Pergamon

*Planet. Space Sci.*, Vol. 44, No. 1, pp. 55–64, 1996

0032–0633/96 $15.00+0.00

0032–0633(95)00106–9

# The Planetary Plasma Interactions Node of the Planetary Data System

**Raymond J. Walker,[1] Steven P. Joy,[1] Todd A. King,[1] Christopher T. Russell,[2] Robert L. McPherron[2] and William S. Kurth[3]**

[1] Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024–1567, U.S.A.
[2] Department of Earth and Space Science and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024–1567, U.S.A.
[3] Department of Physics and Astronomy, University of Iowa, Iowa City, IA 52242–1479, U.S.A.

Received 20 November 1994; accepted 26 July 1995

**Abstract.** Five years ago NASA selected scientists at UCLA to form the Planetary Plasma Interactions (PPI) Node to help the scientific community locate, access and preserve particles and field data from planetary missions. Since planetary plasma data are varied and require expertise in many areas the PPI Node is distributed with an Outer Planets Subnode at the University of Iowa, an Inner Planets Subnode at UCLA and a Radio Astronomy Subnode at GSFC. The PPI Node has tried to serve the science community by providing them with high quality data products. It has worked with missions and individual scientists to secure the highest quality data possible and to thoroughly document it. The PPI Node has validated the data, placed it on long lasting media and made sure it was properly archived for use. So far it has prepared and archived over $10^{11}$ bytes of data and has produced 171 CD-ROMs with peer reviewed data. In so doing an efficient system has been developed to prepare and archive the data and thereby enable to steadily increase the rate at which the data are archived. Although the PPI Node produced a substantial archive during the initial five years, it has an even larger amount of work in progress. This includes preparing CD-ROM data sets with all of the *Voyager*, *Pioneer* and *Ulysses* data at Jupiter and Saturn. It is also completing the *Pioneer Venus* data restoration. The *Galileo Venus* archive and radio science data from *Magellan* will be prepared early in 1995. It is assisting the Small Bodies Node of PDS in the preparation of comet data and with the preparation of data from the comet Shoemaker–Levy 9 collision with Jupiter. Asteroid data from *Galileo* will also be archived. In addition to providing the data, users have been provided with software tools to manage and read the data which are computer, operating system and data format independent. Scalable systems have been developed so that the same software used to manage and access the data for the entire PPI Node can be used by individual investigators to manage the data on a single CD-ROM thereby greatly reducing the software development effort for both the PPI Node and users. This software is delivered with the disks. The PPI Node data holdings are available over the Internet. They can be accessed through the World Wide Web (WWW) at Universal Resource Locator (URL) http://www.igpp.ucla.edu/ssc/pdsppi/Welcome.html. Users without a WWW browser can use the WWW interface by signing into the host "pdsppi.igpp.ucla.edu" as "pdsuser".

*Correspondence to*: R. J. Walker

## 1. Introduction

For over 30 years space plasma physicists have studied the interaction of the Earth's magnetosphere with the solar wind. Although in this process we have developed a rather detailed understanding of many of the phenomena involved, the bulk of our knowledge has been limited to a rather narrow range of parameters that occur naturally at the Earth. More recently we have been challenged to test that understanding by applying our ideas to the vastly wider range of parameters involved in the interaction of the solar wind with other planetary magnetospheres. At Mercury, Jupiter, Saturn, Uranus, and Neptune the interaction is to first order similar to the Earth's in that a super Alfvenic magnetized plasma impinges on a planetary magnetic field forming a magnetospheric cavity. However, there are many important differences and by studying these differences across the suite of planets, we learn more about how planetary magnetospheres work. Also planetary magnetospheres provide a variety of scales over which the processes can take place from the magnetosphere of Mercury 1/20th of the radius of the Earth's magnetosphere to that of Jupiter close to 100 times larger than the Earth's magnetosphere.

At Venus and Titan (when outside of Saturn's magnetosphere) an ionosphere interacts with the flow to form an induced magnetospheric cavity. At Mars too the interaction is most probably between the solar wind and the ionosphere but Mars may have a very weak dynamo field or a remanent field like that of the Moon. The interaction between a comet and the solar wind is different yet because of the small mass of the comet and the large amount of gas that evaporates from its surface. Recent observations near asteroids indicate that they too may have a strong interaction with the solar wind. Io's interaction with Jovian plasma presents a unique case in two respects. First, Io is embedded in a sub-Alfvenic flow and second the plasma impinging on Io was originally of Ionian origin. Finally, the other planetary moons interact with plasma displaying a broad range of interaction parameters.

From these studies of the interaction of space plasmas with extraterrestrial bodies, planetary plasma physics, a new scientific discipline concerned with the physics of magnetospheres and ionospheres and the interaction between various plasmas has evolved. A major goal of planetary plasma physics is to obtain an understanding of the interaction between the solar wind and planetary obstacles (magnetospheres, ionospheres, or even the planet itself) and the processes which occur within the resulting induced or intrinsic magnetospheres. Some of the phenomena studied include magnetospheric dynamics and the aurora, wave–particle interactions and processes taking place in planetary magnetotails. Planetary plasma physics is concerned with the flow of energy and mass throughout these systems. It also is concerned with the interaction between a planetary wind and satellites. While planetary plasma physics is a new discipline, it is not one that is isolated from the other planetary disciplines. Planets are closely coupled systems. Planetary atmospheres evolve due to plasma processes. Planetary surfaces are bombarded by energetic particles and flowing plasma. The interiors of planets can be probed with low frequency magnetic variations.

Several different instrument types are used to provide the many physical parameters required in planetary plasma studies. These include magnetometers and electric field instruments to provide low frequency magnetic and electric field observations. A variety of energetic charged particle instruments provide charged particle (electrons and ions) fluxes and composition in energy ranges from a few keV up to GeV. Plasma detectors provide information about the plasma distribution function and composition for cold and thermal energy particles. Plasma wave instruments and AC magnetometers provide information about plasma waves in the magnetospheres. In addition to these *in situ* observations space plasma physics makes use of remote sensing instruments for radio astronomy measurements, for ultraviolet images of planetary magnetospheres and for images of aurora. Even radio signals are used to probe planetary ionospheres.

The number and variety of planetary plasma physics data sets are large. Five years ago the Planetary Data System (PDS) selected the Planetary Plasma Interactions (PPI) Node at UCLA to help the scientific community locate, access and preserve these data. The data for a given scientific study in planetary plasma physics frequently come from several instruments on a spacecraft as well as ground observations. Since the data from these instruments tends to be distributed at many institutions around the country and around the world, the PPI Node was organized as a distributed node. In addition to the primary node at UCLA there are three subnodes located at the University of Iowa (W. Kurth, subnode leader), the Goddard Space Flight Center (M. Kaiser, subnode leader) and at UCLA (C. Russell, subnode leader). (Note, in 1995 we plan to add a Radio Science Subnode at Stanford University (R. Simpson, subnode leader) and in 1996 the GSFC subnode will phase out its activities after completing the archive of *Voyager* and *Ulysses* Radio Astronomy data.) The subnodes provide scientific expertise on the data, participate in node data selection activities and help with data restoration and documentation activities.

In this paper we describe how the PPI Node is striving to provide the planetary science community with access to high quality data products and how we preserve the data for future use. In Section 2 we describe how we work with planetary missions and individual investigators to secure and prepare high quality data products. We discuss both the archiving of existing data and our interaction with missions. In Section 3 we review the present data holdings and describe the work in progress. We discuss the services available for the planetary science community in Section 4. The organization and distribution of a large quantity of data of varied type from many sources present a major data management challenge. In Section 5 we present a new approach for managing data inventories and describe how we use this approach to manage the varied data useful in planetary plasma physics studies. Finally we discuss some of our future plans for the planetary plasma interactions archive.

## 2. The PPI approach to data archiving

Before we can make data available to the scientific community, we must obtain the data, document it, validate it and prepare archival volumes. The first step in the archiving process is to determine what data are available and to learn about them. Particles and fields data come mainly from instruments developed by a principal investigator working with a small team of engineers and scientists rather than from large facility class instruments (usually remote sensing instruments) which have large teams of scientists and engineers. This means that PPI Node members must first meet with each team separately to determine data availability and location and to learn about the instrument and how the data were processed. Next we design the archival volumes, document the data, and begin the process of validating the data. The volume design process includes organizing the data on the volume, documenting both the data and the instrument that produced it and populating the inventory tables (see Section 5) that

**Major tasks in the preparation of archival data**

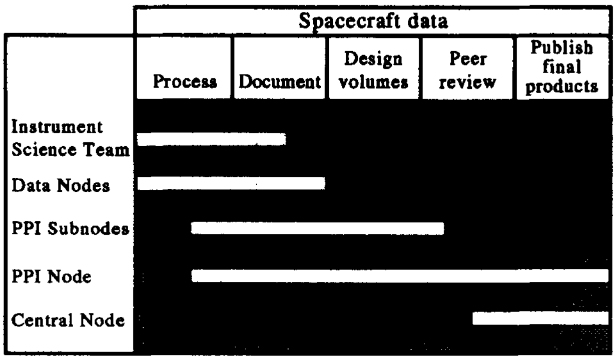


**Fig. 1.** The tasks required for archiving planetary data with the PPI Node. The main teams involved in preparing the archival data are listed on the left and the tasks which they are involved in are indicated by the white bars

can be used to help locate and read the data. The main tool for validating data is the peer review process whereby members of the scientific community evaluate the data (both the science data and the documentation) in much the same way that they peer review an article for publication or a research proposal. When problems are found they are corrected and the process is repeated until the data complete the peer review process. Finally we "publish" the data by making them available to the general scientific community both over the Internet and on CD-ROM. Figure 1 shows the main tasks in the process of archiving the data. The contributions of each team are noted.

Missions such as *Pioneer Venus* or *Pioneer* 10 and 11 or *Voyager* which were flying before PDS present a unique challenge for data archiving. By the time PDS came along they had already started their archival process and usually had little or no resources to change systems in mid-mission. The investigators on these missions have a commitment to deliver data to the National Space Science Data Center (NSSDC) not PDS. The first step in working with these missions is to explain what PDS can do to help them build better archival products. We explain that we can help them build well documented data products on stable random access media (CD-ROM and write once compact disk [CD-WO]). As investigators do research with their data they learn more about the data and refine the data products. We point out that by rearchiving their data with PDS they can update the archive with the latest version of the data. We explain that for existing mission restorations PPI does all of the design work and prepares the documentation. The investigators are asked to concentrate on providing the best possible data and to help check it and the documentation. Finally we ask the investigators to voluntarily submit their data. So far on *Pioneer Venus*, *Pioneer* 10 and 11 and *Voyager* only one team has chosen not to archive their data with PDS.

After the initial contact the PPI team starts to prepare the metadata (data describing data) for the volumes starting with the instrument descriptions and a rough draft of the data descriptions. We meet again with the science teams to refine the data descriptions and arrange for the delivery of data of PDS. Once the data are available the volume design effort quickens. As soon as possible the data are placed on-line. This gets the data into the science community quickly since the remaining steps in the data archiving process can take a year or more. Since the data archiving process is incomplete these data are labelled as "in review" and users are warned that we have not completely checked the data. Comments from users of the on-line data are encouraged. When the volume design is complete we produce CD-WO volumes and send them out for peer review.

Members of the peer review panel are selected by the Node manager. They check both the science data and the documentation. The peer review process can take several months. Most volumes need to be corrected before they can be sent out for final reproduction. We work with the investigators to correct the problems encountered by the peer review panel and the on-line data users. If the problems encountered in the peer review are serious then the volume may require major modifications and the revised volume must be sent out again for another review. After the volume finally completes peer review it is sent to the PDS Central Node for a final check to make sure that it conforms to PDS standards and has no technical problems. It is then sent to the vendor for mastering and production. For data which is published only on CD-WO, the final production is done at UCLA. Copies of all the CD-ROMs are sent to the NSSDC for safe keeping and distribution through their system.

We also publish software to manage and extract the data with the CD-ROMs as well as any special software provided by the investigators. For multiple disk data sets the CD-ROMs are numbered sequentially. The software is published on CD-ROM number 1 in the series. Even with the most careful checking mistakes will be made on the CD-ROMs. We publish an errata containing information about the known errors. It too is included on CD-ROM number 1 which incidentally is the last disk of the series published. Until CD-ROM number 1 is published the software and errata are available through the PPI Node anonymous FTP account.

For data from instruments in their areas of scientific expertise, the PPI Subnodes are involved in all of the data archiving functions except preparing the final volumes.

For restorations after the mission has ended the contact is with the individual experimenters directly. Frequently the investigators are no longer funded to process the data. In some cases small amounts of data node funding have been effective in helping the investigators restore the data. With the aid of PPI Node or Subnode personnel, the data nodes process the science data and prepare the instrument and data descriptions.

For new missions we try to start discussion about the archive early. We encourage the mission personnel to use PDS standards for the data right from the start. For raw data (i.e. at the experimenter data record (EDR) level) we encourage missions to build PDS compatible data volumes for distribution to the investigators. We work with the mission teams to prepare the metadata early and point out that once the metadata are in place it is relatively easy to submit multiple versions of the science data products which incorporate the lessons learned from ongoing research. We also work with mission personnel to prepare mission archive plans. Once data are delivered to PDS the verification procedure is the same as for the older missions.

## 3. The PPI data archive and work in progress

During the past 5 years we have prepared and archived over $10^{11}$ bytes of data. As of March 1995 the PPI Node data holdings included 232 datasets from 74 instruments on 11 spacecraft and consisted of over 160,000 individual files.

The PPI Node development staff, along with PDS system developers at JPL, have been working on a set of standards for archiving data on CD-ROMs (McMahon, 1996). The initial PDS standards were optimized for imaging datasets and did not include the flexibility required for planetary particles and fields data. Space plasma datasets typically consist of time series data and are much smaller than their imaging counterparts. The data from different instruments frequently are analyzed collectively rather than individually in order to properly interpret the physics. The *Voyager* 2 Neptune data were selected as the first data to be prepared to the new standards for non-imaging data. After many re-engineering efforts, in February 1992 the PPI Node issued a CD-ROM containing *Voyager* Neptune data compliant with PDS version 3.0 standards (PDS, 1993). After this we were able to rapidly engineer numerous data volumes and begin the task of archiving the vast stores of planetary plasma data on long lasting media. So far we have published 171 CD-ROM and CD-WO data products.

### 3.1. *Outer planets*

Table 1 lists the PPI Node outer planets data holdings as of March 1995. In addition to the *Voyager* Neptune archive, PPI produced a Uranus CD-ROM in June 1993. The *Voyager* and *Pioneer* 11 Saturn data are available on-line and the CD-ROMs will be available in mid-1995. The data from all five Jupiter encounters are currently being prepared with publication scheduled for early 1996 in time to support analysis of *Galileo* data from Jupiter. Much of the Jovian data is currently available on-line.

**Table 1.** Outer planets data holdings

| Planet | Spacecraft | Instrument[a] |
|---|---|---|
| Jupiter | *Pioneer* 10 | HVM, GTT, TRD, CRT, UVphoto, Orbit-Attitude, SPICE, Radio Science |
| | *Pioneer* 11 | HVM, GTT, FGM, Radio Science, TRD, CRT, CPC, Orbit-Attitude, SPICE |
| | *Voyager* 1 | LECP, MAG, PLS, PRA, PWS, Position, SPICE |
| | *Voyager* 2 | LECP, MAG, PLS, PRA, PWS, Position, SPICE |
| | *Ulysses* | COSPIN, GRB, DUST, SWOOP, HISCALE |
| Saturn | *Pioneer* 11 | HVM, GTT, FGM, Radio Science, TRD, CRT, PA UVphoto, CPC, Orbit-Attitude, SPICE |
| | *Voyager* 1 | CRS, LECP, MAG, PLS, PRA, PWS, SPICE, Position |
| | *Voyager* 2 | CRS, LECP, MAG, PLS, PRA, PWS, SPICE, Position |
| Uranus | *Voyager* 2 | CRS, LECP, MAG, PLS, PRA, PWS, SPICE, Position |
| Neptune | *Voyager* 2 | CRS, LECP, MAG, PLS, PRA, PWS, SPICE, Position |

[a] Please see the glossary in the Appendix for a complete list of instrument acronyms.

### 3.2. *Inner planets* (*Venus*)

The inner planets data archiving effort has focused on Venus datasets (Table 2) with emphasis on the *Pioneer Venus Orbiter* (PVO) data. PVO was actively acquiring data at Venus for nearly 14 years before the orbiter descended into the Venus atmosphere. Over a 2 year period, all of the instruments which fall into the scope of the PPI Node have submitted data for the entire mission (5055 orbits). So far we have produced 67 of 70 CD-ROM volumes of Radio Science original data record (ORD) data and 16 of 17 volumes of reduced data record (RDR) data from the Retarding Potential Analyzer (OPRA) experiment. Magnetometer and Electric Field Detector data are being processed jointly and we have completed 52 of $\sim$75 volumes of processed data from these two experiments. The first of 12 supplemental experimenter data record (SEDR) volumes submitted by the MAG team has been produced. All processed data from the Ion Mass Spectrometer (OIMS—2 volumes), Electron Temperature Probe (OETP—2 volumes) and Neutral Mass Spectrometer (ONMS—2 volumes) experiments are currently in peer review. The bulk of the PVO data will complete the archival process in early 1995. This will include data from the Plasma Analyzer (OPA) and the Gamma Burst Detector (OGBD) in addition to the instruments already mentioned.

The only data from the MGN project which are relevant to the PPI Node of PDS are some of the Radio Science products. The solar wind radio scintillation original data records (ODR) are being preserved.

Data archiving from the *Galileo* Venus encounter has begun. Members of the PPI Node are working with the *Galileo* Data Archive Working Group (DAWG) to prepare data from the Plasma Wave Subsystem (PWS), Energetic Particle Detector (EPD), Magnetometer (MAG) and Plasma (PLS) instruments.

### 3.3. *Comets and asteroids*

The primary responsibility for the archive of data from comets and asteroids lies with the Small Bodies Node (SBN) of PDS. However, many of the observations of these bodies and their interaction with the solar wind are made *in situ*. For these datasets, the PPI Node has an obligation to assist the SBN with the data archiving

**Table 2.** Inner planets data holdings

| Planet | Spacecraft | Instrument[a] |
|---|---|---|
| Venus | *Pioneer Venus Orbiter* | OMAG, OEFD, ORPA, EPHEM, ONMS, ORSE, SEDR, OETP, OIMS, OPA, OGBD |
| | *Galileo* | EPD, MAG, PLS, PWS, EUV, DDS, HIC |
| | *Magellan* | Solar wind radio scintillation data |

[a] Please see the glossary in the Appendix for a complete list of instrument acronyms.

process. The actual working agreement between the two nodes is decided on a case-by-case basis. At present, the PPI Node is assisting the SBN (lead node) in the documentation preparation and data validation for the Johnstone Plasma Analyzer data from the *Giotto* spacecraft encounter with the comets Griggs–Skjellerup and Halley.

We are working on archiving the *Vega* data from comet Halley. The Inner Planets Subnode is helping to prepare the magnetic field (MISCHA) archive. We are also working with the SBN to improve the documentation in the plasma wave data archive.

On October 29, 1991 and August 28, 1993 *Galileo* flew by asteroids 951 Gaspra and 243 Ida, respectively. PPI Node members have been working with the *Galileo* DAWG to prepare the data for archiving.

### 3.4. *Earth*

PDS usually does not archive *in situ* data from the Earth, however one exception is the data from the two *Galileo* fly-bys. By so doing the entire *Galileo* archive will be prepared in the same way. Because of *Galileo's* high speed trajectory the data from the Earth has provided space physicists with a unique view of our planet not available from orbiters. Data from the Magnetometer, Energetic Particle Detector, Heavy Ion Counter, Plasma Wave Subsystem and Plasma Instrument have been received. The PPI Node works closely with the Space Physics Data Systems (SPDS) who are responsible for archiving data from the ionospheric, magnetospheric, heliospheric and solar missions of the NASA Space Physics Division. For instance members of the PPI Node actively consult with SPDS and will make the availability of the *Galileo* Earth data known to the space physics community through SPDS.

## 4. Service to the scientific community

The PPI Node serves the science community in several ways. The first and most common of these services is providing data to researchers. In the case where the data requested exists within the PDS, even if it does not reside within the PPI Node, the request can be filled quickly. At times the data which a particular investigator requires for a research project does not exist within current PDS data holdings. In these instances, the PPI Node team tries to determine if the desired data exist and whether or not they can be made available to the requester. If the data requested is available and can be well documented easily, the PPI Node will acquire the data quickly and make it available to other researchers as well. Sometimes data have not been archived because complex instrument response functions make processing of the data difficult and subject to scientific debate. In these cases, we understand the need for close interaction between the expert and the data user and try to get the requester in direct contact with the investigator. Later we will work with the investigator to obtain the data and place it in the archive along with documentation describing the controversial elements.

The next most common type of service that the PPI Node provides to the science community is additional data processing or manipulation. The PPI Node receives many requests for data in a form other than the original archival form. We respond to these requests as resources are available. One of the most common requests is to have vector data transformed into another coordinate system. Another common request is for data file reformatting, conversion to ASCII, or translation from one binary representation to another. The PPI Node has both the scientific knowledge and the computing skills to perform these tasks easily and quickly when they are requested.

If there is sufficient demand we create software tools to assist researchers in accessing and manipulating data (see Section 5). For instance we now provide software which performs the most frequently requested coordinate transformations. A distributable software package called DET (see Section 5) enables researchers to access data and documentation from all the PPI CD-ROMs. The PPI Node has provided another software package, SEDR2ASC, which reads the raw PVO SEDR (orbit and attitude) data files and produces flat ASCII data tables with detached header files.

## 5. Data organization and management

In this section we discuss our efforts to organize and manage the planetary particles and fields data. In particular we discuss the tools we have developed to help scientists locate and access the data available through the PPI Node. These tools are scalable so the same tools which we use to manage the data for the entire PPI Node can be used by an individual scientist to access the data on a single CD-ROM. We are working to make these tools platform, operating system, and data format independent.

As data are submitted to the PPI Node we organize them into appropriately sized collections to facilitate their "publication" on easily distributed media. Currently the media of choice is CD-ROM. In addition to the publication of the data we also must maintain as much of the data on-line as possible to provide quick access to the data. Due to the volume of the data and constraints in the amount of on-line storage available we can only maintain a limited amount of the data on-line. This results in a need to be able to quickly bring data on-line and off-line while still presenting to the individual an accurate representation of all available data holdings. This has been a challenging undertaking which has required us to develop a unique information management model. In general, our model involves a management approach which minimizes the amount of effort to generate a "product", permits the information system to be scalable in size, and adheres to PDS standards.

Our approach for handling large volumes of data is to minimize the amount of direct processing the data must undergo as part of the product generation process. This is accomplished through a cooperative effort between the PPI Node and the data provider. As part of the cooperative effort an agreement is reached with determines the specific format the data provider will use to provide the data to the PPI Node. In deciding on the specific format for a given data set we consider the impact its format has on the science community. For instance if there is a great deal of software in active use which uses the existing format we would be reluctant to change the format. We also consider the amount of effort required to reformat the data. Once a specific format is chosen resources are allocated to implement software solutions so that the data may be fully integrated into the PPI Node information system.

Since our management model requires us to deal with data in a variety of formats the data and information are managed at the inventory level. That is, we track every indivisible unit of information (i.e. every file containing information). The inventory information we maintain includes information about the storage format of the data, the type of information the data contains (i.e. raw data, sampled data, documentation, etc.), the file in which the data are stored, the curator (someone who is familiar with the details of the data), and the relationship between the data units (e.g. what documentation is available for given scientific data).

Each data "product" is a completely self-contained collection. That is, every "product" contains all the necessary documentation and inventory information required to use the science data. Since at this time each "product" is a CD-ROM, each "product" is essentially an information system in a jewel case. Larger information systems are built by using the "products" as building blocks. Since each "product" is itself an inventory item, larger information systems are built from inventories of products. The on-line information system for the PPI Node is, in its simplest form, an inventory and access system for every product offered by the PPI Node.

### 5.1. *Inventory tracking*

The management of large scale databases presents many interesting challenges. One challenge is how to track all the data regardless of its physical location. Another challenge is how to track a dataset through revisions or modifications. Yet another challenge is how to indicate associations between various data (i.e. multi-media datasets which can include data, metadata, journal reprints, etc.). The challenge of managing large scale databases is made more complex when you want to be able to distribute pieces of the entire database and allow people to assemble these pieces into new databases. We have met all of these challenges by developing a management model based on scalable data systems and by developing and deploying a software system which implements that model. The management model and software system adhere to a set of specifications we call the Distributed Inventory Tracking and Data Ordering Specification (DITDOS). The DITDOS compliant software we have developed consist of tools to assist in the assembly of the metadata, servers for delivering metadata and data in a network environment, clients to access the server, report generation tools for accounting purposes, and stand alone applications which run on multiple platforms to assist the individual users who prefer not to use the servers.

DITDOS consists of five specifications (King *et al.*, 1994a,b). The first specification defines the required metadata which must be included with every "product" in order to manage and track the contents of that "product". This is called the "inventory" information and it is contained in a series of tables. There are three types of inventory information: databases, products and datasets. A database describes collections of products (Table 3). Databases are usually defined by spacecraft and planet (i.e. PVO data from Venus or *Voyager* 2 data from Neptune). A product is a collection of datasets which are contained on a single deliverable medium (i.e. CD-ROM) (Table 4). In the product table, the connection to the database is maintained through the dbname parameter. A dataset describes collections of physical data holdings and includes information about the data curator and access privileges. Datasets are defined by the investigators of a given instrument and there may be more than one dataset from an instrument. Each physical data holding is referred to as a member of a dataset. A given data holding can be

**Table 3.** Database table

| Field name | Description |
|---|---|
| refname | The reference name for the database |
| desc | A textual description of the contents of the database |

**Table 4.** Product table

| Field name | Description |
|---|---|
| dbname | The reference name of the database which includes this product |
| refname | The external reference name for the product |
| sysname | The location within the system of the inventory tables for the product |
| desc | A textual description of the contents of the product |

**Table 5.** Dataset table

| Field name | Description |
|---|---|
| refname | The external reference name for the dataset |
| curator | The reference name for the curator for the dataset |
| desc | A textual description of the contents of the dataset |

**Table 6.** Member table

| Field name | Description |
|---|---|
| dataset | The external reference name of the dataset of which this file is a member |
| sysname | The system name (path) to the data holding which is referred to by this entry |
| type | The data content type of a member. Examples include data, document, image, animation, etc. |
| status | A description of the status of the member file. Examples include "on-line" and "off-line", etc. |
| class | The class is the name of a data storage format. Examples include ascii cr/lf, UCLA IGPP flatfile, Vicar, etc. |
| instance | The instance is the version of a format (i.e. UCLA IGPP flatfile version 2) |
| desc | A textual description of the contents of the member |

**Table 7.** Curator table

| Field name | Description |
|---|---|
| curator | The reference name of the curator for the dataset |
| name | The full name of the curator |
| inst | Institution or affiliation name |
| address | The full mailing address (street, city, state, zip and country) |
| phone | Telephone number of the curator |
| e-mail | Full e-mail address of the curator |

a member of more than one dataset. Datasets are described by the contents of four tables (Tables 5–8). Figure 2 shows the relationships between the dataset tables.

The second specification is the inventory description language which defines a portable means of describing inventories. This description language provides a simple textual description of how to construct an inventory and has a command/directive/options syntax. For example, to create the inventory entry for a dataset which will contain the *Voyager* 2, Low Energy Charged Particle data from January 25, 1986 which has been provided at 15 min averages (PDS dataset ID "VG2-U-LECP-3-DDR-T860125_AVERAGE-15M-1_0") the text would read:

**Table 8.** Privileges table

| Field name | Description |
|---|---|
| refname | The external reference for the dataset for which this set of privileges apply |
| user | The name of the user granted this set of privileges |
| host | The name of the host the user must originate from in order for this set of privileges to be granted. This entry may contain wild cards |
| priv | This field is a bit map of the privileges granted: bit 1 select; bit 2 insert; bit 3 update; and bit 4 delete |

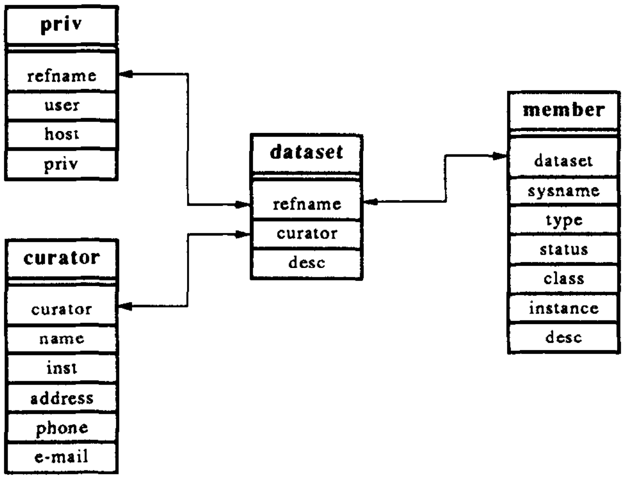


**Fig. 2.** The relationships between the dataset tables (from King *et al.*, 1994a). The table names are bold

In this example the data are stored as "UCLA/IGPP Flatfiles". This is a simple format in which the data are stored as binary tables in which all rows have the same format. Within a DITDOS inventory there is no restriction on the formats which can be supported. All that is required is that the format be uniquely identified by the class and instance entries.

The third specification is the protocol used between a client and a server. This protocol consists of tagged packets of data. There is a small set of well defined tags and request/response requirements. For example, whenever a client submits a request to the server it sends a packet containing a Structured Query Language (SQL) (Date, 1987) query which is tagged as QUERY. The server must respond with an OK tag to indicate receipt of a valid query. This is then followed with one or more DATA tagged packets. The end of the response is indicated by the receipt by the client of an END tag.

```
define dataset refname=VG2-U-LECP-3-DDR-T860125_AVERAGE-15M-1_0 \
   sysname=/LECP/AVERAGE/15M/DATA/T860125 \
   status=online curator=ditdos class="UCLA/IGPP Flatfile" \
   instance="Version 1.0" \
desc="Voyager 2 LECP 15M average data near Uranus"
```

The fourth specification defines the interface between the client or server and agents. An agent is an application which is separate from the client or server, but which is called by the client or server to perform a specific task. To the agent the client or server looks like a file so it is easy to use preexisting applications as agents.

The fifth specification is the syntax used in specifying which agents (applications) are used to perform specific operations on data in a specific format. This is a textual format consisting of a set of keyword-equals-value instructions that define the association between applications and possible actions. For example, the agents for use with data stored in a version 1.0 UCLA/IGPP Flatfile are defined as follows:

```
class = UCLA/IGPP Flatfile
instance = Version 1.0
   reader.native = ffconvert
   writer.native = ffwrite
   writer.ascii = ffwrite -ascii
   viewer.text = xffviewer
   viewer.graphic =  xffgraph
   request.offline = /usr/ucb/mail
```

This set of definitions defines the agents which will read the data in the version 1.0 UCLA/IGPP Flatfile format (reader.native), which will write the data in the same format (writer.native), convert it to ASCII (writer.ascii), list the data in a textual form on the screen (viewer.text), plot the data on the screen (viewer.graphic), and request that the data be brought on-line (request.offline).

All data handled by the PPI Node are catalogued and tracked by using DITDOS compliant inventories. In the operation of the PPI Node DITDOS is used in two ways. First, our public, networked data access system utilizes a DITDOS compliant server. This server provides a single access point to all our data holdings and tracks all requests for data. The most common method of access to the networked data holdings is through our DITDOS compliant client (described in more detail below) which provides access to the data holdings over the World Wide Web (WWW). Second, every product we generate has included with it a DITDOS compliant inventory description and pre-made inventory files. By packaging DITDOS inventory information on every product we enable individuals to use a variety of tools to browse the contents of the product and to extract selected datasets from the product for subsequent analysis.

DITDOS can be summarized as a virtual file system with strong relational database management elements. In both DITDOS and a database system data holdings are organized into databases by grouping them into datasets. However, with DITDOS a dataset is an abstraction while in a database management system the data holding and datasets are the same. DITDOS is like a file system in that a volume is self contained and self consistent. Both provide a hierarchical method for organizing the individual files. DITDOS is somewhat like the University of Minnesota's Gopher service (Anlesaria *et al.*, 1993), but has more explicit content classification and is more reliant on a database model. DITDOS provides a level of location independence much like the Wide Area Information Servers (WAIS) (Kahle and Medlar, 1991), but imposes a level of structure (databases, products, datasets, etc.) on the location independent data holdings.

### 5.2. *Tools to access and manage the data*

In December 1993 we became a World Wide Web server. The Universal Resource Locator (URL) for the PPI Node WWW home page is "http://www.igpp.ucla.edu/ssc/pdsppi/Welcome.html". (Note users without a WWW browser can sign into the host "pdsppi.igpp.ucla.edu" as "pdsuser" and access the system.) DITDOS and WWW browsers such as Mosaic work well together. Through a series of DITDOS gateway agents (DITDOS applications which serve as paths or "gateways" in WWW terminology to data) we provide users with a wide range of functions. Users can find an overview of the PPI Node, a summary of data holdings, information about Node personnel, and access to software available from the PPI Node. Most importantly users can access all of the data holdings of the PPI Node through WWW browsers. The data can be either displayed on the screen or delivered to the user's home computer via the Internet. Finally users can order CD-ROMs through the system.

A unique feature of our WWW document is that the DITDOS gateway agents generate the WWW page each time a request is received. In response to each request the agents read the DITDOS inventory tables maintained on the system and format Hypertext Markup Language (HTML) pages containing the information extracted from the DITDOS inventory. Thus the system updates itself automatically. Links to our PPI home page have been made from WWW documents at a variety of sites. Sites include the PDS Central Node (http://stardust.jpl.nasa.gov/pds_home.html), the University of Iowa Department of Physics and Astronomy Plasma Wave Group (http://www.physics.uiowa.edu) and the Space Physics Data System (http://nssdca.gsfc.nasa.gov/spds.spds.html).

The construction of DITDOS inventories is made easy through the use of tools to automate or streamline the process of generating the inventories. First we organize a product by placing the data files in appropriate locations in accordance with PDS standards for volume design. Once the product is fully assembled, we run the inventory building tool. This tool scans the volume and utilizes the information which is contained with the data and metadata (i.e. PDS labels), and the actual location of the data within the volume, to construct an inventory. The result is a DITDOS compliant inventory description. Therefore, the generation of the DITDOS inventory descriptions can be accomplished automatically from PDS compliant data.

We also have implemented and deployed a stand alone, menu driven, text based interface (DET). Through a set of successive menus an individual can use DET to explore the contents of an inventory and selectively extract data.

During the extraction process DET can convert the binary representation of the data to a variety of machine dependent formats, as well as to ASCII. Great care was taken in the design and implementation of DET to insure true portability between a variety of platforms. DET is supported on the following platforms: HP-UX, MS-DOS, SPARC SunOS, SPARC Solaris, VAX/VMS and MacOS.

## 6. Summary and future directions

The Planetary Plasma Interactions Node of the Planetary Data System is now a mature working tool for planetary scientists. We have assembled a substantial database and have a system in place to aid scientists in locating, acquiring and using planetary particles and fields data. In the future our efforts will be directed almost exclusively to increasing the available data. In the near future we are planning to increase the inner planets archive with data from Mars, Mercury and the moon. For Mars we will make the Phobos data available to the community and will work with the *Mars* 96 and *Mars Surveyor* teams. For Mercury we are planning to restore the *Mariner* 10 data. For those interested in studying the moon, we will make the *Clementine* energetic particle data available and work with Geosciences Node to prepare the Lunar Prospector archive. For the outer planets we will make the *Galileo* Jupiter data available as well as appropriate data from the International Jupiter Watch (IJW). We will be assisting the Small Bodies Node in archiving data from the collision of comet Shoemaker–Levy 9 (SL9) with Jupiter. The IJW and SL9 data will include the first data from ground observations in the PPI archive. We also will assist the Small Bodies Node in archiving data from the Near Earth Asteroid Rendezvous (NEAR) mission. For Saturn we are working with the *Cassini* project to design their archival system. Finally PDS and the PPI Node are prepared to aid future *Discovery* missions in preparing their archival products.

*Acknowledgements.* The assistance of Eva Fried, Sophie Wong, Muriel Kniffin and Bruce Rezin with the data volume design and production is gratefully acknowledged. We thank Larry Granroth, Douglas Menietti and Julie Dowell for their assistance with the work at the University of Iowa. The PPI Node activities at UCLA are funded by NASA through Jet Propulsion Laboratory contract JPL 959026 while those at the University of Iowa are funded by JPL contract 957665.

## References

**Acton Jr, C. H.,** Ancillary data services of NASA's Navigation and Ancillary Information Facility. *Planet. Space Sci.* **44,** 65–70, 1996.

**Anlesaria, F., McCahill, M., Lindner, P., Johnson D. and Torrey, D.,** F.Y.I. on the Internet Gopher Protocol. Network Working Group, University of Minnesota, 1993.

**Date, C. J.,** *A Guide to the SQL Standard.* Addison-Wesley, Reading, Massachusetts, 1987.

**Kahle, B. and Medlar, A.,** An information system for corporate users: wide area information servers, *ConneXsions—The Interoperability Report*, 5(11), pp. 2–9. Interop, Inc., 1991.

**King, T. A., Walker, R. J. and Joy, S. P.,** DITDOS: a set of design specifications for distributed data inventories, in *Developments and Applications of Data Handling and Visualization Tools to Space and Atmospheric Sciences* (edited by E. P. Szeszczewicz and J. Bredekamp), 1994a (in press).

**King, T. A., Joy, S. P. and Walker, R. J.,** The design, development and operation of a distributed data inventory system, *Proceedings of the Seventh International Working Conference on Scientific and Statistical Database Management* (edited by J. C. French and H. Hinterberger), p. 287. IEEE Computer Society Press, Los Alamitos, California, 1994b.

**McMahon, S. K.,** Overview of the Planetary Data System. *Planet. Space Sci.* **44,** 3–11, 1996.

**Planetary Data System,** Planetary Data System standards reference, in *Planetary Data System Data Preparation Workbook*, JPL D-7669, part 2, version 3.1, 1993.

## Appendix

*Acronym list*

| | |
|---|---|
| CD-ROM | Compact Disk Read Only Memory |
| CD-WO | Compact Disk Write Once |
| DAWG | *Galileo* Data Archiving Working Group |
| DET | DITDOS Extraction Tool |
| DITDOS | Distributed Inventory Tracking and Data Ordering Specification |
| EDR | Experimenter Data Record |
| GSFC | Goddard Space Flight Center |
| HTML | Hyper Text Markup Language |
| IGPP | Institute of Geophysics and Planetary Physics at UCLA |
| IJW | International Jupiter Watch |
| JPL | Jet Propulsion Laboratory |
| MGN | *Magellan* |
| NEAR | Near Earth Asteroid Rendezvous |
| NSSDC | National Space Science Data Center |
| ODR | Original Data Record |
| PDS | Planetary Data System |
| PPI | Planetary Plasma Interactions |
| RDR | Reduced Data Record |
| SBN | Small Bodies Node |
| SEDR | Supplemental Experimenter Data Record |
| SEDR2ASC | Supplemental Experimenter Data Record to ASCII |
| SL9 | Comet Shoemaker–Levy 9 |
| SQL | Structured Query Language |
| UCLA | University of California at Los Angeles |
| URL | Universal Resource Locator |
| WAIS | Wide Area Information Server |
| WWW | World Wide Web. |

*Acronym list for spacecraft instruments*

*Galileo.*

| | |
|---|---|
| DDS | Dust Detector System |
| EPD | Energetic Particle Detector |
| EUV | Extreme Ultra Violet |
| HIC | Heavy Ion Counter |
| MAG | Magnetometer |
| PLS | Plasma Analyzer |
| PWS | Plasma Wave Subsystem |
| UVS | UltraViolet Spectrometer |
| SPICE | Spacecraft trajectory, Planetary ephemerides, Instrument parameters, C-matrix (pointing) and Event data (Acton, 1996). |

*Pioneer* 10 *and* 11.

| | |
|---|---|
| CPI (CPC) | Charged Particle Instrument (Counter) |
| CRT | Cosmic Ray Telescope |
| FGM | Fluxgate Magnetometer |
| GTT | Geiger Tube Telescope |
| HVM | Helium Vector Magnetometer |
| PA | Plasma Analyzer |
| TRD | Trapped Radiation Detector |
| UV Photo | Ultraviolet Photometer |
| Orbit-Attitude | Refers to spacecraft trajectory and orientation data |
| SPICE | Spacecraft trajectory, Planetary ephemerides, Instrument parameters, C-matrix (pointing) and Event data (Acton, 1996). |

*Pioneer Venus Orbiter.*

| | |
|---|---|
| OEFD | Orbiter Electric Field Detector |
| OETP | Orbiter Electron Temperature Probe |
| OGBD | Orbiter Gamma Burst Detector |
| OIMS | Orbiter Ion Mass Spectrometer |
| OMAG | Orbiter Magnetometer |
| ONMS | Orbiter Neutral Mass Spectrometer |
| OPA | Orbiter Plasma Analyzer |
| ORPA | Orbiter Retarding Potential Analyzer |
| ORSE | Orbiter Radio Science Experiment |
| SEDR | Supplemental Experimenter Data Records |
| Ephem | Spacecraft trajectory data prepared by the Magnetometer team |

*Voyager* 1 *and* 2.

| | |
|---|---|
| CRS | Cosmic Ray Subsystem |
| LECP | Low Energy Charged Particle |
| MAG | Magnetometer |
| PLS | Plasma Analyzer |
| PRA | Planetary Radio Astronomy |
| PWS | Plasma Wave Subsystem |
| position | Spacecraft trajectory data |
| SPICE | Spacecraft trajectory, Planetary ephemerides, Instrument parameters, C-matrix (pointing) and Event data (Acton, 1996). |

*Ulysses.*

| | |
|---|---|
| COSPIN | COsmic and Solar Particle INvestigations |
| DUST | Dust Detector System |
| EPAC | Energetic PArticle Composition instrument |
| FGM | FluxGate Magnetometer |
| HISCALE | Heliosphere Instrument for Spectra, Composition, Anisotropy at Low Energies |
| HVM | Helium Vector Magnetometer |
| SWICS | Solar Wind Ion Composition Spectrometer |
| SWOOP | Solar wind plasma |
| URAP | Unified Radio And Plasma wave experiment. |