

团 体 标 准

T/CSES XXXX—2021

流域水环境管理大数据存储 与交换技术规范

The technical specification for big data storage and
exchange of water environmental management of watershed
(征求意见稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国环境科学学会 发布

目 录

1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
4.1 HTTP /HTTPS	2
4.2 XML	2
4.3 FTP/SFTP	2
4.4 GML	2
5 流域水环境管理大数据组成	2
6 流域水环境管理大数据存储	3
6.1 存储体系	3
6.2 存储模式	4
6.3 存储要求	4
6.4 资源存储结构	5
7 流域水环境管理大数据交换	7
7.1 交换总则	7
7.2 交换模式	7
7.3 交换模型	8
7.4 网络环境	9
7.5 安全要求	9
7.6 交换方式	9
7.7 空间数据交换	10
附录 A（规范性） 交换接口定义	13

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由北京师范大学提出。

本文件由中国环境科学学会归口。

本文件起草单位：北京师范大学、北控水务（中国）投资有限公司、山东省生态环境规划研究院、中科宇图科技股份有限公司。

本文件主要起草人：王国强、王运涛、冯艳霞、王宝刚、黄鹏飞、朱宜、谢刚、彭岩波。

流域水环境管理大数据存储与交换技术规范

1 范围

本文件规定了流域水环境信息系统建设或集成过程中流域水环境管理大数据组成、流域水环境管理按数据存储、流域水环境管理大数据交换等内容。

本文件适用于流域水环境管理业务系统的数据存储与交换。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 18793 信息技术可扩展置标语言（XML） 1.0
GB/T 22239 信息安全技术 信息系统安全等级保护基本要求
GB/T 37722 信息技术 大数据存储与处理系统功能要求
GB/T 37737 信息技术 云计算 分布式块存储系统总体技术要求
HJ/T 416 环境信息术语
HJ 726 环境空间数据交换技术规范
HJ 727 环境信息交换技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

水环境 water environment

自然界中水的形成、分布和转化所处空间的环境。

3.2

水环境业务数据 business data of water environment

水环境保护信息化应用领域的与业务相关的数据。

3.3

水环境空间数据 spatial data of water environment

水环境保护信息化应用领域的所有与空间位置直接或间接相关的数据，包括基础空间数据和专题空间数据。

3.4

结构化数据 structured data

一种数据表示形式，按此种形式，由数据元素汇集而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。

3.5

非结构化数据 unstructured data

不具有预定义模型或术以预定义方式组织的数据。

3.6

元数据 metadata

描述数据的数据。

3.7

节点 node

参与数据交换的部门或单位。

4 缩略语

4.1 HTTP/HTTPS

超文本传输协议 (Hypertext Transfer Protocol)

4.2 XML

可扩展标记语言 (eXtensible Markup Language)

4.3 FTP/SFTP

安全文件传输协议 (Secret File Transfer Protocol)

4.4 GML

GML 地理标记语言 (Geography Markup Language)

5 流域水环境管理大数据组成

流域水环境管理大数据可分为水环境空间数据和水环境非空间数据，其中水环境空间数据包含基础空间数据和专题空间数据；水环境非空间数据包含水环境业务数据和水环境非业务数据（如图1）。流域水环境管理大数据可按结构化和非结构化形式进行存储。

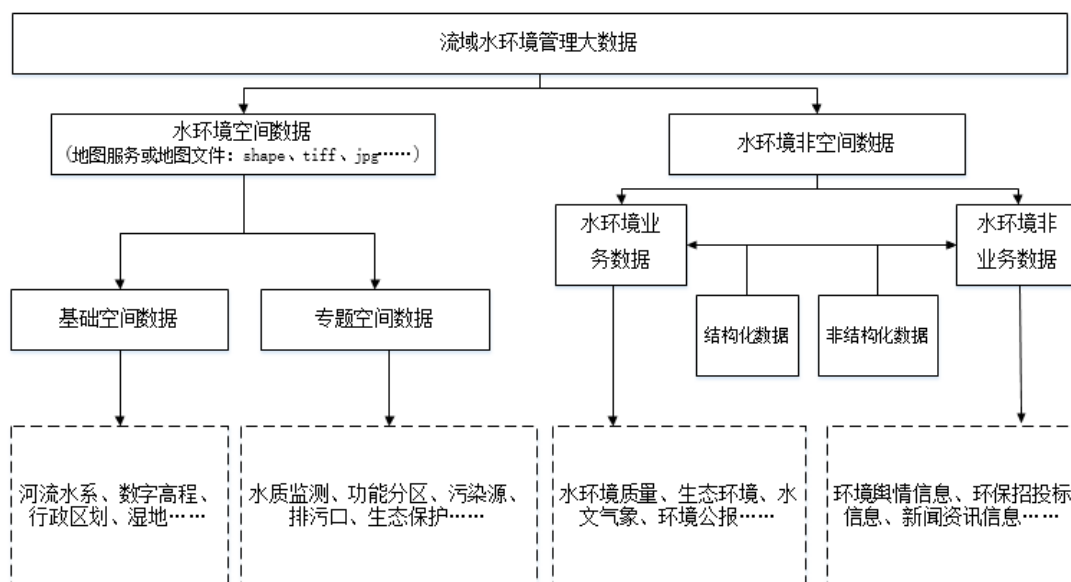


图1 流域水环境管理大数据组成

6 流域水环境管理大数据存储

6.1 存储体系

- a) 流域水环境管理大数据的数据来源可从物联网、互联网、生态环境业务专网和政务网中采集。
- b) 大数据初始化可按离线计算和流式计算处理。
- c) 大数据数据库宜按流域水环境管理的业务功能划分（如图2）。

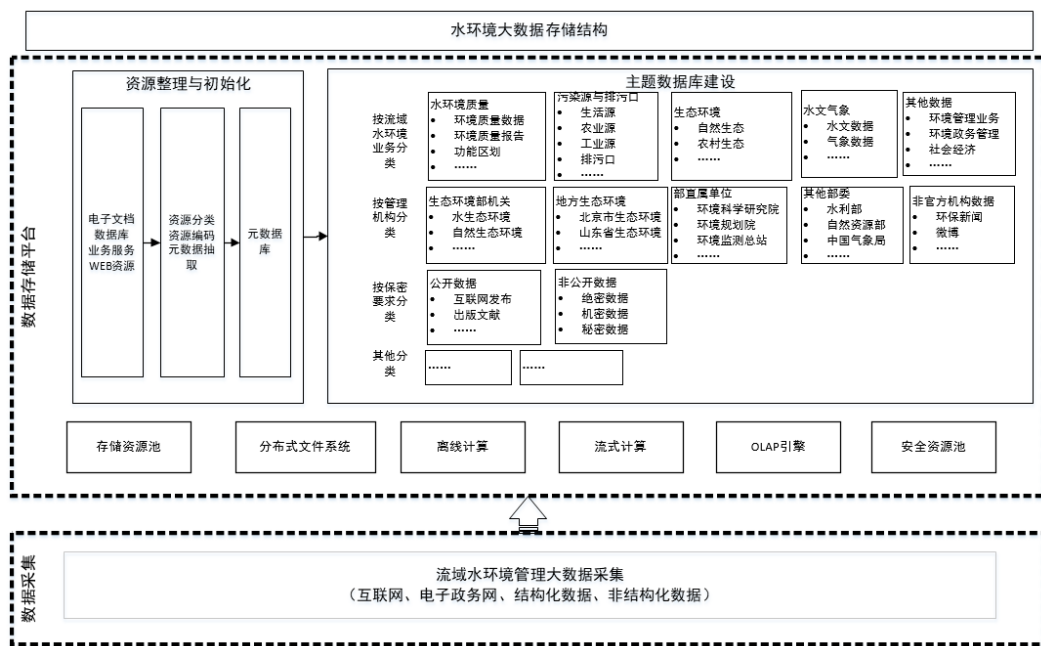


图2 流域水环境管理大数据存储体系结构图

6.2 存储模式

- 流域水环境管理大数据存储模式宜为分布式存储、集中式存储和内存存储。
- 分布式存储规模需按数据类型、数据体量而定，独立设备宜不低于5台，设备内存不低于2GB。
- 水环境管理中数据规模达GB级、流式数据和非结构数据宜按分布式存储。
- 水环境管理中分级上报、审核、汇集的业务数据宜按分布式存储。
- 集中式存储规模需按业务属性、数据内容而定，流域水环境管理核心业务数据宜按集中式存储。
- 水环境管理中的环境监察、行政处罚、复议、诉讼、日常政务等数据宜按集中式存储。
- 内存式存储规模需按数据访问频率、更新频率而定。
- 水环境管理中的在线监控等数据宜按内存式存储。

6.3 存储要求

6.3.1 流域水环境管理大数据存储模块的基本要求

- 应支持数据上传、数据下载、目录查看、目录创建、目录删除、权限修改等操作。
- 应支持标准、开放的数据访问API对数据进行操作。
- 应提供数据加载工具功能，满足大数据存储与处理系统和传统关系型数据库其他文件系统之间交换数据和文件。
- 应具备关键节点(部件)高可用性设计与要求。

- e) 宜提供流域水环境管理核心业务数据自动备份和手动备份功能。
- f) 宜支持流域水环境管理核心业务数据的批量更新、删除等数据管理功能。
- g) 宜支持流域水环境管理视频监控流式数据的实时入库，支持实时查询。

6.3.2 大数据数据存储数据节点要求

- a) 大数据可靠性存储需将数据切分成等大数据块，并存储到多台机器（数据节点）。
- b) 数据节点中主节点应包含主备节点。
- c) 数据节点应不少于 5 个数据节点。
- d) 数据节点数据块大小应不少于 128MB。
- e) 数据节点备份应不少于 3 个。
- f) 数据节点存储大小应不少于 2TB。

6.3.3 大数据数据存储安全要求

- a) 对敏感数据须进行加密，应采用对称加密或不对称加密。
- b) 包含重要、敏感或关键数据信息的传输须加密并经过专人审核。
- c) 删除敏感或关键数据时，应使用专用工具进行彻底删除。
- d) 数据入库或出库应经过授权。

6.4 资源存储结构

6.4.1 资源存储

资源存储包括资源目录中涵盖内容的结构化数据和非结构化数据的存储，两类数据的逻辑存储结构为资源目录存储结构。

6.4.2 资源存储基本结构

资源按一定的资源目录进行组织，资源目录相关的数据文件按不同的媒体类型归类，基本结构如表 1 所示：

表 1 资源目录基本结构

文件根目录	媒体文件	资源目录	说明
文件【dataFile】			文件根目录
	音频【audio】		音频文件根目录
		分组文件目录【水环境管理业务、环境政务管理、……】	流域水环境资源目录音频文件分组
	视频【video】		视频文件根目录

文件根目录	媒体文件	资源目录	说明
		分组文件目录【污染源与排口、流域生态环境、水环境管理业务、……】	流域水环境资源目录视频文件分组
	图像【image】		图像文件根目录
		分组文件目录【水环境质量、污染源与排口、流域生态环境、环境管理业务、环境政务管理、……】	流域水环境资源目图像文件分组
	文档【document】		文档文件根目录
		分组文件目录【水环境质量、污染源与排口、流域生态环境、环境管理业务、环境政务管理、社会经济、……】	流域水环境资源目录文档文件分组
	关系型数据【relationalData】		关系型数据资源根目录
		分组文件目录【水环境质量、污染源与排污口、水文与气象、水模型、社会经济、水环境管理业务、……】	流域水环境关系型资源分组
	其他【other】		其他类资源根目录
		分组文件目录【水资源、水生态、经济、……】	流域水环境其他类资源分组
注 1: 【 】代表存储时所定义的存储分类名称。			
注 2: 数据文件存于分组的对应资源目录下。			
注 3: 当某一类数据文件数量较多时, 在对应的媒体文件目录下, 可以定义多个分组文件目录, 并建立分组关系, 本部分未对分组文件目录的数量层级等做出要求。			

6.4.3 关系型资源存储

关系数据库以二维表结构对实例数据进行组织和存储。当数据存储于关系数据库中时, 每个资源作为独立的文本单元, 存储于数据库表中。表结构见表 2 所示。

表 2 资源索引表

序号	字段名	说明	类型	非空	备注
1	ID	当前资源本地标识	字符串型	是	主键
2	Code	当前资源编码	字符串型	是	唯一键
3	Description	资源描述	长文本型	是	

资源索引与实体资源表结构见表 3 所示。

表 3 资源关系表

序号	字段名	说明	类型	非空	备注
1	ID	当前关系标识	字符串型	是	主键

2	SubjectCode	资源索引编码	字符串型	是	资源索引编码
3	SubjectType	实体资源类型	字符串型	是	资源类型
4	ObjectCode	实体资源编码	字符串型	是	实体资源编码

6.4.4 非关系型资源存储

完整的非关系型资源文件可单独存储于文件系统中。如果文件中涉及媒体对象相关数据文件的具体存储位置时，该存储位置可以是当前实体文件存储位置的相对路径，也可以是其他路径。

7 流域水环境管理大数据交换

7.1 交换总则

流域水环境管理大数据交换应遵循以下要求：

- a) 涉及接入的业务流程较简单且稳定。
- b) 接入的数据相应数据结构明确清晰。
- c) 数据能够独立体现数据的业务性。
- d) 具备支持多种异构数据的接收和整合能力。

7.2 交换模式

7.2.1 点对点模式

简单的系统间数据交换宜采用点对点交换方式，如图 3 所示。

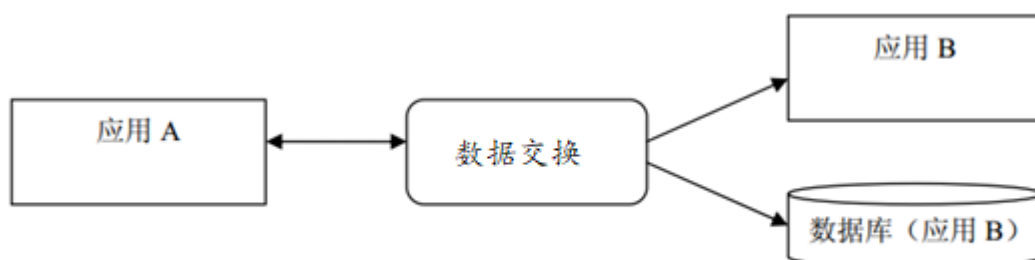


图 3 点对点模式

7.2.2 数据集成接口模式

数据的报送和接收宜采用数据集成方式，如图 4 所示。

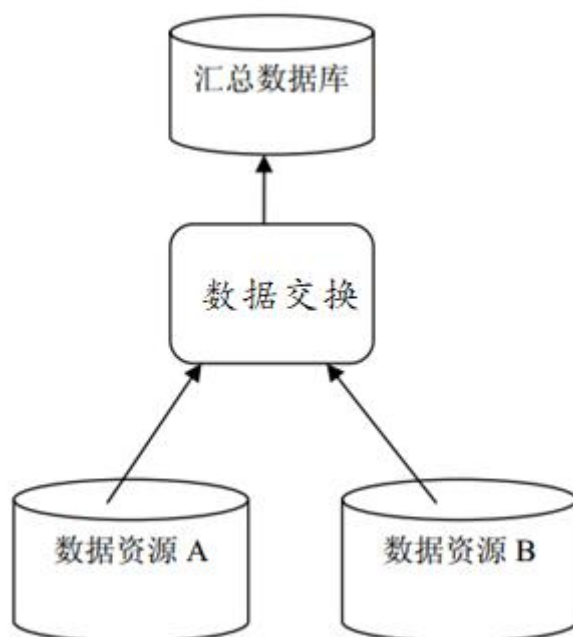


图4 数据集成接口模式

7.2.3 临时库模式

互联网或其他专网数据的交换宜采用临时库的交换方式，如图5所示。

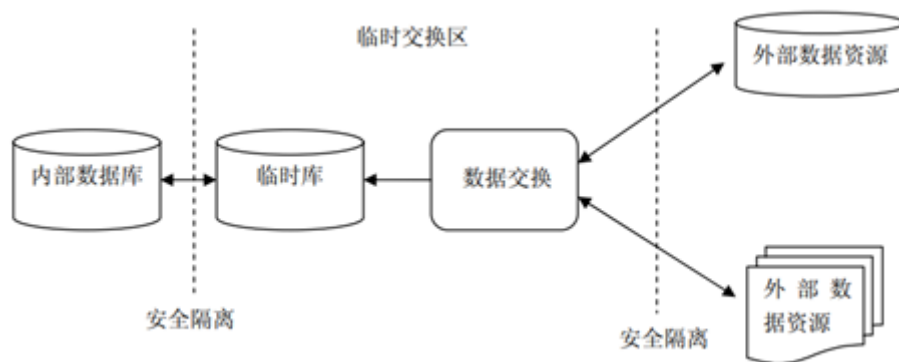


图5 临时库模式

7.3 交换模型

7.3.1 交换资源包模型

资源信息包用以封装资源对象及其所包含的单一媒体对象（相关数据文件）信息，其结构如下：

Object_Local_Code.zip/

mimetype

CONTAINER/

doc_rdf_intance.xml(实体文件)

dataFile/

[media_type_label]/

某一媒体类型数据文件

“Object_Locall_Code.zip”为资源对象的本地标识符；交换信息包以“Object_Locall_Code.zip”命名，采用 ZIP 压缩格式对信息包进行压缩。

“mimetype”资源的类型。

7.3.2 传输要求

- a) 应采用数据交换格式标准，宜采用 JSON 形式进行传输。
- b) 数据传输内容，应满足基于 XML 的资源包结构要求。

7.4 网络环境

流域水环境管理业务包含多种网络环境的数据，交换模式宜支持互联网环境交换、互联网环境与生态环境业务专网交换、生态环境业务专网内交换。

7.5 安全要求

信息交换的安全性和准确性要求，应遵从以下 5 个方面：

- a) 数据保密：防止信息被截获或非法存取而泄密。
- b) 对象认证：通信双方对各自通信对象的合法性、真实性进行确认，以防第三者假冒。
- c) 数据完整性：阻止对交换数据的非法修改、插入、删除并防止数据丢失。
- d) 防抗抵赖：用于证实已发生过的操作，防止交换双方否认发生的行为。
- e) 访问控制：防止非授权用户非法访问和使用系统资源。

互联网环境与电子政务网交换、电子政务网内交换宜采用前置机模式，具体要求应遵循 GB/T 22239 的规定。

7.6 交换方式

7.6.1 FTP/SFTP 交换

文档、多媒体数据等文件型数据宜采用 FTP/SFTP 方式进行交换。

7.6.2 接口交换

应用系统之间可直接利用交换接口传送“交换信息包”文件来进行交换，交换接口定义规定参见附件 A，所包含接口如下：

- 资源列表获取接口(ListObject)：获取符合查询条件的资源标列表。
- 资源数据获取接口(getMetadata)：获取指定资源的数据，支持多种元数据格式。
- 数据流获取接口(getDataStream)：获取数据文件对应的数字内容的字节流。
- 资源上传接口(uploadZIP)：将资源包及交换信息包上传至资源存储系统。

7.7 空间数据交换

流域水环境空间矢量数据交换，应遵循 HJ 726 的规定。

7.7.1 空间数据交换模型

水环境空间数据交换模型通过将环境空间数据接入所需的适配器，如 FTP 适配器、GML 适配器、空间数据库适配器、空间服务适配器等，同时依托交换平台将前置机上用于交换的环境空间数据进行传输交换，如图 6 所示。

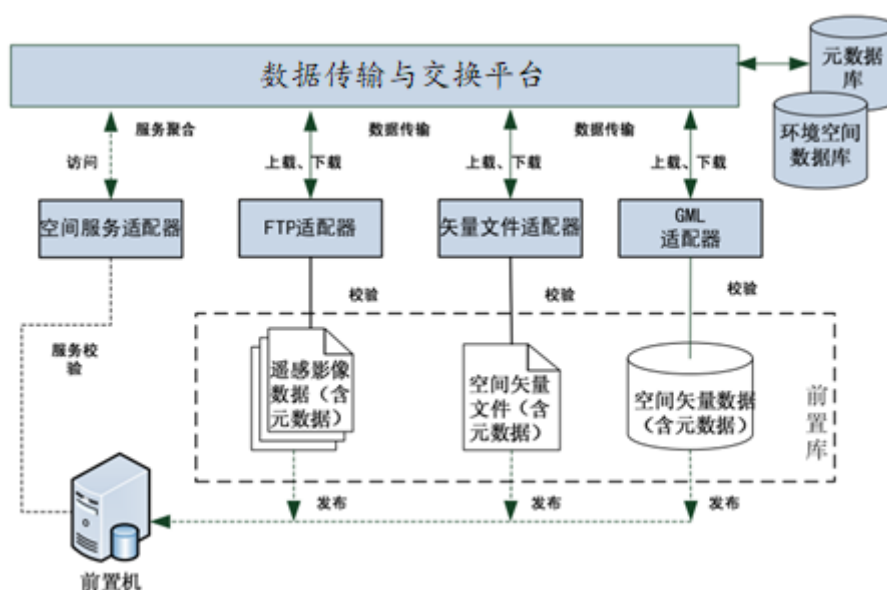


图 6 空间数据交换模型

7.7.2 空间数据交换格式

水环境矢量数据包括矢量数据库、矢量文件两种，矢量数据库宜采用 GML 作为交换格式，环境矢量文件包括 .shp、.tab、.CVS 等，需支持格式间的转换。

水环境栅格数据宜采用通用性强，且无损无压缩的 GeoTIF 格式。

水环境空间数据附加的数据信息宜采用 GeoJson 格式进行交换。

7.7.3 空间数据交换流程

空间数据交换流程可分为下载流程（如图 7）、上传流程（如图 8）。

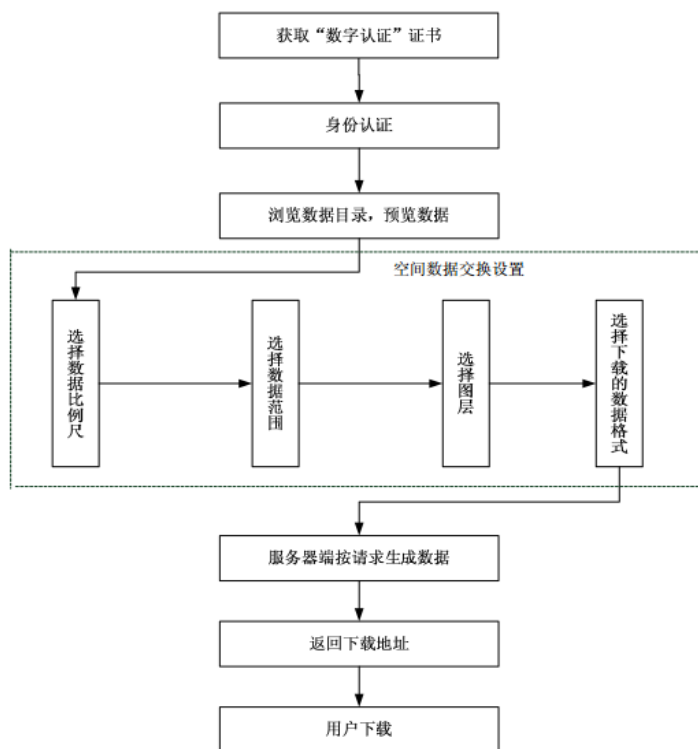


图 7 下载流程

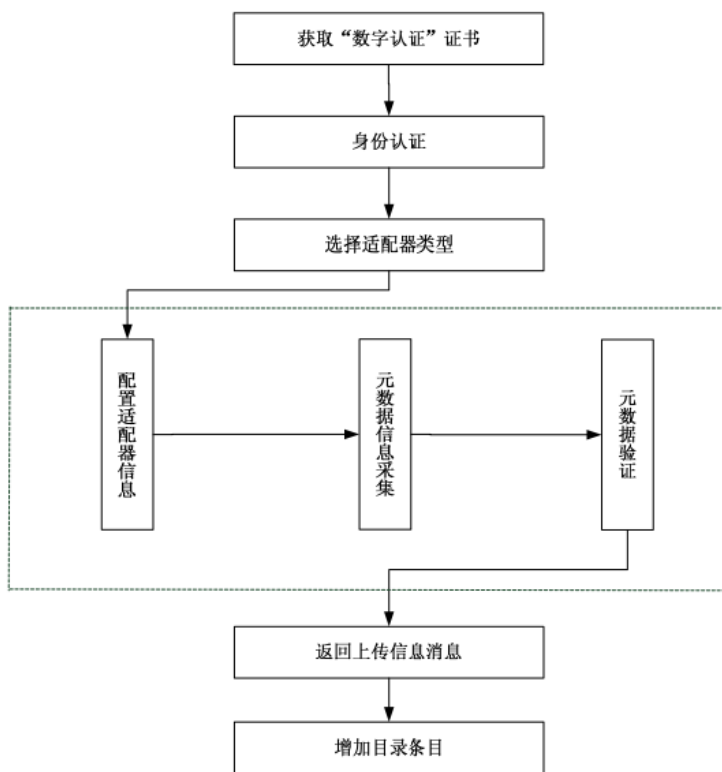


图 8 上传流程

7.7.4 空间数据交换质量

空间数据交换质量应满足如下要求：

- a) 空间数据实体无丢失。
- b) 空间实体的空间分辨率符合要求。
- c) 空间实体属性内容无缺失。
- d) 栅格数据的格网单元（或像元）尺寸、行列数以及单元值应保持不变。

附录 A
(规范性)
交换接口定义

A.1 资源列表获取接口

表 A.1 资源列表获取接口定义

项名	说明
名称	ListObject
描述	返回符合条件的资源列表
句法	http://hostname:port/listobject?type=IDType&value=IDValue
参数	type: 必备, 资源标识符类型 value: 必备, 标识符的值

A.2 资源数据获取接口

表 A.2 资源数据获取接口定义

项名	说明
名称	getMetadata
描述	通过资源的唯一编码获取资源的实体内容
句法	http://hostname:port/getMetadat?ID=resourceID&format=metaFormat
参数	Format: 可选, 表示元数据格式。 注: format 为空时返网当前资源的所有格式的元数据文件的 ZIP 格式压缩包

A.3 数据流获取接口

表 A.3 数据流获取接口定义

项名	说明
名称	getDataStream
描述	通过数据文件的内容链接获取其对应的数字内容的字节流
句法	http://hostname:port/getDataStream?url=contentURL
参数	url: 必填, 数据文件的内容链接

A.4 资源上传接口

表 A.4 资源上传接口定义

项名	说明
名称	uploadZIP
描述	将 EIP 上传至服务器或数据库
句法	http://hostname:port/uploadZIP

项名	说明
请求方法	Post
发送内容	ZIP 的二进制内容以 multipart/ form-data 方式编码，对应的 name 为 zip. Filename， Zip.filename 为本地真实文件名