

# The Study on the Multiple Data Centers Data Integrated Storage Model Based on a Feature Class

Guanzhu Wang<sup>1</sup>, Haochuan Li<sup>2</sup>, Fei Huang<sup>3</sup>

<sup>1</sup>China Centre for Resources Satellite Data and Application, Beijing

<sup>2</sup>State Information Center, Beijing

<sup>3</sup>Beijing GEOWAY Software Co., Ltd., Beijing

Email: 13810011471@139.com

Received: Feb. 4<sup>th</sup>, 2016; accepted: Feb. 23<sup>rd</sup>, 2016; published: Feb. 26<sup>th</sup>, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

According to the multiple integrated data storage model, this paper proposed the integration of a multiple data center data storage model based on a feature class, combined with the actual needs of multiple data centers and distributed heterogeneous environment. It also built the model based on the project of the National Natural Resources and Geospatial Basic Information Database. Modeling results showed that the model could meet the demands of integrated storage and management of massive multi-source data under the multiple data centers and distributed heterogeneous storage environments.

## Keywords

Multiple Data Centers, Integrated Data Storage, Multi-Source Heterogeneous Data, Feature Class

---

# 基于要素类的多数据中心数据一体化存储模型 研究

王冠珠<sup>1</sup>, 李浩川<sup>2</sup>, 黄非<sup>3</sup>

<sup>1</sup>中国资源卫星应用中心, 北京

<sup>2</sup>国家信息中心, 北京

<sup>3</sup>北京吉威数源信息技术有限公司, 北京

Email: 13810011471@139.com

收稿日期: 2016年2月4日; 录用日期: 2016年2月23日; 发布日期: 2016年2月26日

---

## 摘要

本文根据多个一体化数据存储模型, 结合多数据中心、分布式异构环境等实际需求, 提出了基于要素类的多数据中心数据一体化存储模型, 并以国家自然资源和地理信息基础信息库项目为基础, 进行了建模。建模结果显示: 该模型可以满足多数据中心、分布式异构环境下的海量多源数据一体化存储与管理。

## 关键词

多数据中心, 一体化存储, 多源异构数据, 要素类

---

## 1. 引言

国家自然资源和地理信息基础信息库是国家发改委联合 11 个相关部门开展的建设项目, 旨在整合资源, 切实提高资源的开发利用和共享水平, 为社会提供标准化、规模化和可持续开发利用的自然资源和地理空间基础信息。

由于 11 个部门的数据本身自成体系, 汇集在一起形成服务时, 在数据来源、种类、时间序列上均相比单个部门复杂。如何根据多数据中心、分布式异构环境下的应用需求, 确立数据(矢量、影像、格网、图表、图片、文档等)的存储模型, 形成各类数据的存储模型及实现技术方案是本文研究的重点。

## 2. 一体化数据模型研究现状

关于一体化数据模型的研究, 有如下几种情况:

1) 提出一种新的空间数据组织模型——矢量栅格混合的数据模型[1]。从数据结构模型层次考虑[2], 能兼顾矢量、栅格数据模型的优点, 矢量、栅格数据可转化到新模型中存储, 从物理层次实现一体化。

2) 基于某一种软件技术对矢量、栅格、表格等数据的统一管理, 物理上各种类型的数据选择软件提供的相应的物理结构存储, 例如应用 ArcSDE 这一成熟的海量空间数据库引擎, 结合数据库管理系统能实现一体化的管理; 或者基于 OracleSpatial 实现了在 Oracle 数据库中同时存储和管理栅格文件与矢量文件并无须借助其他工具[3]。

3) 从数据的组织管理考虑的, 以某些软件技术为支撑, 对分布式的多源海量的矢量、栅格、表格等数据的统一管理, 不同的数据采用不同数据存储方式、通过数据索引、元数据等的建立, 在组织管理层面达到一体化的效果。

## 3. 基于要素类的一体化数据模型设计

统一时空框架下的多维海量异构多类型数据一体化存储组织模型, 从本质上来说是数据模型。内容上涵盖包括空间数据和与空间相关的经济统计等非空间数据, 空间数据又具有多种模型实现方式。然而, 不论是何种数据, 研究的最终目标是实现一体化的存储组织管理, 本文提出基于要素类的一体化数据模型设计方法, 以解决多中心多源异构数据的一体化存储。

多类型数据包括矢量、栅格、格网、表格、多媒体数据等，对于多媒体数据通过数据库字段存储，在数据模型中可以归到表格数据处理。

面向要素类的多类型数据一体化存储组织，即以要素类作为对象进行数据建模，矢量、栅格、格网、表格数据在组织管理上实现一体化，在存储上，需要实现数据关系的存储和管理。

### 3.1. 多数据中心存储结构设计

#### 3.1.1. 数据中心组织模型

对于多数据中心的组织管理，采用分布式、多层次的方法。可以将数据中心分为数据主中心和数据分中心。数据分中心是数据提供单位，独立的存储自有的专题数据。数据分中心之间相互独立，数据分中心与数据主中心存在连接。数据分中心通过主中心实现其与其他分中心数据的共享交换。数据中心的层次关系如图 1。

数据主中心和数据分中心具有一定的相对性，一个数据中心是某个数据中心的分中心，而其下也可以有多个数据中心，相对于这些数据中心，该数据中心又是数据主中心。

数据分中心可以独立对外提供信息服务，服务内容为本中心自有信息；数据主中心对外提供综合的信息服务，服务内容为其所有分中心的共享信息。

#### 3.1.2. 数据中心数据库组织结构

数据中心包含多个数据库，数据库可以进一步划分为多个子库，数据库可划分为子库，子库下包含多个数据集，数据集由要素类或对象类组成。同时数据库应具有元数据、数据编目、数据字典等信息，支撑数据库的运行管理。

元数据描述对象的基本信息，来源、组成等内容。数据编目是对要素类对象的描述，既有关于数据精度、空间范围、专题信息的描述，也有关于要素类之间的关系、要素类支持的操作的说明。数据字典描述的是数据本身了，主要是用来解释数据表、数据字段等数据结构意义，数据字段的取值范围，数据值代表意义等等。数据库组织结构如图 2。

#### 3.1.3. 基于要素类的数据集设计

根据需求，项目过程中涉及到的数据包括：矢量数据、影像数据、格网数据、表格数据、多媒体数据、三维可视化数据、数字字典、数据编目数据、元数据、数据关系等，下面以影像数据的存储为例说

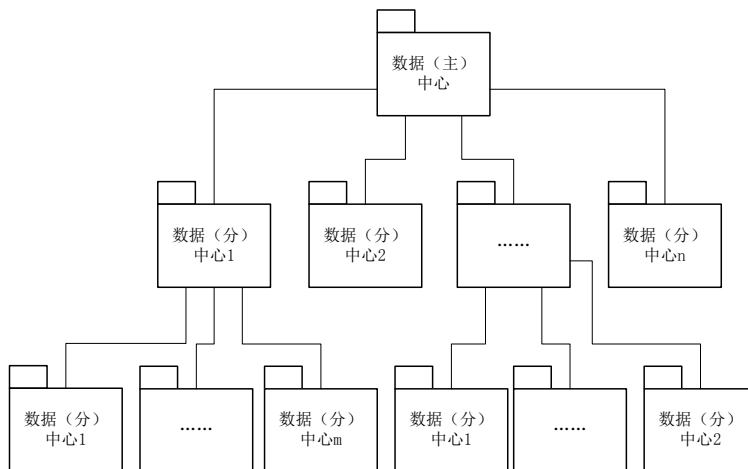


Figure 1. Data center hierarchical relation

图 1. 数据中心层次关系

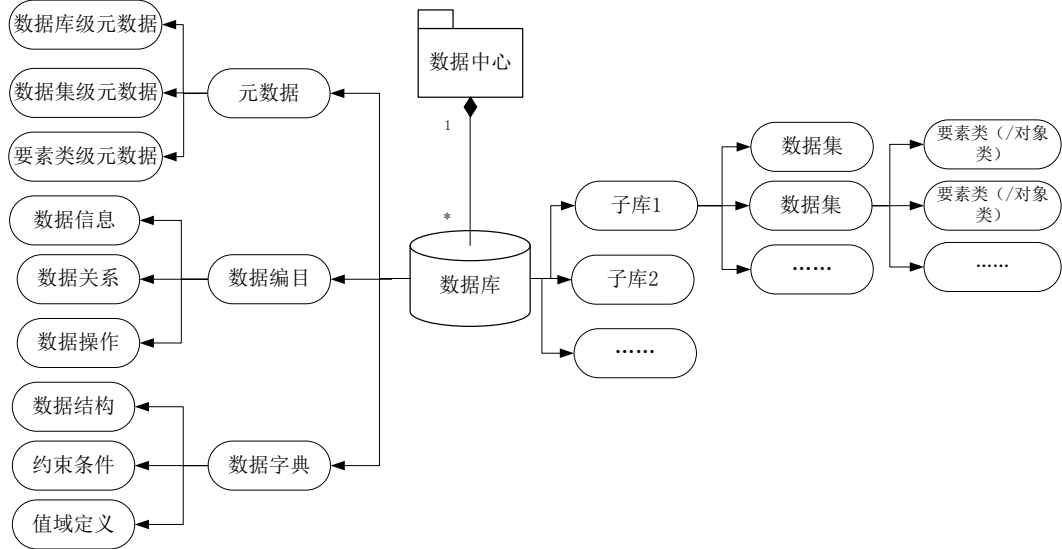


Figure 2. Data organization structure of database  
图 2. 数据库数据组织结构

明影像数据集的设计。

影像数据是栅格数据中的一种，影像数据表达地理要素时，一个象元所指向的地理对象是一定的，不存在二义性的。

影像数据按照波段、空间范围来分层、分块，它作为空间数据集下的栅格要素类建模；栅格作为一个要素类，每一个分块代表一个要素。一个栅格要素类具有确定的时间参考和空间参考，栅格要素类分块或者分层存储栅格数据实体，实体内包含多个象元，每个象元具有固定的空间位置，和描述该象元所代表的基本信息。作为栅格数据种类中一种的影像数据，具有传感器信息、波段范围信息，或其他扩充信息。栅格数据模型如图 3。

栅格数据建模支持面向多种空间数据源的各种栅格目录模型的定义，采用通用的空间数据库存储技术，提供对栅格数据存储方式、压缩方法、拼接方式等参数的设置，栅格数据建模支持目前常见的空间数据模型，如 OracleSpatial、ArcSDE、PGDB 多种。

在具体存储实现时，影像实体根据标准分景、分幅等方式组织，以文件或数据库形式存储；其他信息通过附加表进行存储；数据间的关系通过关系类实现。

其它类型的数据集设计的原理与影像数据集的设计类似。

### 3.2. 多数据中心的存储模型设计

根据数据的特点，以关系数据库、空间数据库、文件的方式进行存储[4]。矢量数据在空间数据库中存储。栅格数据中数据量较小的、使用频率高的，以空间数据库形式存储；数据量大、不经常使用的，以文件形式存储。格网数据分为矢量格网和栅格格网，矢量格网存储与矢量数据相同，栅格格网存储与栅格数据相同。表格数据为非空间数据，在关系数据库中存储。多媒体数据以关系数据库存储时，体现为表格；也可以以文件形式存储，数据库中仅记录路径及描述信息。文件数据集包含了影像数据、多媒体数据等，通过文件夹组织管理。

## 4. 面向自然资源的领域建模

以全国自然资源和地理空间基础信息库为例，数据中心、数据库、数据集及各类数据依据基于要素

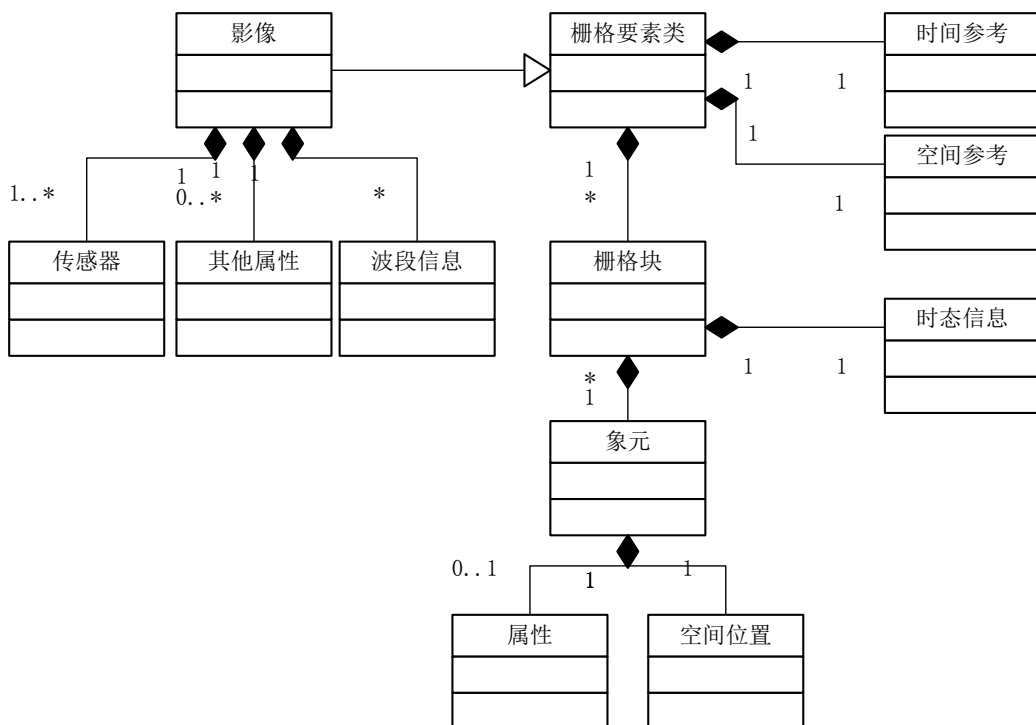


Figure 3. Raster data model  
图 3. 栅格数据模型

类的一体化数据模型组织[5]。

#### 4.1. 数据中心逻辑组织

国家自然资源和地理空间库数据中心包含一个数据主中心和多个数据分中心。数据分中心与数据主中心之间存在联系，进行共享与交换。在国家部委层面，包含 11 个数据中心，各个数据中心下，又有数据分中心。省级试点层面，包含三个数据分中心。逻辑组织如图 4 所示。

#### 4.2. 数据库逻辑组织

其中，综合信息库逻辑划分为 9 个一级库，每个一级库下包含若干数据集。数据主中心综合信息库逻辑组织图如图 5 所示。

对于栅格数据集及栅格要素类进行建模，以资源卫星遥感数据标准产品数据集为例。资源卫星遥感数据标准产品库包括：CCD、IRMSS 和 WFI 三个相机遥感数据的标准产品。每个相机遥感数据标准产品包括：0 级数据、1 级标准产品、2 级标准产品、3 级标准产品等 4 个级别的标准产品。逻辑组织模型如图 6 所示。

该模型由主中心覆盖设计模型和分中心具体应用模型两部分组成，其中主中心覆盖设计模型包括栅格类(Raster\_Class)和要素类(Geo\_Class)，分中心具体应用模型包括卫星遥感数据标准景 0 级产品信息、标准产品信息、信息库、产品处理步骤、资料引用、单位信息、控制点信息、卫星信息、传感器信息、波段信息、覆盖范围信息等十一个要素类组成。其中卫星遥感数据标准产品是栅格类(Raster\_Class)的派生类，会全部继承栅格类(Raster\_Class)的属性，但其自身属性和相关联要素类作为子库设计的重点。

#### 4.3. 数据存储

数据存储同时涉及空间数据库、关系数据库和文件数据库。矢量、格网和部分栅格数据存储于空间

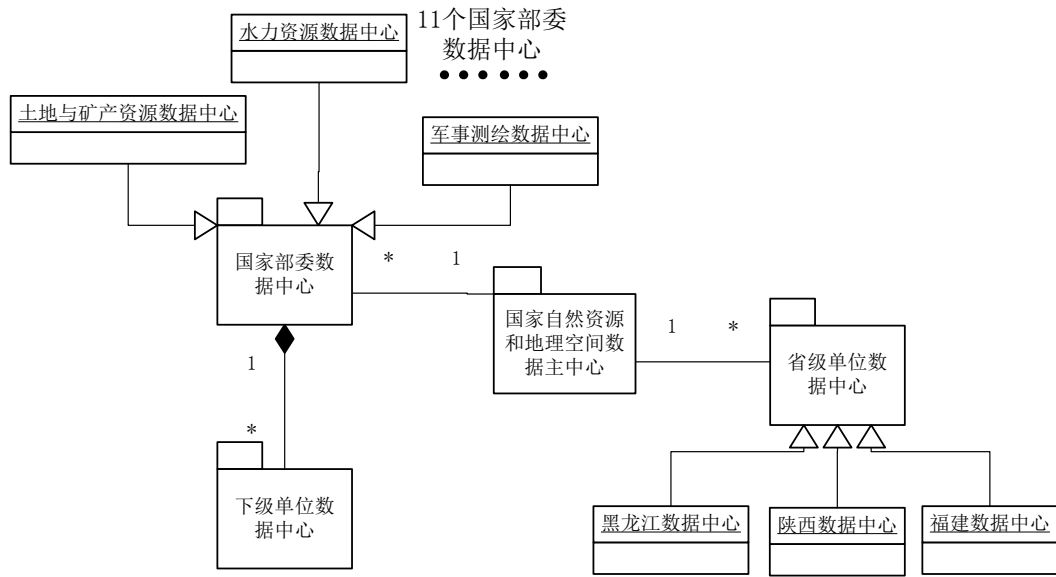


Figure 4. The center logical model chart of the national natural resources and geospatial basic information database

图 4. 国家自然资源和地理空间数据库中心逻辑模型图

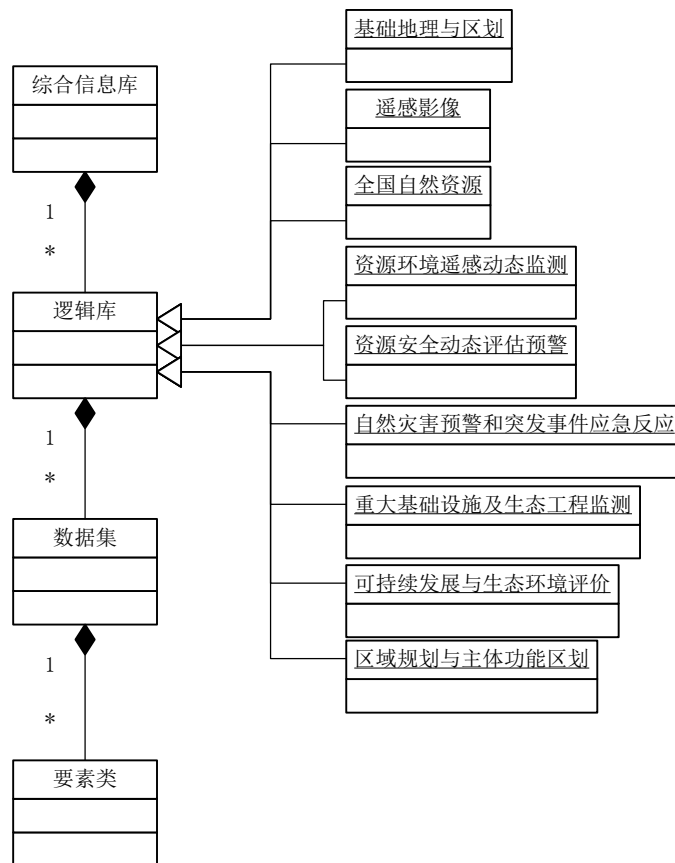


Figure 5. Logical organization chart of integrated information database in data center

图 5. 数据主中心综合信息库逻辑组织图

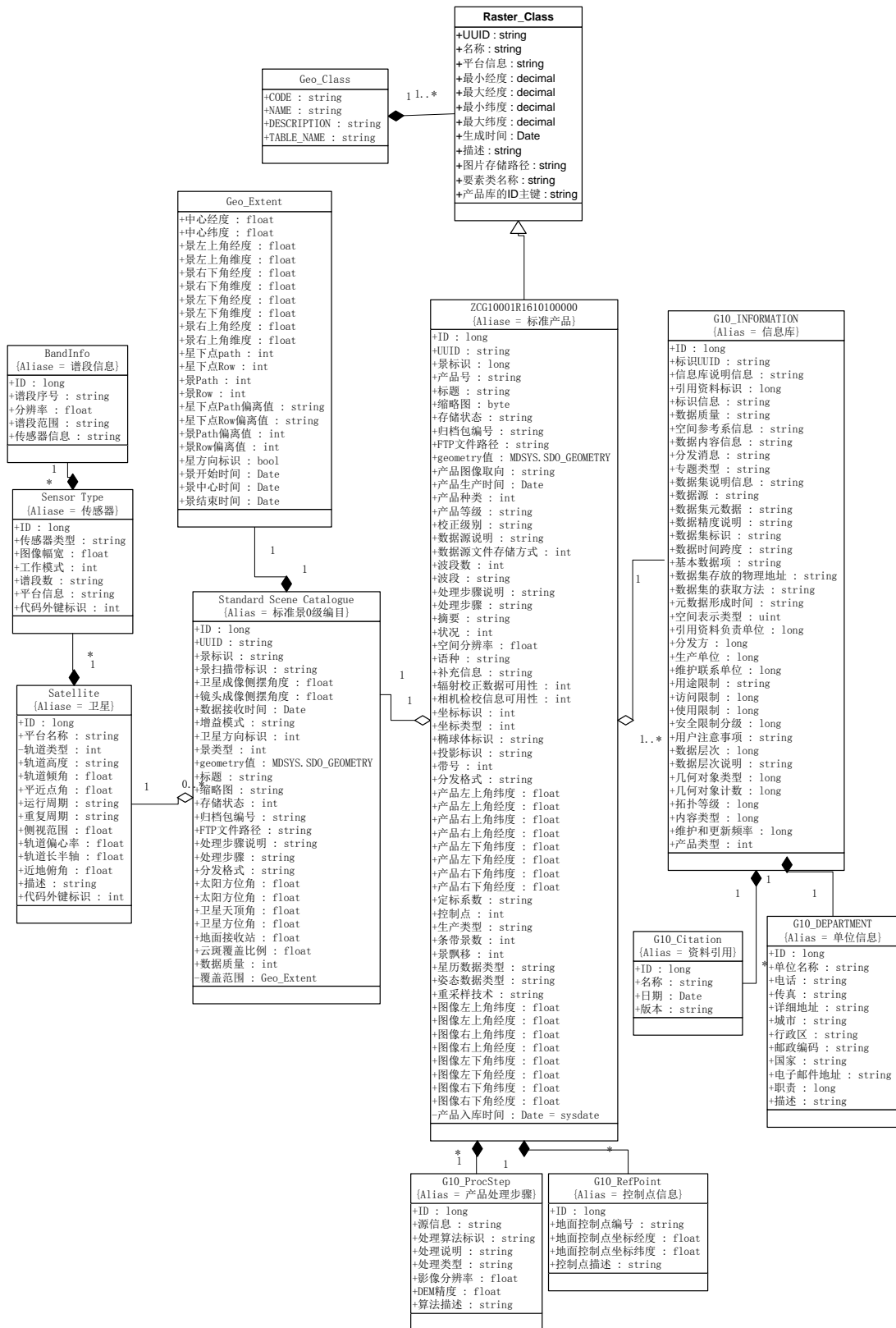


Figure 6. Model structure of standard product of CRESDA remote sensing data  
图 6. 资源卫星遥感数据标准产品模型结构

数据库；表格数据、数据字典、编目、元数据等存储于关系数据库；分格网数据、多媒体数据等存储于文件数据库。

以选用 Oracle 数据库和 ArcSDE 中间引擎为例，空间数据库采用 ArcSDE for Oracle 的 Geodatabase，关系数据库为 Oracle 数据库。矢量数据集采用 FeatureDataset (要素数据集)，矢量要素类用 FeatureClass (要素类)存储；栅格数据采用 RasterCatalog (栅格目录)或 RasterDataset (栅格数据集)存储；矢量格网采用 FeatureClass 存储；与这些空间数据紧密关系的属性数据，也可在空间数据库中用 Table 存储[6]。其中，要素数据集用来存放多个要素类，栅格数据集用来存储多个栅格影像。数据字典、数据编目、数据关系，元数据在 Oracle 表中实现。影像、瓦片、多媒体等数据，以文件数据库方式存储。

## 5. 结论与展望

本文结合国家自然资源和地理空间数据的特点，从面向数据中心、海量异构数据对象出发，研究各种类型数据的存储组织方式，统一概念，提出了满足各类数据存储模型，并开展了建模研究。

在研究基础上，初步实现了一个主中心与十一个分中心的数据统一存储与管理，并实现了矢量、影像、格网、图表、图片、文档等各类型数据的存储管理。主中心与各分中心正在良好的运行当中。应用结果显示，在物理分散、逻辑统一的整体框架下，基于要素类的多数据中心数据一体化存储模型能很好地满足的项目的各项应用要求。

在新的需求导向下，如何扩展数据的一体化存储模型，以适应云计算技术为基础，大数据分析为手段的新技术架构，进一步提升数据一体化存储的广度与深度，是进一步研究的重点。

## 基金项目

本研究由 863 课题资助，课题号：2012AA12A408。

## 参考文献 (References)

- [1] 王伟, 杜道生, 熊汉江. 分布式栅格数据库的一体化管理[J]. 测绘通报, 2007(3): 57-59.
- [2] 李云, 戴长华. GIS 中空间数据结构和空间数据模型一体化研究[J]. 计算机与现代化, 2003(9): 7-10.
- [3] 孙荣辉. 基于 OracleSpatial 的空间数据一体化存储研究[D]: [硕士学位论文]. 北京: 首都师范大学, 2006.
- [4] 李景朝, 王永志, 林燕, 何文娜. 全国矿业权实地核查成果数据一体化存储与管理[J]. 地理信息世界, 2011(5): 23-25.
- [5] 黄士力, 金惠英, 周凡, 冯径, 潘崇伦. 面向数据驱动应用的水务信息一体化数据组织与管理[J]. 水利水文自动化, 2008(3): 1-5.
- [6] 黄成, 丁琳, 张保钢, 侯敏. OracleGeoraster 的数据存储研究及其与 ArcSDE 性能的对比[J]. 北京测绘, 2007(1): 5-11.