

行星数据编目模型的设计与实现

李莎莎, 陈 萃, 陈宏盛, 钟志农

(国防科学技术大学电子科学与工程学院, 长沙 410073)

摘 要: 根据行星数据系统(PDS)标准规范、行星数据特点及 PDS 元数据模型, 设计一个行星数据编目模型。通过扩展传统数据编目模型支持数据展示和空间查询, 采用对象-关系数据库的方式存储编目信息以实现数据的管理、归档和分发。通过建立行星数据编目检索系统验证该模型的正确性和有效性, 并表明该模型能提供一致的目录访问接口, 使用户方便快速地检索到感兴趣的行星数据。

关键词: 行星数据系统; 海量数据管理; 元数据模型; 数据编目模型; 空间数据库

Design and Implementation of Planetary Data Cataloging Model

LI Sha-sha, CHEN Luo, CHEN Hong-sheng, ZHONG Zhi-nong

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

[Abstract] According to the standard specification, planetary data characteristics of Planetary Data System(PDS) and PDS metadata model, this paper designs a planetary data cataloging model. By extending traditional data cataloging model to support data show and space inquires, using the method of object-relational database to store cataloging information to realize data management, filing and distribution. The planetary data cataloging index system validates the correctness and effectiveness of the model, and shows that this model can provide a consistent directory access interface, makes the user quickly and easily to retrieve the interested planet data.

[Key words] Planetary Data System(PDS); massive data management; metadata model; data cataloging model; spatial database

DOI: 10.3969/j.issn.1000-3428.2011.17.076

1 概述

随着深空探测工程的起步和发展, 我国已积累并掌握了大量的行星探测数据, 这些数据是我国空间探测实验基础, 是十分宝贵的科学资料。如何实现海量行星数据管理, 并从中快速获取用户感兴趣的数据, 是实现行星科学数据有效利用和面向应用的关键。行星数据编目管理体系的建立, 为解决这些问题提供了有效工具。

行星数据编目是对行星数据属性、质量等进行描述的结构化信息, 是行星数据的目录。其目的在为分布在网络各节点的行星数据提供一致的目录访问接口, 使用户快速、方便地获取感兴趣的数据。行星数据系统(Planetary Data System, PDS)标准规范作为美国国家航天航空局(National Aeronautics and Space Administration, NASA)提出的行星科学数据存储组织和分发标准, 以其高效性和权威性已得到世界许多国家的支持与认可。本文基于 PDS 标准规范, 设计并实现行星数据编目模型, 建立支持该模型的原型系统, 实现行星数据的有效归档和管理。

2 行星数据系统及其标准规范研究

2.1 行星数据系统

为保证行星科学数据的长期可读性与可利用性, 更规范、有效地进行行星科学数据存储、交换与服务。NASA 科学委员会行星科学部于 1989 年提出建设行星数据归档和分发系统, 即 PDS。其主要任务是对行星任务中所获取的大量科学数据进行备份归档, 同时依据 PDS 标准规范, 为全球科研机构、科学家和公众分发数据^[1]。

PDS 及其标准规范在十几年的发展不断成熟与完善, 已成功应用于美国 NASA 所有行星科学数据的管理, 并在国

际行星探测数据标准方面发挥了重要作用。欧空局行星探测任务数据、印度 ISRO 的 Chandrayaan-1 探测数据、日本 JAXA 的金星任务探测数据、中国 CNSA 的嫦娥-1、嫦娥-2 月球探测任务数据、萤火火星探测任务数据, 均接受并采用 PDS 标准进行数据的归档和分发。至此, PDS 已成为实际存在的国际标准^[2]。

然而, PDS 是和万维网(World Wide Web, WWW)一同发展起来的, PDS 的设计之初是针对以磁带和 CD/DVD 为分发介质的数据集, 随着数据量和数据复杂性的增加, 数据的共享和分发面临新的问题。PDS 研究小组于 2008 年制定了 PDS 2010 Project, 提出未来 10 年的 PDS 发展建议, 建议采用现代数据库技术和 Web2.0 等技术保证和改进新的数据标准, 支持行星科学数据的快速高效地存储、查询、检索和分发^[2]。

2.2 行星数据系统标准规范

PDS 标准规范强调数据归档中必须包含必要的能解释多种数据存储格式和理解数据内容的元数据信息, 以保证行星科学数据在分发中易于解释和共享, 保证数据长期归档的可读性和可用性。

PDS 标准规范定义了数据集命名规范、数据组织方式、数据产品标签标记方法等, 以保证数据质量和数据完整性。数据集定义为一组具有相同来源、历史和应用的数据库产品

基金项目: 国家“863”计划基金资助重大项目(2008AA12A211)

作者简介: 李莎莎(1986—), 女, 硕士研究生, 主研方向: 地理信息系统, 数据库技术; 陈 萃, 副教授、博士; 陈宏盛, 副教授、硕士; 钟志农, 副教授、博士

收稿日期: 2011-03-10 **E-mail:** shasha.gis@gmail.com

的集合。完整的数据集包括原始数据以及辅助数据、软件和说明文件等^[3]。数据产品是数据集的组成部分，分为原始数据产品和辅助数据产品。数据产品由数据对象和 PDS 标签组成。其关系表示如图 1 所示。数据集中的数据产品按照一定的目录组织标准规范方式存储于数据卷中，进行数据的归档、存储与分发，依据 PDS 标准组织的行星探测数据，本文简称为 PDS 数据。

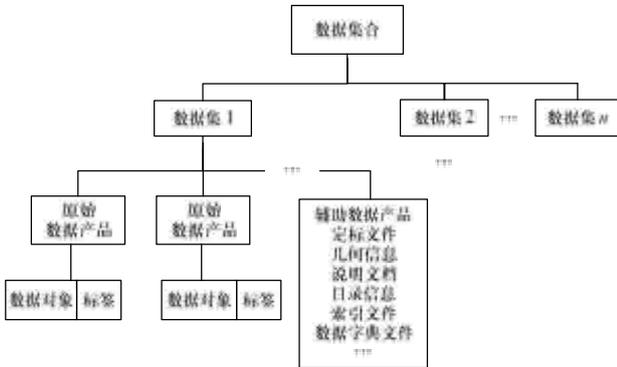


图 1 数据集与数据产品、数据对象的关系

2.3 行星数据系统的数据特点

随着深空探测任务的开展以及探测技术的进步，PDS 数据呈现如下特点：

(1)海量性：PDS 数据包含大量的原始影像、浏览图像、文本描述以及解析软件等，并随着深空探测的不断发展，数据量还将急剧增加。

(2)多样性：PDS 数据多样性来自于数据采集和数据生产 2 个方面。不同的目标、探测器、传感器以及数据生产涉及学科、机构、工具，使 PDS 数据呈现多源、多格式、多尺度的特点。

(3)归档性：行星探测任务代价高、难以重复，因此，PDS 数据作为宝贵的历史资料，要求保持数据原始性和完整性，不允许外部破坏。

(4)复杂性：PDS 数据集涉及不同的学科领域，所关注的目标特点和描述方法不同。

(5)空间性：PDS 数据包含属性信息、空间信息和时间信息，具有空间数据特征。

(6)分布式存储：随着数据量的增加和网络技术的发展，PDS 数据具有分布式存储的特点。

3 元数据模型及行星数据编目模型

行星数据编目面向的对象是广泛分布在网络各节点的行星数据。其目的是提供一致的目录访问接口，使用户方便、快速地检索到感兴趣的数据。数据编目的主要功能包括查找、识别、选择、获取和导航等。

优化的目录体系和编目方法可以最大化地缩短用户的查询路径，并具有良好的可扩展性和可维护性。因此，编目模型的设计原则是：

- (1)支持规范、一致的数据检索。
- (2)支持空间信息的浏览和复杂的空间查询。
- (3)具有良好的扩展性。

本文通过对行星数据的分析提取其元数据框架，建立编目模型以实现海量分布行星数据的统一编目与管理。

3.1 元数据模型

科学元数据是科学数据的描述数据，它以结构化的形式

描述了科学数据文件的内容、质量、表示方式、科学用途、属性参考、管理方式以及数据集的其他特征。科学元数据是行星数据编目的重要依据，是实现信息共享的基础。

PDS 数据模型是基于对象的数据模型，它采用对象描述语言(Object Description Language, ODL)标记其数据产品和目录信息^[3]。ODL 强调人和机器的可读性，其描述方法为：Keyword=Values。

PDS 数据的复杂性决定了数据描述难以达到完全一致。抽象其共性特征，描述 PDS 数据集的信息，包括航天器信息、目标信息、传感器信息、数据加工等级、数据集类型、数据集描述以及版本信息等。数据产品的描述信息包括行参数、列参数、影像时间、滤波器参数等。

3.2 行星数据编目模型

根据 PDS 元数据模型，数据集涉及的对象有航天器、目标和传感器，因此，行星数据编目模型的逻辑设计如图 2 所示。

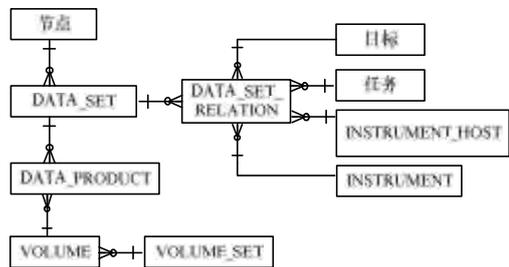


图 2 行星数据编目模型的结构设计

编目信息应该体现行星数据的空间性。由于行星数据多样性为标准化存储带来难度，传统遥感影像编目模型的设计通常忽略了这一特性。但是，浏览图像作为对原始影像的快视图，可以作为编目信息的一部分。相对于原始数据，浏览图像具有数据量小、格式简单的优点；其不足是分辨率低、无空间信息。如果浏览图像的空间特性可以通过数据处理获得，那么它可以作为空间数据提供浏览与空间查询。同时，数据卷的空间属性可以用面状要素描述，以表示其地理位置信息，建立与其他空间数据源的拓扑关系实现空间查询和空间分析。

因此，本文对图 2 模型进行扩展，使其包含空间对象信息。扩展后的数据集和数据卷逻辑结构设计如图 3 所示，其中，数据集属性 RASTER_DATASET_ID 指向数据集浏览图像空间对象，数据卷属性 VOLUME_OBJECT_ID 指向表示数据卷空间范围的面状要素。这样设计的目的是以空间对象扩展传统编目模型，使其支持数据展示和空间查询。

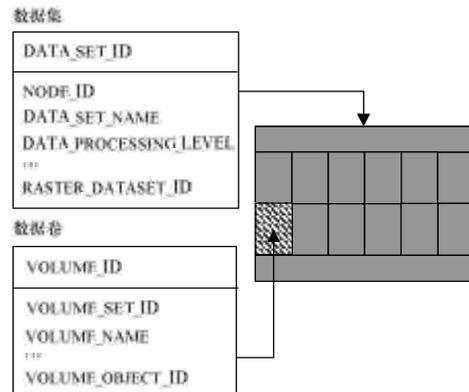


图 3 行星数据编目模型的扩展结构设计

3.3 编目信息的存储组织

行星数据特点决定了其编目信息的海量性、多样性、复杂性和空间性。编目信息的存储组织将直接影响未来数据的查询检索效率^[4]。

根据上文 PDS 数据编目信息的逻辑性描述, 编目信息的存储要考虑非空间信息与空间信息的存储, 并使其具有一致性。通常方案是将编目信息与浏览图像分别存放在关系数据库和文件系统中, 该方案在数据完整性与一致约束、并发控制、数据库恢复等方面存在不足, 系统维护难度大, 难以满足海量行星编目信息存储需要。基于对象-关系数据库的方式存储编目信息是一致性存储的解决方案, 将数据产品级的空间数据和属性信息作为一个整体以关系表的形式统一存储, 空间数据采用特殊数据类型存储。例如, 采用 Oracle 的 GeoRaster 实现影像数据的存储和管理。在数据查询时, 只需采用标准的结构化查询语言(Structured Query Language, SQL)命令就可访问和操纵数据库中的数据。采用这种数据存储组织方案的优点是: 以一致的方式维护了非空间信息与空间信息存储与管理, 既可以充分利用关系数据库的检索功能, 又可以实现编目信息的检索查询; 同时支持复杂的空间操作与空间查询。

4 行星数据编目检索系统设计与实现

4.1 系统设计

行星数据编目模型的主要作用是支持基于编目的行星数据检索, 支持行星数据中心对外提供一致的数据检索访问接口。行星数据中心是网络中各行星数据归档节点的管理节点, 负责所有节点的管理、编目信息的维护以及数据的分发。

因此, 在数据中心全局编目数据库基础上, 构建行星数据编目模型, 提供行星数据的统一查询视图, 可以屏蔽底层数据存储细节, 帮助用户快速有效地获得分布在各个节点的行星数据。

一个典型的数据检索用例是: 用户进入数据中心检索入口, 根据行星编目数据导航选择感兴趣的目标数据集; 通过地图窗口选择感兴趣的区域, 向数据中心提交查询请求; 数据中心组织查询, 获取包围该区域的数据卷的编目信息, 向存储节点发送数据请求, 存储节点返回目标数据卷, 提供给用户。基于行星数据编目模型设计如图 4 所示的行星数据编目检索系统。

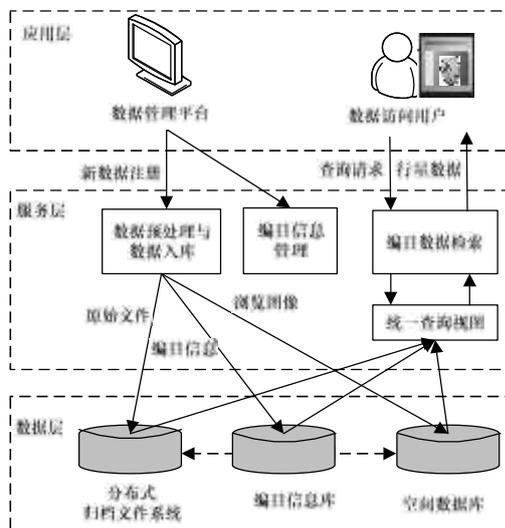


图 4 行星数据编目检索系统

4.2 系统实现

4.2.1 编目信息自动提取方法

编目信息的主要提取与构建方法如下:

(1)数据预处理

数据预处理的主要目的是提取 PDS 数据的浏览图像, 并为其配置空间信息文件, 使其具有空间数据特征。具体方法是: 扫描并获得数据卷中每一个数据产品标签的空间参考和坐标信息, 获得对应浏览图像的空间参考系信息并计算其空间位置, 生成对应的 Word File 文件。Word File 是 ArcGIS、Microstation 和 AutoCAD 等支持的影像数据坐标信息的文本文件, 该文件定义了影像像素坐标与实际地理坐标的仿射关系, 具体原理如下:

$$x' = Ax + By + C$$

$$y' = Dx + Ey + F$$

其中, x' 、 y' 是像素的地理或投影坐标; x 、 y 是像素坐标列号和行号; A 、 E 是水平和垂直方向上的像素分辨率; D 、 B 是平移和旋转系数; C 、 F 是栅格影像左上角像素中心的 X 、 Y 坐标值。

根据 PDS 标签中的参数, 计算 Word File 文件中的 6 个参数:

$$A = \text{LineSamples} \times \text{resolution} / \text{Width}$$

$$D = \text{Reference_longtitude}$$

$$B = \text{Map_projection_rotation}$$

$$E = -(\text{Lines} \times \text{resolution} / \text{Heigh})$$

$$C = -(\text{Sample_projection_offset} \times \text{resolution})$$

$$F = \text{Line_projection_offset} \times \text{resolution}$$

其中, $Width$ 、 $Heigh$ 是浏览图像的宽和高; $resolution$ 是原始影像的分辨率; $LineSamples$ 、 $Lines$ 是原始影像宽和高; $Sample_projection_offset$ 、 $Line_projection_offset$ 是 PDS 标签中标记原始影像左上像素的坐标值; $Reference_longtitude$ 、 $Map_projection_rotation$ 是 PDS 标签中原始影像的平移和旋转参数。

(2)编目信息自动提取与入库

数据的标准化是自动导入的基础, 通过构建 PDS 数据关键字的数据字典可以实现编目信息的自动提取, 提高数据管理效率。

编目信息提取流程是: 以卷为单位进行数据导入, 主要包括: 解析并获取 PDS 数据卷中的编目信息, 构建编目关系; 录入卷中数据产品编目信息及浏览图像记录; 在空间数据库中添加标记该卷范围的要素。

4.2.2 行星数据编目树的构建

行星数据编目的构建是基于全局数据库, 以目录树的方式提供用户检索, 可将分散的数据整合成一个单独的目录^[5]。数据集目录节点来自于与数据集相关的对象或数据集本身的属性信息, 并可根据用户关注的不同对目录树节点进行调整以及扩展。默认的目录结构是: 目标-任务-探测器-传感器-数据集类型-数据集。

4.2.3 行星数据的检索查询

基于本文的编目模型和编目检索系统可实现如下查询检索:

(1)编目检索: 以目录的方式, 引导用户检索到感兴趣的 PDS 数据集。

(2)关键字检索: 关键字检索是系统最基本的功能, 用户

(下转第 232 页)