

基于元数据规范的高分遥感时空大数据智能服务平台研制

文/李翔翔 李集林

摘要

为了低耗高效的充分开发高分专项的数据应用价值,为领域用户提供一套规范化、综合性、一站式的数服务系统,研究高分专项元数据规范的设计框架、提出了一种基于元数据规范的应用主题数据字典模型,基于该模型设计了一套高分遥感时空大数据智能服务平台系统,通过实际项目的应用,初步验证了高分专项元数据及字典规范合理性和本平台系统设计的有效。

【关键词】高分专项 元数据规范 主题数据字典 时空大数据 智能服务

1 概述

高分辨率对地观测系统重大专项是《国家中长期科学与技术发展规划纲要(2006-2020)》所部署的16个重大专项之一。该专项将统筹建设基于卫星、平流层飞艇和飞机的高分辨率对地观测系统,完善地面资源,与其他观测手段结合,形成全天候、全天时、全球覆盖的对地观测能力。

在此趋势下,如何低耗高效的充分开发高分专项的应用价值成为行业内共同关注的热点问题。低耗的数据传输、运转、应用机制需要建立在强化统一的卫星数据接收、生产、存储、交换、服务标准规范基础之上,从而推动遥感卫星数据的开发和共享,形成长期、连续、高效的数据应用服务体系。高分元数据规范体系的建立可以用于许多方面,包括数据文档建立、数据发布、数据浏览、数据转换等。如图1所示,其主要作用可以总结为以下几点:

(1) 帮助数据生产单位有效地管理和维护高分遥感影像数据,建立详细的数据文档,即使工作人员变动,也可以根据元数据详细地了解高分影像信息;

(2) 为各级、各类高分数据中心、应用中心体系(如地理信息系统)提供整合的工具与纽带,元数据规范既是数据集成的核心技术,也是系统集成的有利工具;

(3) 提供数据生产单位各方面的信息,便于用户查询检索地理空间数据,其中包括信息内容、存储位置、数据分类、质量、交换网络等;

(4) 提供便于用户转换和处理有用数据的帮助说明信息;

(5) 提供网络查询、检索数据的方法及途径以及与数据交换和传输相关的辅助信息标准。

2 需求分析

高分信息元数据规范需要按照国际、国内标准规范编制的框架和要求,科学规范的制定,并考虑与国际、国内已有标准的继承性,保证数据共享和交换的快速和方便,具体需求内容包括以下几个方面:

2.1 国内外标准兼容需求

在遵循国际主流遥感数据格式标准的基础上,针对我国高分遥感器的特点,制定适合于我国的遥感数据格式标准,既有利于统一我国各部门、行业遥感数据格式,也有利于标准将来与国际接轨,甚至上升为国际标准。

2.2 跨平台、异构环境兼容需求

我国正建立自己的遥感数据分发共享的高分数据服务体系,以实现不同部门、不同行业对不同遥感数据源的交换与共享。高分遥感元数据格式规范需要支持跨平台、异构环境的应用需求,包括不同软件平台(操作系统和应用软件)和不同硬件平台的兼容能力。以操作系统为例,由于不同的操作系统(如Windows、Linux和Unix等)在文件编码和数据存储方面存在一定的差异。因此高分遥感元数据格式规范应能够支持用户在不同的操作系统上进行遥感数据的互操作。随着我国对标准化工作的日益重视,特别是遥感数据的交换、共享的需求不断增加,越来越要求有一个能够统一各行业和部门的遥感数据格式标准,从而避免资源的不必要的浪费,节省人力、物力和财力。

2.3 元数据类别定义需求

为了最大化的提升高分数据的应用价值,拓宽并推广行业、区域应用成果为公众服务,解决产业链长、应用范围广、“数据-信息-知识”转换带来的信息集成与共享的问题,并最终推动高分数据进入行业、区域广泛应用,高分元数据规范应按照数据的应用领域特性与产品数据固有特性分层、分类的定义基础属性类、扩展属性类、应用特性类,从而满足数据的共享服务、扩展性、应用融合管理的需求。

2.4 内容扩展需求

高分数据的特性包括应用领域特性与产品数据固有特性。应用领域包括气象、林业、减灾、国土、地震、测绘、水利、农业、卫生、交通运输等,各领域用户管理的高分深加工产品具有特定的应用价值,元数据规范需要对各类应用定义具有唯一识别性的属性扩展集合,方便用户共享和获取数据。高分产品数据的固有特性包括云量、观测模式、幅宽、数据结构等,固有属性数量多,属性的定义域和值域变化较大,因此在元数据规范中应考虑属性内容扩展能力。

2.5 时态版本关联需求

高分数据最高时间分辨率将达到分钟级,

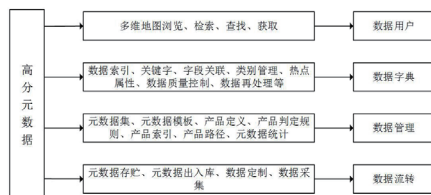


图1: 基于元数据高分遥感元信息可视化应用示例

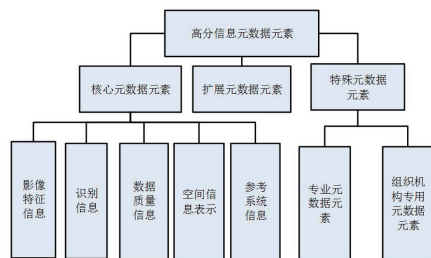


图2: 高分信息元数据元素组成图

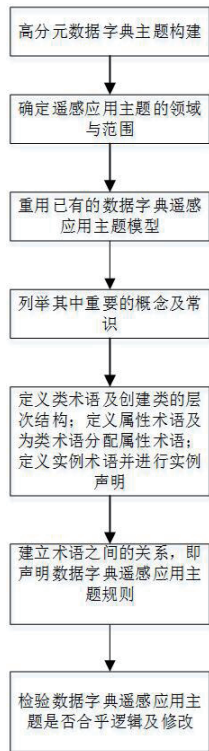


图3: 高分元数据字典主题构建流程

因此对海量多时相数据、多版本数据的管理需求将成为高分数据中心首要解决的问题。历史数据的迁移、备份,如何通过自动化的手段帮助用户准确获取目标数据最终都将归结到时态版本关联这一核心问题上。

2.6 数据与应用的关联需求

目前主流遥感数据管理系统都欠缺对应用的关注,导致影像数据检索条件复杂,无法

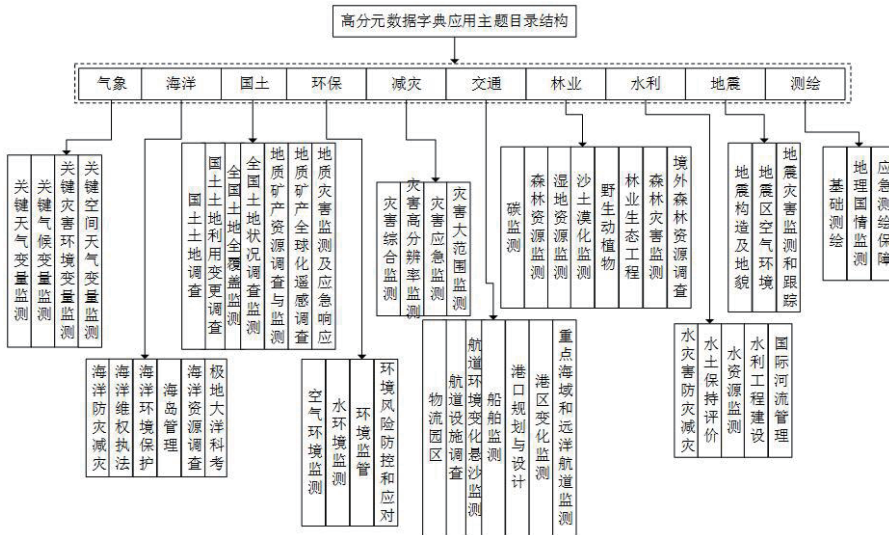


图4：高分元数据字典应用主题目录结构



图6：高分遥感时空大数据智能服务平台系统功能

做到精确检索。因此每个数据订购需求必将大于用户的实际需求，进一步导致数据管理系统共享数据的压力过大。

2.7 共享服务需求

用户对影像数据的需求是多种多样的，一个影像数据提供者不可能同时满足所有用户的需求。但是，用户总是希望一个共享体系能够为他们提供所有的元数据信息，以便及时获得所需要的影像数据。所以，一个部门单独提供的数据共享服务对于用户的需求是微不足道的，甚至没有现实意义。因此，数据中心、应用技术中心和12个行业应用系统应该在数据共享层面上通过技术手段实现合作，以便建立广泛适应用户需求的影像数据共享体系。在高分地面应用体系中，各参与部门不仅是影像数据的提供者，同时也是共享体系的用户。他们在为用户提供影像共享服务的同时，也可以及时从其他科研机构获取自己所需的影像数据或其他有价值的信息。这对于影像数据的重复生产，提高效益极其有力。

3 元数据及字典规范设计规范设计

通过前述需求分析，笔者将高分信息元数据分解为图2的三大部分：核心元数据元素、扩展元数据元素和特殊元数据元素。核心元素定义高分影像的元数据最小集，包括影像特征信息、识别信息、数据质量信息、空间信息、参考系信息。特殊类元素包括专业元素和组织机构专用的元素。扩展类元素则定义描述高分影像包括了高分信息的进一步详细特征和参考数据等信息。

传统的信息系统模型通常是建立实体关系模型和数据库信息模式，将信息数据按照建立好的模式输入至一个数据库中，这种信息系统的程序、数据以及编写方式互相依赖，一旦系统投入使用，资源信息组织方式不能改变，该系统的不可扩展性较大的限制着资源更新。基于这种情况，建立可扩展的数据字典体系对于高分遥感时空大数据智能服务平台有着极为重要的意义，一方面可扩展的数据字典体系可实现数据一致性，另一方面能有效提高数据结构和系统功能的独立性，最终增强系统通用性。因此，高分信息元数据规范设计了基于应用主题的数据字典规范，这部分结合了高分遥感信息本质特征和行业需求，以高分遥感卫星应用领域主题为切入点进行元数据字典规范设计。如图3所示，高分元数据字典主题构建可简单描述为六个流程。

随着高分分辨率卫星应用的发展，高分卫星能够全天候和全天时实现全球海洋和陆地信息的监视监测，可服务海洋、减灾、水利及气象等多个行业及业务部门。高分元数据内容涉及大气、海洋、国土、气象气候、地震、林业、水利、灾害、生态脆弱带、全球变化等领域。高分遥感时空大数据智能服务平台将结合高分遥感信息本质特征和行业需求，以高分遥感卫星应用领域主题为切入点进行元数据字典目录结构设计，数据字典目录结构图见图4。

4 规范应用设计

研究高分信息数据字典目录和应用流程的目标是建立一套更先进、更面向用户的元数据规范服务模式，并从规范的角度提炼出一套

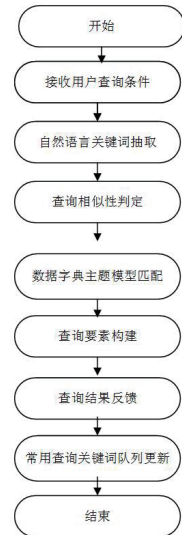


图5：基于应用主题的高分遥感时空大数据智能服务平台数据检索工作流程

基于语义级别上的应用主题查询检索策略目录。通过建立关于遥感数据信息源的应用主题将所有已有的数据源高度整合，挖掘信息之间的内在联系。图5描述了高分遥感时空大数据智能服务平台基于数据字典主题检索的工作流程。

基于应用主题的高分遥感时空大数据智能服务平台数据检索工作流程设计目标是：即可以通过正向的在用户查询的过程中，基于数据字典主题分类进行高分信息的快速检索获取，也可以逆向的通过用户的检索，主动对热点数据对应的查询关键词执行加权处理。

5 平台设计

基于元数据规范的高分遥感时空大数据智能服务平台面向元数据规范化应用，实现空间数据、遥感数据、情报数据等多类数据及产品的一体化存储管理。提供统一的面向应用主题的大数据关联存取组织、检索、展示和辅助研判功能，实现遥感时空数据的“即插即用”注册、数据管理类别的动态扩展能力与安全控制等功能。将解决多星、多载荷、高分辨率海量数据接入、处理压力与系统高效定制化服务的冲突，充分发挥遥感卫星的大数据资源优势，作为综合性的遥感时空大数据中心是构建天地一体化卫星应用工程的数据服务基础。其亮点主要有以下四个方面：

- (1) 面向签约用户的定制化离线数据、应用服务；
- (2) 基于主题检索的影像数据智能搜索引擎服务；
- (3) 精准时空数据检索及热点影像服务；
- (4) 基于影像数据特征和用户热点需求的PB级文件生命周期管理。

系统功能组成见图6。

数据生命周期管理功能包括：数据存储目录维护、数据迁移恢复、冗余数据清理三项。数据存储目录维护功能实现文件存储类数据目录的深度和子文件数量平衡管理，数据迁移恢

基于 Web 的专业气象服务平台的设计与实现

文/宋永鹏¹ 李兰兰² 黎亮¹

摘要

使用 PHP 和 ASP 开发语言, Mysql 和 Access 数据库, 在 PHPCMS 框架中开发基于 web 的淄博市专业气象服务平台, 平台域名: <http://www.zbszyqxt.com>。介绍了平台的总体结构, 并对系统所包含的专业气象服务数据库系统、专业气象产品制作系统、专业气象产品显示系统、平台及用户管理系统和淄博专业气象台门户系统进行了概述。

【关键词】淄博专业气象 服务产品 Web 技术 PHPCMS

淄博市专业服务经过 10 多年的发展, 已经拥有了大量的专业用户, 并建立了良好的合作关系, 为了适应经济社会发展需求, 结合淄博市的实际气象服务情况, 设计并实现了淄博专业气象服务平台。平台提供的气象服务产品

包括: 气象预警信息、48 小时短期预报、三小时预报、各区县天气预报、五天滚动预报、今日生活指数预报、天黑天亮日出日落、十七城市天气预报、天气实况、电业局专用天气实况、日降水量统计、月降水量统计、中期预报、长期预报、区县三天预报、逐时预报、卫星云图和雷达图等十数种。平台的建设, 较好地规范了工作流程, 提高了工作效率, 丰富了服务产品, 有效地提高了专业气象服务水平。

1 平台设计

平台主要包括专业气象产品制作系统、平台及用户管理系统、专业气象服务数据库系统、专业气象产品显示系统和淄博市专业气象门户系统 5 部分。平台针对专业气象服务特点, 依托多种气象服务产品, 以专业气象产品数据库为支撑, 在专业预报的基础上, 通过专业气象产品制作系统形成预报服务产品并保存在专业气象产品数据库, 通过专业气象产品显示系统为行业用户提供服务, 建立自动化程度较高、以网络传输为主要途径的专业气象服务工作平台。平台总体结构如图 1。

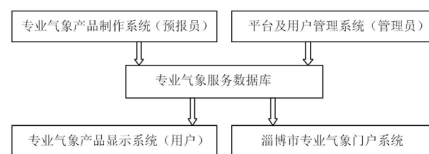


图 1: 平台总体结构

2 专业气象产品制作系统

专业气象产品制作系统使用 ASP 语言 (Active Server Pages) + Access 数据库开发; 这套开发环境结合了 ASP 的语言成熟和 Access 数据库的体积小反应快的优势, 并且有 Windows 系统对该开发环境的支持。

预报员将各种专业气象服务产品所需要的气象数据和资料文件输入图 2 的专业气象产品制作系统, 系统在后台根据用户的需求和产品的规则生成服务产品并上传保存到专业气象服务数据库。

3 平台及用户管理系统

平台及用户管理系统使用 PHP 语言 (Hypertext Preprocessor) + Mysql 数据库, 以

<< 上接 197 页

复功能实现冷数据降级、升级存储和数据动态降级规则维护, 动态降级规则因子包括数据质量、云盖指数、用户焦点度、时延等属性。冗余数据清理功能实现对错误数据、低质量数据、重复数据、常规历史日志文件的逻辑删除和物理删除功能。

系统资源安全监控对象包括: 数据库、业务服务器、存储、消息通道和平台内核。能够对这些资源对象的在线状态、运行服务、关键目录进行实时监控, 并在故障、恶意攻击发生时及时报警, 为用户提供详细故障日志, 以便问题定位。

数据共享服务接口监管功能包括数据规范维护、接入及共享推送、任务监控三项功能。数据规范维护为针对遥感影像、矢量、专题地图、文本对象特征进行识别策略建模和数据验证。接入及共享推送功能实现遥感影像、矢量、专题地图、文本数据的自动编目接入、数据批量共享推送、下载, 任务监控则具备对接入、共享推送任务实时监控及图形界面展示功能。

主题建模与应用领域知识库维护包括规则验证数据字典管理、行政区域及空间逻辑字典管理、主题模型创建及关联管理。规则验证数据字典管理包括各类数据属性有效阈值定义及检验规则管理。行政区域及空间逻辑字典管理实现全球各级有效行政区域、空间关系目录对象化及维护。主题模型创建及关联管理包括

主题词库管理、领域主题建模、领域知识树型目录管理。

精准时空数据库检索功能实现世界范围的国家和省/州/郡两级行政区查询、国内行政区实现省区/市/县三级查询, 支持多级行政区组合查询; 具备经纬度点坐标范围、矩形区域、多边形区域和矢量文件为空间搜索模板的空间查询; 所有空间查询支持“覆盖”, “包含于”和“相交”空间位置二次筛选。针对所有类型空间搜索模板提供查询历史记忆和查询历史清理功能。针对文件型的空间搜索模板提供上传功能。

热点影像服务包括热点权重规则维护、热点数据升降级管理、热点数据图形化展示、时空对比及快捷下载服务功能。

影像时空数据库主题搜索引擎包括数据索引存档和查询检索两个基本功能。数据索引归档是在数据接入过程中基于接入元信息内容进行语义分词, 并创建归档数据索引库。基本检索是指根据用户输入的关键词序列, 进行语义分解和关键词智能提示, 并能够自动将关键词与领域知识库进行匹配, 生成搜索命令输出查询结果。

系统配置包括系统初始化配置管理、用户权限管理、数据权限管理、日志管理。

时空数据统计分析包括周期性热点数据报表、在库数据覆盖统计、接收计划完备性统

计、数据生命周期管理报表、及数据周期维护计划报表功能。

闪电下载及离线服务包括三个方面, 首先是针对签约客户提供远程高速下载服务, 其次针对离线业务需求提供的主题数据服务包导出功能、离线瘦客户端服务功能。

7 结论

基于元数据规范的高分遥感时空大数据智能服务平台是一套面向领域用户的规范化、综合性、一站式的数据服务系统, 低耗高效的充分开发高分专项的数据应用价值是该平台的立项初衷。本文从这几项核心需求出发, 根据高分专项元数据及字典规范的设计思路、当前用户需求特点和历史项目研发经验, 设计并初步实现了平台的核心功能, 合并入系列产品的升级版本中, 并且应用到同类的在研项目中, 降低了目标用户的专业技术门槛, 提高了数据服务系统的用户友好度和可靠性, 初步验证了高分专项元数据及字典规范合理性。平台后续的优化研究工作将主要集中在领域知识发现和主题目录扩展验证、海量时空数据查询响应优化三个方面。

作者单位

航天恒星科技有限公司 北京市 100000