

团 体 标 准

T/GDDL 20—2024

地理科学数据共享平台元数据编写规范

Specification for metadata writing of geographic science data sharing  
platform

2024-06-28 发布

2024-06-30 实施



## 目 次

前 言 .....	II
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 元数据编写规范 .....	2
5 质量检查与发布 .....	5

全国团体标准信息平台



## 前 言

本文件按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。

本文件由广东省科学院广州地理研究所提出。

本文件由广东省地理学会归口。

本文件起草单位：广东省科学院广州地理研究所、广东工业大学、广西壮族自治区海洋环境监测中心站、南方海洋科学与工程广东省实验室（广州）、广州中科云图智能科技有限公司、中科珠江（广州）技术有限公司、自然资源部深圳海洋中心、广东省有色地质测绘院。

本文件主要起草人：荆文龙、侯志伟、舒思京、杨骥、周文能、邓应彬、潘屹峰、蓝文陆、王丽丽、杨传训、洪鸣、周敬峰、赵晓丹、孙嘉、杨爽、李勇、陆山、胡泓达、徐健淞、黄国娟、黄吴蒙、尹小玲、王晓旭、邓丽明、魏巍、雷建锋、吴燕华、魏雷、邵光贺、王聪强、王梓斌、罗涛、刘伟。



# 地理科学数据共享平台元数据编写规范

## 1 范围

本文件规定了地理科学数据共享平台中数据集元数据的编写规范、质量检查与发布。

本文件适用于地理科学数据采集、处理、存储、交换，以及各类地理科学数据共享平台、地理信息系统的元数据著录和发布。其他学科亦可参考使用。

## 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 2659.1-2022 世界各国和地区及其行政区划名称代码 第1部分：国家和地区代码

GB/T 7408.1-2023 日期和时间 信息交换表示法 第1部分：基本原则

GB/T 19710.1-2023 地理信息 元数据 第1部分：基础

GB/T 31075-2014 科技平台 通用术语

## 3 术语和定义

GB/T 31075-2014界定的以及下列术语和定义适用于本文件。

### 3.1

#### 科学数据 scientific data

在自然科学、工程技术科学等领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据。

### 3.2

#### 数据实体 data entity

记录数据的最基本单元。

例：一个压缩包文件、一个ARC/INFO-COVERAGE, GRID, 一个EXCEL文件，一幅数字遥感影像文件等称为一个数据实体。

### 3.3

#### 数据集 dataset

在数据的专题内容、数据源、数据格式、数据加工整理方法等方面具有一定关联的数据实体的集合。一个数据集可以包含一个或多个数据实体。

### 3.4

#### 元数据 metadata

描述数据的数据。即数据的标识、覆盖范围、质量、空间和时间模式、空间参照系和分发等信息。

[来源：GB/T 19710-2023]

### 3.5

#### 元数据元素 metadata element

元数据的基本单元。

[来源：GB/T 19710-2023]

### 3.6

**核心元数据元素 core metadata element**

一组说明数据系统特性的元数据元素。

[来源：GB/T 19710-2023]

## 4 元数据编写规范

### 4.1 数据集名称

数据集名称应包含时间、区域、精细程度（例如分辨率）、专题要素等内容，其顺序可根据习惯排列。

注：其中时间是指原始数据获取的时间。精细程度是指数据描述的详细程度，可以是分辨率、重要程度、等级或数量，同时符合 GB/T 7408.1-2023 的相关规定。

### 4.2 数据集摘要

#### 4.2.1 内容说明

数据集摘要部分必须包含数据来源以及数据的加工生产与整理方法的描述。

#### 4.2.2 数据来源

数据来源应描述原始数据的出处。数据来源信息的编写规则如下：

- a) 数据如果来自正式出版的文献，应说明文献的名称、出版单位、出版时间；
- b) 数据如果来自非正式出版的文献，应当说明文献的名称、时间；
- c) 数据如果来自未正式出版的课题成果，应当说明课题的名称、性质、课题的执行时间；
- d) 数据来源于其它渠道（机构、网络），应当详细说明。数据源的说明应具备可复查性，不能笼统地以“科学考察”、“课题研究”等方式说明。

#### 4.2.3 数据的加工生产与整理方法

描述在数据源的基础上进行了哪些加工，使用户了解数据的形成过程，在此基础上用户认识数据对其研究的可用性。一般加工方法的描述应简单、明了，方法应当具备可操作性和可重复性，特别复杂的方法同时应当指明相关的参考文献。

### 4.3 数据集核心元数据

核心元数据为数据集元数据描述中主要的元素，可用于数据集编目、数据交换网站活动和对数据集的描述。核心元数据包括标识信息、内容信息、分发信息、数据质量信息、图示表达目录信息、数据模式信息、元数据维护信息和元数据参考信息。数据集核心元数据内容如图 1 所示。

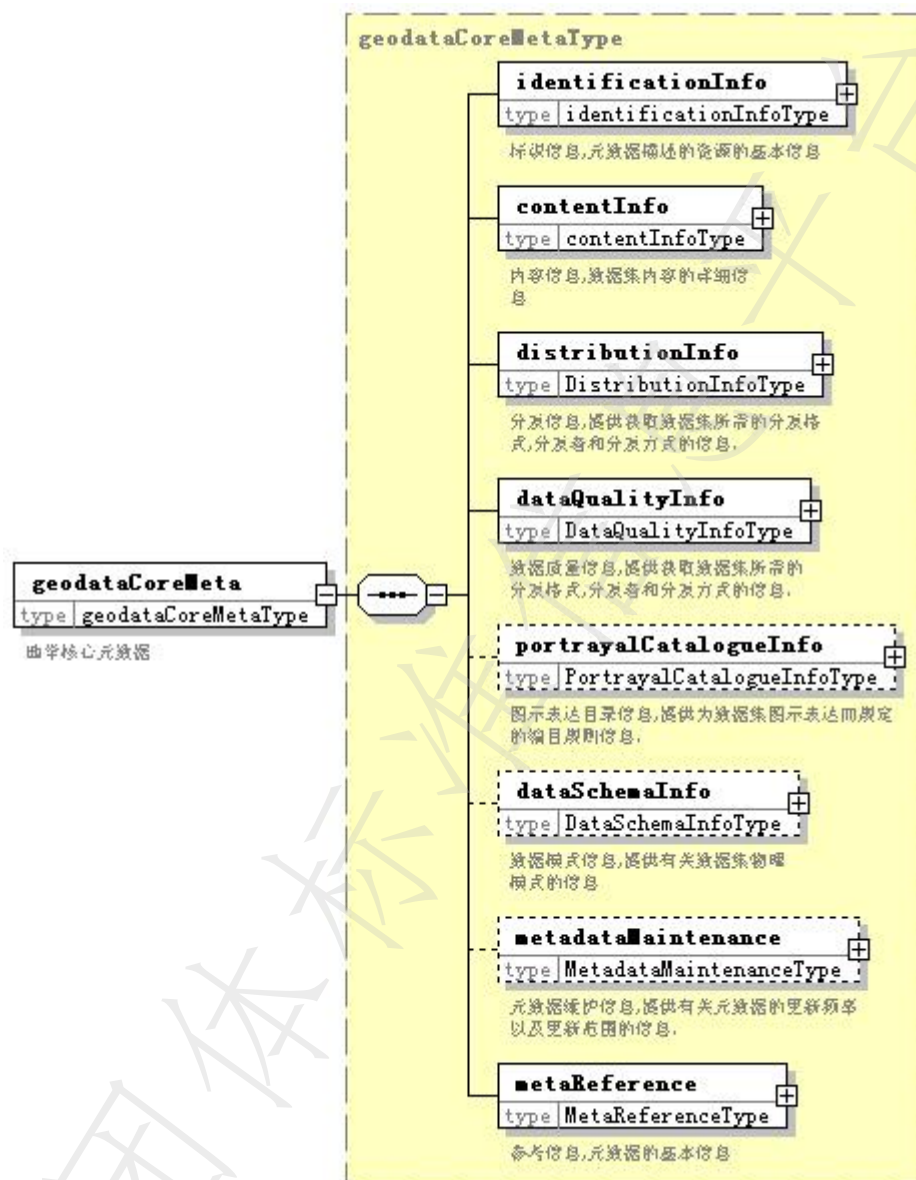


图 1 数据集核心元数据内容

#### 4.3.1 标识信息

4.3.1.1 标识信息包含唯一标识数据的信息，内容包括资源的引用、数据集摘要、目的、状态和联系办法等信息。

4.3.1.2 标识实体是必选的。

4.3.1.3 标识信息的必选元素内容包括：

- a) 数据集名称；
- b) 数据集日期；
- c) 数据集摘要；
- d) 数据集语种；
- e) 数据集维护和更新频率；
- f) 关键词。

4.3.1.4 标识信息的可选元素内容包括：

- a) 数据集负责方，包括数据集负责方姓名、单位、电话信息、传真、地址、详细地址、城市、行政区（按 GB/T 2659.1-2022 的规定）、邮政编码、国家、电子邮件地址等；
- b) 数据存储格式，包括名称与版本（日期、版本号等）；
- c) 数据集字符集，即字符编码标准全称；
- d) 数据集开发的说明；
- e) 数据集状况，即数据集生产与完成情况；
- f) 引用资源的表达方式；
- g) 数据集维护信息与维护方联系信息；
- h) 浏览图，即概要性说明数据集及其图例的图形，包括文件名称、文件说明、文件类型；
- i) 关键词说明与关键词词典基本信息；
- j) 数据集限制信息，包括访问限制、使用限制、安全限制分级。

#### 4.3.2 内容信息

4.3.2.1 内容信息提供数据内容特征的描述信息。内容信息实体是可选的。内容信息的“资源域”属性是制定领域科学数据共享元数据内容标准时必选的元数据元素，以表明数据集所在的资源范围。

4.3.2.2 内容信息的必选要素包括：

- a) 资源域，即数据资源所在的资源范围；
- b) 数据集分类名称。

4.3.2.3 内容信息的可选要素包括：

- a) 空间范围，包括地理名称、平面范围（西经、东经、北纬、南纬、坐标系统、投影）、垂向范围；垂向范围又包括垂向最大值、垂向最小值、高程系统；
- b) 时间范围，包括开始时间、结束时间；
- c) 空间分辨率；
- d) 采样间隔。

#### 4.3.3 分发信息

4.3.3.1 分发信息包含数据集分发格式信息、分发者信息以及用户获取数据集的途径。

4.3.3.2 分发信息实体是可选的。

4.3.3.3 数据集分发信息内容包括：

- a) 分发格式的名称和版本；
- b) 分发者联系信息；
- c) 传送选项（包括传送量、传送方式，如在线、离线）。

#### 4.3.4 数据质量信息

4.3.4.1 辐射定标应根据遥感器定标公式和各波段定标系数进行。

4.3.4.2 数据质量信息包含数据集的数据志信息或/和数据质量报告。

4.3.4.3 数据质量信息实体是可选的。

4.3.4.4 数据志是非量化的数据质量信息，包括数据志信息的一般说明、处理步骤、数据源。

4.3.4.5 数据处理步骤包括的内容如下：

- a) 处理步骤说明；
- b) 基本原理，即处理步骤的要求或目的；
- c) 日期时间；
- d) 处理者；
- e) 数据源。

4.3.4.6 数据质量报告是数据的量化的质量信息，主要是数据的多余性，即数据集中超出范围规定的多余数据。数据质量报告包括的内容如下：

- a) 度量名称；
- b) 度量说明；

- c) 评价方法类型；
- d) 评价方法说明；
- e) 评价程序；
- f) 日期时间；
- g) 评价结果；
- h) 量化评价结果；
- i) 评价人。

#### 4.3.5 图示表达目录信息

图示表达目录信息提供为数据集图示表达而规定的编目规则信息，包括图示表达目录引用的参考文献目录。图示表达目录信息实体是可选的。

#### 4.3.6 元数据扩展信息

扩展信息是领域在本文件中定义的公共元数据和参考元数据之外扩展的元数据内容的描述。扩展信息实体是可选的。

#### 4.3.7 数据模式信息

数据模式信息包含有关数据集物理模式的信息。数据模式信息实体是可选的。数据模式信息包括数据模式名称、模式语言、约束语言。

#### 4.3.8 元数据维护信息

维护信息包含有关元数据或数据集的更新频率及更新范围的信息。更新频率可以是连续、按日、按周、按两周、按月、按季、按半年、按年、按需要、不固定、无计划或未知。

#### 4.3.9 元数据参考信息

4.3.9.1 有关元数据自身的描述信息。

4.3.9.2 元数据参考信息的必选项包括：元数据标识符、元数据语种、元数据字符集、元数据联系方、元数据创建日期；

4.3.9.3 元数据参考信息的可选项包括：元数据标准名称和元数据标准版本。

### 4.4 元数据粒度

#### 4.4.1 内容说明

元数据粒度是指元数据编写的粗细程度，即在哪个层次上（数据实体、数据集、数据库）编写元数据。

#### 4.4.2 编写规则

在元数据编写过程中遵循以下几条原则：

a) 数据实体的内容、来源、格式及数据产生方法完全相同，只是区域、时间上的差异，应将这些数据实体组织成一个数据集编写一条元数据。有新的数据实体增加时，更新元数据，而不是新增元数据。

b) 数据内容相同，但格式、加工方法不同：如矢量格式的土地利用数据、分县土地利用数据、公里网格土地利用数据等，应分别按不同的数据集编写元数据。

c) 数据内容不同，但数据来源、数据格式相同，且这些数据是围绕某个相同的主题（如港口数据、岛屿数据、定位观测数据、某些出版物（不含统计年鉴）中的数据），可视为一个数据集，编写一条元数据。

## 5 质量检查与发布

## 5.1 质量检查方式

5.1.1 对编写的元数据，采用基于规则的系统自动检查与工作人员人工检查相结合的方式。首先根据本文件第4部分规定，编写共享平台的数据检查规则，对编写提交的元数据进行初步检查。

5.1.2 提交并通过系统自动检查后，由共享平台数据审核员人工进行检查，给出检查结论，并编写质量检查反馈意见。

## 5.2 元数据的发布

经质量检查不合格的元数据不予发布，元数据编写人员必须进行修改，并再次检查合格才予发布。