Rev C, Change 13, 22 July 2003

[5] D. Conner, Cassini / Huygens Program Archive Plan for Science Data Initial Release Version 1, JPL D-15976, (August 3, 2000)

[6] Planetary Data System Standards Reference, Version 3.6, JPL D-7669, Part 2. (August 1, 2003).

[7] Navigation Ancillary Information Facility SPICE toolkit download web site ftp://naif.jpl.nasa.gov/pub/naif/toolkit

[8] Navigation Ancillary Information Facility SPICE tutorial ftp://naif.jpl.nasa.gov/pub/naif/tutorial

[9] Cassini anonymous ftp site for downloading SPICE files ftp://naif.jpl.nasa.gov/pub/naif/CASSINI/kernels

[10] Seidelmann, et al., Celestial Mechanics and Dynamical Astronomy, 82, pp. 83-110 (2002).

Twelve Experiments, One Goal: Archive Usable Science Data for the Cassini Mission to Saturn

D. L. Conner[1], R. F. Beebe[2],

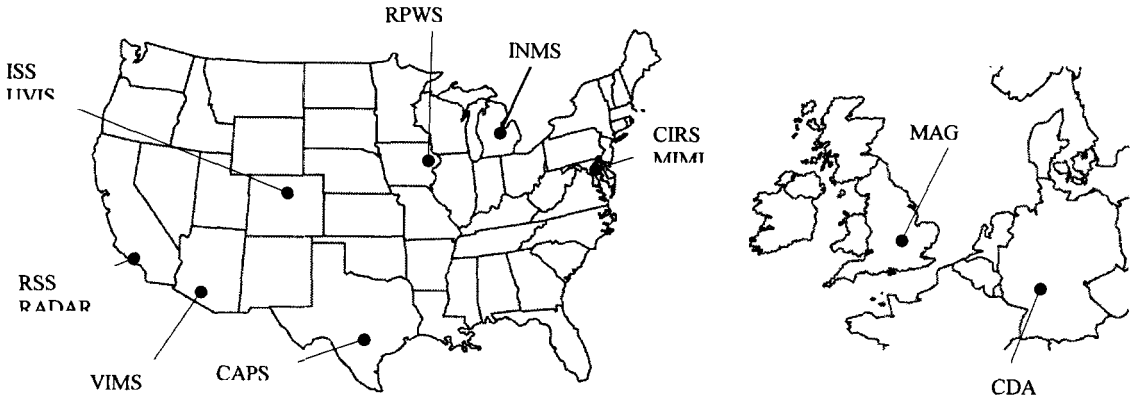
[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA.
[2] *New Mexico State University, Las Cruces, NM 88003, USA.*

The Cassini Mission to Saturn will significantly expand our understanding of the Saturn system. The importance of this unique tour of the Saturn system will be revealed in the wealth of scientific findings obtained from the data acquired. Archiving this data in a usable form is critical for future scientific studies of this remarkable system. With NASA's current two decadal schedule of recurring visits to the individual outer planets, it is necessary to ensure access to the data for future users. A reliable Cassini archive will support a data analysis program, allow comparisons of the Saturnian system with the Jovian system, (the Galileo and Jupiter Icy Moons Orbiter (JIMO) missions), and provide a basis for future mission planning.

Mission Background

The Cassini spacecraft with the ESA Huygens Probe was launched on 15 October 1997 [1,2,3]. After nearly 7 years of cruise, Saturn Orbit Insertion (SOI) will occur on 1 July 2004. The probe will be released from the orbiter on 24 December 2004, and on 14 January 2005 the Huygens probe will descend through the atmosphere of Titan and relay data to the orbiter for up to 2.5 hours. The Cassini orbiter will tour the Saturn system for 4 years, with the baseline mission ending mid 2008, for a total mission duration of 10.7 years.

The task of archiving Cassini science data requires significant resources, development of critical timelines, and the management of twelve instrument teams who were individually selected and are distributed around the world. Developing a data archive that is consistent in quality and detail has been a significant challenge in the Cassini distributed operations environment (fig. 1). The downlinked data are cleaned and transmitted to each instrument home site where they are processed into usable formats. The individual teams have a one-year validation period to organize the data and solve data processing and calibration problems before submitting data to the Planetary Data System (PDS) for release to the science community and archiving (fig. 2). It is assumed that early science analysis of selected components of the data will be carried out as data is processed through the pipeline. Incorporating consistent standards into the various systems requires a great deal of communication and teamwork. It also requires a mechanism for distributed teams to report archiving progress to the project to facilitate upward status reporting.



CAPS	Cassini Plasma Spectrometer, *Southwest Research Institute, San Antonio, USA*
CDA	Cosmic Dust Analyzer, *Max Plank Institute für Kemphysi, Heidelberg, Germany*
CIRS	Composite Infrared Spectrometer, *Goddard Space Flight Center, Greenbelt, USA*
ISS	Imaging Science Subsystem, *Space Science Institute, Boulder, USA*
INMS	Ion And Neutral Mass Spectrometer, *University of Michigan, Ann Arbor, USA*
MAG	Dual Technique Magnetometer, *Imperial College, London, UK*
MIMI	Magnetospheric Imaging Instrument, *Johns Hopkins University, Baltimore, USA*
RADAR	RADAR, *Jet Propulsion Laboratory, Pasadena, USA*
RPWS	Radio And Plasma Wave Spectrometer, *University of Iowa, USA*
RSS	Radio Science Subsystem, *Jet Propulsion Laboratory, Pasadena, USA*
UVIS	Ultraviolet Imaging Spectrograph, *University of Colorado, Boulder, USA*
VIMS	Visible And Infrared Mapping Spectrometer, *University of Arizona, Tucson, USA*

Figure 1. Cassini Distributed Operations Environment

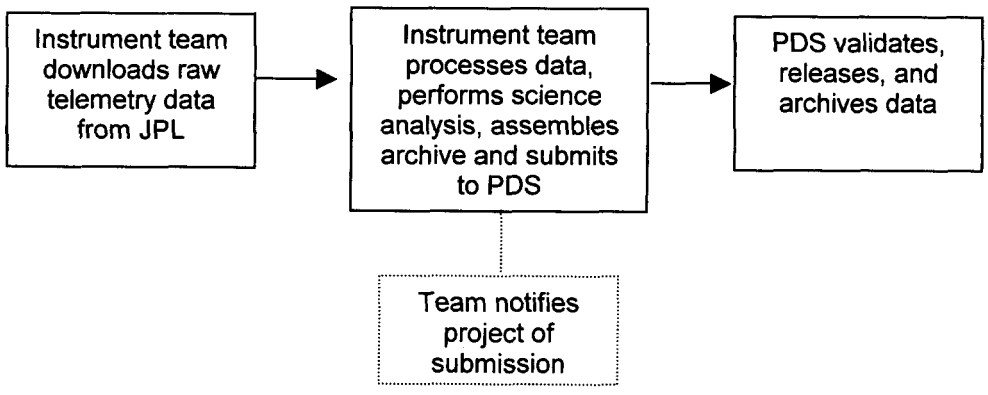


Figure 2. Distributed Operations Data Flow

Communication has been the key to the successful development of the Cassini archive system in a distributed environment. The timeline in figure 3 illustrates how archive development has progressed through the mission and the various working groups that were formed to advance development. The Cassini tour will consist of a series of passes through the Saturnian system leading to a complex observing sequence that will place huge demands on the mission staff; therefore it was deemed necessary to define data products and pipelines prior to the Saturn Approach Phase of the mission.

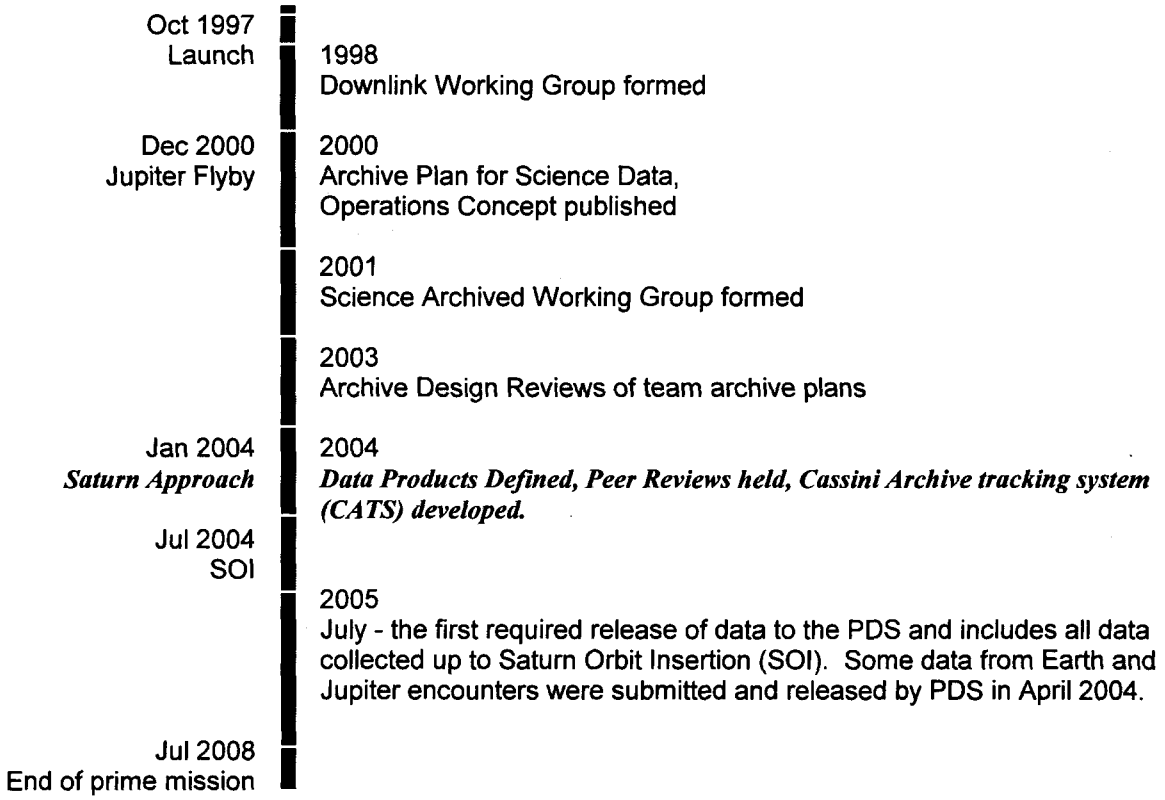


Figure 3. Archive Development Timeline

In 1998 the Cassini Downlink Working Group (DWG) began meetings with members from the Cassini operations system team, instrument teams and a PDS representative. The group began preliminary discussions to define an Archive Generation and Validation Operations Concept, and finalize requirements that were later incorporated in the Cassini Functional Design Document [4]. The operations concept defines a direct interface between the instrument teams and the PDS to develop, produce, and deliver archive datasets. Teams are required to archive raw and calibrated data, or raw data and calibration files and algorithms that can be used to produce calibrated data. The DWG forum was also used to introduce instrument teams to PDS standards and requirements. Doing this early ensured that the operational cost of archiving data with the PDS was understood and could be budgeted. It was equally important that PDS understood the volume and types of data returned by instruments so they could plan for the operations of archiving, preserving, and distributing this data to the science community. The DWG also developed the

Cassini/Huygens Archive Plan for Science Data (APSD) [5], which was signed by the Project, PDS, and NASA HQ in the year 2000. The APSD defines roles and responsibilities, the validation concept, the datasets teams plan to archive and provides data volume estimates. The work done in the DWG laid the foundation for a more concentrated development effort that began in January 2001 with the formation of the Science Archive Working Group (SAWG). The SAWG was chartered to complete the development of the archive system and included representatives from the project, each instrument team and appropriate PDS discipline nodes. The working group concentrated on providing instrument archiving teams with PDS standards [6] training, defining Cassini standards for representing parameters such as time and coordinate systems, and developing the master index keyword parameters. Archiving teams worked with the PDS nodes that specialize in their science discipline to define data products and archive volume layouts in interface specification documents. Over 100 types of data products have been defined and an estimated 10 Terabytes of data will be archived.

By January 2003 preliminary archive design reviews began, and each instrument team's data product definitions, data formats, label files, documentation, and validation plans were scrutinized by a review board with membership from PDS and the science user community. The reviews resulted in nearly 100 action items that have been addressed by the project, instrument teams, and the PDS. These detailed archive design reviews, proved to be an effective means for communicating archive designs and identifying problem areas. As a result, data formats were clarified, documentation plans improved, and requirements better understood. Teams gained the knowledge to focus on the areas, (outlined in the below table), that required the most attention and would have the greatest influence on the usability of data.

DOCUMENTATION	of the mission, spacecraft, instruments and datasets need to provide a detailed history of the mission and instrument capabilities.
STANDARDIZATION	of data formats - to ensure long-term access, data formats should be as simple as possible.
PRESERVATION	of ancillary and calibration - retrieval of pointing information and calibration need to be planned early and accomplished in a monitored incremental process.
VALIDATION	of data products – validation for science content is best accomplished by distributing data products in their archive form to the science team for analysis.
OPTIMIZATION	for correlative studies - coordination among experiments to define standards for data products, indexing and documentation to ensure correlative studies are supported.

DOCUMENTATION

Providing appropriate context for the data is critical to understanding how that data can and cannot be used. Several types of documentation are important for understanding archived data. These include: the mission plan, spacecraft subsystems descriptions, instrument design and operations descriptions, and dataset characteristics. Maintaining the mission plan with details of the mission as it progresses is crucial to understanding the possible contents of datasets. The mission plan should be periodically updated to reflect changes in the mission and science objectives, tour geometry, and spacecraft and instrument operations. Spacecraft subsystems, individual instrument assemblies and hardware components, instrument orientations and interfaces with the spacecraft as well as operational capabilities and instrument modes should be thoroughly described. Detailing instrument component characteristics is important because specific hardware technology may come into play when analyzing data and initial documentation may be lost for outdated instrumentation. Performance specifications of instrument components and individual detectors should be accompanied by a history of operational capabilities.

Two important aspects of dataset documentation involve data format specifications and calibration algorithms. Datasets must be described at the bit level in interface specifications and human-readable calibration/reduction algorithms must be provided to ensure that data access and algorithms can be implemented in future programming languages and computing environments. The amount of detail that is needed in the dataset documentation is usually underestimated and requires significant resources to produce.

In addition to data formats and calibration algorithms, specific efforts should include the maintenance and preservation of an observer's notebook. The observer's notebook should include science goals associated with specific observations, unplanned data outages, processing anomalies, instrument anomalies and their effect on the data, and other ongoing errata information pertaining to the dataset.

Cassini uses a centralized system to capture high level science goals associated with observations during the science planning process so that science goals are associated with resource usage. These data will be archived as a set of searchable Event SPICE [7,8,9] files that could be combined with the observer's notebook to relate desired science goals to data.

Including detailed documentation that describes the mission plan, spacecraft subsystems, instrument design and operations, and characteristics of the dataset are imperative to understanding the dataset and the amount of detail needed is typically underestimated.

STANDARDIZATION

Complicated data structures that require specialized software for access are problematic for long-term archives. If data are formatted for use in an environment that requires specific software they tend to be proprietary in nature and their accessibility is inherently short-lived. Because individual teams tend to rely on familiar formats, (and they have strong opinions of how databases should be designed), it is important to emphasize standardization of formats among instruments and across missions, if possible, long before

data processing systems are conceived. For most non-imaging instruments, tables or time series ASCII formats provide the most flexibility. However, many modern instruments involve various modes of spatially resolved, hyper-spectral imaging, yielding data that are best stored in arrays with multiple back planes (images and cubes). For higher-level spatially resolved products it may be convenient to store reduced geometry parameters related to each pixel in additional back planes, further complicating the data, but greatly enhancing its ease of use. Consideration of the possible complexity of these datasets illustrates the challenge in establishing standards. However, by using formats that meet PDS standards, developers can ensure their data will be accessible using PDS maintained tools.

In a mission like Cassini that involves remote sensing, (CIRS, ISS, RADAR, RSS, UVIS, and VIMS), and in situ measurements, (CAPS, CDA, INMS, MAG, MIMI, RPWS), (fig. 3), correlative keyword parameters are not common to all instruments. Time of observation is the most fundamentally shared parameter and should be consistently expressed among all datasets. Other parameters, such as latitude and longitude ranges, may be shared among remote sensing instruments and should be consistently represented and formatted.

An example of data format standardization that has worked very well is the PDS Navigation Ancillary Information Facility (NAIF) SPICE System [7,8,9]. The SPICE system defines data file formats for storing spacecraft and planetary ephemerides, spacecraft pointing, planetary constants, and instrument field-of-view and mounting offsets. All planetary missions are now using SPICE files to store and distribute these ancillary data. NAIF also maintains a toolkit in many languages to access these files in most computing environments. Usage of the NAIF SPICE system ensures common geometric parameters can be incorporated into the dataset.

It is necessary to define standard formats that are adaptable enough to accommodate different types of data and to ensure that keyword parameters are consistently represented among instruments so that datasets can be associated with one another based upon common temporal and spatial coverage.

PRESERVATION

The process for creating and distributing ancillary data such as navigation and spacecraft pointing was planned early in the mission so that the schedule of availability could be factored into data processing schedules. Operational designs such as the order and frequency of downlinked data and length and position of individual orbits, within the Saturnian system, affect the production schedules for creating ancillary data and must be considered when establishing requirements on the data archiving schedule. Spacecraft navigation, pointing reconstruction, and planetary orbital elements will be updated periodically as the mission progresses so they can be used as inputs to the data processing pipeline and be archived with science data [7,8,9]. One of the most difficult problems for a user of data after it has been archived is to determine the observing geometry for a specific data file. This emphasizes the critical need to ensure ancillary geometric data is included with datasets as parameters within data files for in situ experiments and within keyword parameters for remote sensing experiments. However, improvements of these geometric ancillary data through optical navigation to improve pointing knowledge and ephemeris

refinement efforts may not be available at the time of science data archiving. The preservation of these ancillary data and inclusion of documentation and subroutines for producing typical parameters by the PDS NAIF node will serve the science community for the long term.

Calibration is an area that continues to be refined as the mission progresses. The importance of preserving early pre-flight calibration and associated documentation cannot be stressed enough since these can be used a baselines for future calibrations. As calibration techniques progress and become consolidated throughout the mission, the algorithms used to calibrate data and any calibration files or models will be completed. It is important to preserve a calibration history that describes temporal changes in instrument performance and identifies which calibration scenarios apply. Without this critical calibration information, data may be unusable.

Project and instrument team vigilance is required to maintain ancillary data and calibration information. Without the proper archiving these data, datasets may be short-lived and have limited usability.

VALIDATION

Cassini 's validation concept implements three types of validation. These include: sample volume review, PDS peer review, and production validation.

A sample volume, created prior to pipeline production, can be used to validate the functional interface between the instrument team archivist and the PDS discipline node partner. The exercise of producing a sample volume, which can be an iterative process and may or may not include science data, provides the science team archivist experience in assembling a PDS compliant archive and the PDS partner experience with accessing the data files and understanding the archive volume layout. The sample volume has proven to be an invaluable tool for developing and validating the archive design.

Once the archivist is ready to produce a complete sample volume for external review, PDS coordinates a peer review of the dataset. A PDS peer review includes members of the science community that are internal and external to the team and function as the initial validation that a dataset can be used to perform science analysis. Periodic PDS reviews of the documentation should be held during the systematic production of datasets to verify temporal changes such as instrument performance, data acquisition schedules, observation plans, and mission extensions are reflected accurately. A well-maintained observers notebook can be used to check that the documentation represents the content of the dataset and verify that the team acquired the data planned.

The data processing pipeline should be constructed so that team members use the data files destined for archive to perform preliminary science analysis. This may be considered the most important form of validation since it proves the usability of data. Team members can provide feedback to the archivist that helps ensure the data content is optimized for science analysis. Examples of feedback include: data confidence statements, requests for format changes, additions to the dataset, and calibration changes. Software can be developed to verify that data products conform to defined formats; however, the development of consistent documentation and methods of reliable calibration requires a strong link between the science team and the archivist.

Pre-production sample volumes and peer reviews have been used to validate archive designs. However, the most important validation is obtained by active use of the dataset to perform science analysis during the mission, which ensures its usability once the data are archived.

OPTIMIZATION

The ultimate goal is to integrate each Cassini dataset into the PDS archive. Integration requires adequate long-term standardization that allows users to carry out correlative studies utilizing multiple instruments, possibly from multiple missions, to address a variety of scientific questions. PDS has developed long-term standards that define the keyword parameters needed to support cross correlative studies in disciplines such as atmospheres, surface morphology, and magnetic field and particle interaction. These keyword parameters map to attributes in the master index. The master index is the PDS database that enables a user to search for and access only those datasets that satisfy the search criteria. Early in the planning phase, the PDS discipline partner must work with the instrument archivist to ensure that keywords, data formats and map projections (if applicable), coordinate systems [10], and time representation are well understood and consistent with PDS standards. As the mission progresses, the archiving team should populate the master index with parameters that define temporal, spectral, and spatial resolution and coverage as well as illumination and viewing geometries.

Including the correct set of keyword parameters with datasets to provide a useful master index ensures that users can perform discriminating searches that limit data downloads and correlates datasets across experiments.

In summary, developing a quality archive of the Cassini data has been challenging in a distributed operations environment for a mission with a broad range of experiments. Communication through regular working group meetings has been the key to success for developing a cohesive system and identifying solutions to problems. Science teams now have mature designs for producing PDS compliant archives. Although, some development to implement those plans into operational systems still remains, the design is solid, and in many cases has been validated. Some teams have already produced archive volumes that have undergone PDS peer review, using data acquired at Earth or Jupiter, which demonstrates that their system is ready for pipeline production. There have been many lessons learned with respect to documentation, standardization of data formats, the preservation of ancillary data, and optimization of datasets, which we have detailed in this paper. And, although the Cassini tour is composed of a series of unique orbits, each utilizing the instruments differently, the archiving challenge is not unique. Constructing an archive that provides prompt and standardized data access is a task faced by all NASA and ESA missions. The length of time between a mission conception and final completion of the dataset ensures that there will be staff turnover, new budget constraints and various instrument and spacecraft anomalies. In addition, many of the current missions include new instrument groups who are participating in a mission for the first time, which requires more efficient archive training procedures. The lessons learned by Cassini will be used to improve PDS standards and the archive planning process for future planetary missions.

Acknowledgment: This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References:

[1] C. Russell, Ed. SSR, Vol 104, Nos 1-4 (2002)

[2] D. Seal, Cassini Mission Plan, JPL D-5564 (Internal Document), Revision O (October 2003)

[3] R. Mitchell, The Cassini/Huygens Mission, 54th International Astronautical Congress, The World Space Congress (2003)

[4] Cassini Functional Design Document, JPL 699-500-3-GS/D (Internal Document), Rev C, Change 13, 22 July 2003

[5] D. Conner, Cassini / Huygens Program Archive Plan for Science Data Initial Release Version 1, JPL D-15976 (Internal Document), (August 3, 2000)

[6] Planetary Data System Standards Reference, Version 3.6, JPL D-7669 (Internal Document), Part 2. (August 1, 2003).

[7] Navigation Ancillary Information Facility SPICE toolkit download web site ftp://naif.jpl.nasa.gov/pub/naif/toolkit

[8] Navigation Ancillary Information Facility SPICE tutorial ftp://naif.jpl.nasa.gov/pub/naif/tutorial

[9] Cassini anonymous ftp site for downloading SPICE files ftp://naif.jpl.nasa.gov/pub/naif/CASSINI/kernels

[10] Seidelmann, et al., Celestial Mechanics and Dynamical Astronomy, 82, pp. 83-110 (2002).