

ICS

CCS 点击此处添加 CCS 号

DB51

四川省地方标准

DB51/T XXXX—XXXX

卫星遥感解译样本数据库建设规范

Specification for sample database construction of satellite remote
sensing image

（征求意见稿）

XXXX—XX—XX 发布

XXXX—XX—XX 实施

四川省市场监督管理局 发布

目 次

前 言 II

引 言 III

1 范围 4

2 规范性引用文件 4

3 术语和定义 4

4 缩略语 5

5 总体技术要求 5

 5.1 总体原则 5

 5.2 数学基础 6

 5.3 数据格式 6

6 数据内容与质量要求 6

 6.1 数据组成 7

 6.2 数据内容 7

 6.3 数据质量要求 7

7 数据库设计 8

 7.1 概念设计 8

 7.2 逻辑设计 8

 7.3 物理设计 9

8 建库流程 9

 8.1 数据准备 9

 8.2 样本处理 9

 8.3 库体创建 10

 8.4 样本入库前检查 10

 8.5 数据入库 10

9 质量检查 10

 9.1 检查制度 10

 9.2 检查方法 11

 9.3 检查内容 11

附 录 A （规范性） 卫星遥感解译样本标识符 12

附 录 B （资料性） 卫星遥感解译样本数据库数据表 14

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。

本文件由四川省测绘地理信息局提出、归口并解释。

本文件起草单位：自然资源部第三地理信息制图院、四川省国土科学技术研究院（四川省卫星应用技术中心）、中国科学院成都山地灾害与环境研究所。

本文件主要起草人：周启、刘立、董先敏、葛洁、曹伟超、王德富、李永鑫、范建蓉、廖洋洋、白欣、雷瑜、张茜戡。

引 言

为促进四川省卫星遥感解译样本的有序积累，指导全省卫星遥感解译样本数据建库工作的开展，满足卫星遥感解译样本数据库建设各项工作中的标准化和规范化的工作需求，特制定本技术规范。

四川省因其特有的地形特征，地表起伏悬殊，地形复杂多样，对自然资源管理工作提出更高的需求。传统的自然资源要素供给方式是基于卫星影像人机交互式目视解译，在处理时效、解译精度和数据价值挖掘方面存在掣肘。缺乏大规模、多类型、多尺度的卫星遥感解译样本数据库，无法实现面向特定需求开展样本数据的检索与管理，限制了人工智能技术在遥感领域的应用深度。为了向四川省自然资源管理工作提供高质量的数据供给，亟需研制面向省域典型地理特点的解译样本数据库建设规范，构建解译样本组织与存储模型，建立解译样本的入库流程，实现大规模、多类型、多尺度的卫星遥感解译样本数据库的构建，促进解译样本在自然资源行业的有序积累，助推四川省自然资源要素保障工作的高质量开展。

本文件共分为八部分，包括范围、规范性引用文件、术语和定义、总体技术要求、数据内容与质量要求、数据库设计、建库流程与质量检查等内容与要求。

卫星遥感解译样本数据库建设规范

1 范围

本文件规定了卫星遥感解译样本数据库建设的总体技术要求、数据内容、数据库设计、建库流程和质量检查等内容。

本文件适用于指导四川省范围内卫星遥感解译样本（以下简称为“样本”）数据库建设、更新与共享，其他相关样本数据库建设可参照执行。

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 33453-2016 基础地理信息数据库建设规范
- GB/T 39477-2020 信息安全技术 政务信息共享 数据安全技术要求
- GB/T 39608-2020 基础地理信息数字成果元数据

3 术语和定义

以下术语和定义适用于本文件。

3.1

样本影像 sample image

一定区域范围内影像数据。

3.2

样本标签 sample label

与样本影像对应的标识特定类别信息的矢量或者栅格数据。

3.3

样本标识符 sample identifier

一定范围内唯一识别样本的符号或代号，标识符通常有一定的命名规则。

3.4

样本数据集 sample data set

一组特定样本类别的由样本影像和样本标签组成的数据集合。

3.5

原始样本数据集 original sample data set

样本影像与矢量样本标签组合的数据集合，是样本生产的数据成果。

3.6

标准样本数据集 standard sample data set

原始样本数据集中的样本影像和样本标签经过影像位深、样本尺寸标准化处理形成的结构化数据。

3.7

训练样本数据集 management sample data set

标准样本数据集中的样本标签经过栅格化处理形成的适用于机器学习的结构化数据。

3.8

元数据 metadata

关于数据的数据，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。
[GB/T 39608-2020，定义3.3]

4 缩略语

下列缩略语适用于本文件。

IMG：一种图像数据格式（ERDAS IMAGE Image File Format）

TIFF：标签图像文件格式（Tag Image File Format）

GDB：文件地理数据库格式（ESRI File Geodatabase）

MDB：个人地理数据库格式（Personal Geodatabase）

GeoJSON：基于JavaScript对象表示法对地理数据结构进行编码的格式（Geo-JavaScript Object Notation）

NonSQL：非关系型的结构化查询语言（Not Only Structured Query Language）

XML：可扩展标记语言（eXtensible Markup Language）

5 总体技术要求

5.1 总体原则

5.1.1 完整性要求

样本数据库建设在技术指标、标准体系、成果模式、数据库模式等方面应充分考虑到不同的遥感应用需求；

样本数据库成果内容完整，包含标准样本数据集或者训练样本数据集以及样本描述元数据；

样本数据库成果可以正常查看。

5.1.2 规范性要求

- a) 样本标识符、样本结构描述、标签字段内容及取值按照本技术规范执行；
- b) 样本标识符与样本内容应保持逻辑一致性，样本标签属性内容应保持逻辑一致性；
- c) 样本影像与样本标签的数学基础应符合本技术规范内容；
- d) 空间要素拓扑关系正确。各要素不存在线段自相交、碎片以及不规则多边形、悬挂点或伪节点等拓扑错误；
- e) 元数据满足本技术规范要求。

5.1.3 安全性要求

在数据库设计、建立、系统运行和管理等方面都应有严格的安全和保密措施，确保整个数据库系统的安全、正常和有效地运行和使用。

数据安全应符合GB/T 39477-2020的规定。

5.1.4 系统性要求

样本数据库建设应总体符合GB/T 33453-2016的规定，在技术指标、标准体系、成果模式、库体结构、服务方式等方面具有系统性，数据库系统整体上具有良好的集成性。

5.1.5 先进性要求

充分利用当前先进、实用的技术手段，采用成熟的设计方案、技术标准、硬件平台和软件环境，实现对多尺度、多数据源、多时相样本数据的无缝管理，保障系统稳定、可靠地运行。

5.1.6 扩展性要求

数据库建设时应考虑后期运行维护阶段数据扩展工作，数据库服务器的存储空间、数据库支撑软件等应预留相应的容量扩展接口和功能接口。

5.2 数学基础

时间参考：日期采用公历纪元，时间采用北京时间。

坐标系统：采用2000国家大地坐标系。

高程基准：采用1985国家高程基准。

计量单位：面积计算单位采用平方米（m²），保留两位小数。

5.3 数据格式

原始样本数据集的样本影像应为IMG、TIFF等常见的栅格数据格式，样本标签应为ESRI Shapefile或者MDB或GDB要素类等矢量数据格式。

标准样本数据集的样本影像应为IMG、TIFF等常见的栅格数据格式，样本标签应为ESRI Shapefile、GeoJSON等常见的矢量数据格式。

训练样本数据集的样本影像应为IMG、TIFF等常见的栅格数据格式，样本标签应为IMG、PNG等常见的栅格数据格式。

标准样本数据集入库应采用NonSQL非关系型数据库存储，满足适用于机器学习的训练样本数据集的导出。

样本元数据文件应为XML格式，采用元数据项英文名称作为元数据项标记。

6 数据内容与质量要求

6.1 数据组成

卫星遥感解译样本数据库存储若干样本类别的标准样本数据集，包括地物分类样本数据集、变化检测样本数据集、目标检测样本数据集、场景分类样本数据集和其他专题样本数据集。每一类样本数据集由若干张样本影像、样本标签和样本元数据组成。

6.2 数据内容

6.2.1 地物分类样本数据集

地物分类样本数据集是若干个地物分类样本的数据集合。地物分类样本是将单时相遥感影像的每个像素关联至预先定义的地表覆盖类别。样本数据包括单时相遥感影像和其对应的逐像素分类矢量标签。

6.2.2 变化检测样本数据集

变化检测样本数据集是若干个变化检测样本的数据集合。变化检测样本是根据不同时相的遥感影像来提取、描述感兴趣物体或现象随时间变化的影像特征。样本组成包括融合多时相的遥感影像和描述变化信息的矢量标签。

6.2.3 目标检测样本数据集

目标检测样本数据集是若干个目标检测样本的数据集合。目标检测样本是根据单时相遥感影像来提取与描述感兴趣物体或现象的影像特征。样本组成包括单时相遥感影像和描述感兴趣信息的矢量标签。

6.2.4 场景分类样本数据集

场景分类样本数据集是若干个场景分类样本的数据集合。场景分类样本根据单时相遥感影像来提取、描述感兴趣场景的影像特征。样本组成包括场景区域对应的单时相遥感影像和场景分类矢量标签。利用遥感影像，综合识别场景中多类地物及其空间关系，对整体语义场景进行分类标注。

6.2.5 其他专题样本数据集

其他专题样本是根据样本应用场景，按需建设符合机器学习内容的样本数据。

6.2.6 元数据

元数据内容记录样本文件的样本基本信息、区域信息、影像基本信息、标签基本信息和产品归属等方面的信息。具体内容如下：

- a) 样本基本信息：记录样本类型、样本尺寸、样本标识符、分类编码、分类名称、空间参考等；
- b) 区域信息：记录样本区域行政区划、地形特征、区域特征等；
- c) 影像基本信息：记录影像来源、影像分辨率、影像通道数、影像时相、影像色彩模式等；
- d) 标签基本信息：记录标签类别数、类别字段索引等；
- e) 产品归属：记录样本生产单位、样本采集人员和样本采集时间等。

6.2.7 数据字典

数据字典是对元数据的总体描述，对元数据每项内容的结构与做出详细的说明。

6.3 数据质量要求

6.3.1 入库前数据质量要求

入库前样本数据应满足以下要求：

- a) 空间坐标系：样本数据的数学基础应与5.2技术要求保持一致；
- b) 完整性：样本数据应为标准样本数据集，样本标识应符合规范，样本标签和影像应完整正确，按照规范结构组织，不应有遗漏；
- c) 概念一致性：样本的分类编码、文件结构、属性构成应保持一致，不同类型数据应符合统一的体系规则；
- d) 格式一致性：样本各类数据格式应与规定格式保持一致；
- e) 影像质量：样本影像的色彩特征应能明显体现地物特征，不存在不清楚、噪声、模糊和扭曲情况；
- f) 属性精度：样本标签的属性值应保持正确性；
- g) 几何位置精度：样本标签边界与影像套合精度应符合技术规定，变化检测样本的多时相影像几何配准精度应符合技术规定；
- h) 逻辑一致性：样本标识符与样本内容应保持逻辑一致性，样本标签属性内容应保持逻辑一致性；
- i) 拓扑一致性：样本标签几何应保持正确的拓扑关系；
- j) 元数据：元数据项填写应真实、准确、完整，当元数据项无值时，应填写“无”。

6.3.2 数据库数据质量要求

样本数据的入库及处理应满足下列要求：

- a) 入库数据应为标准样本数据集，样本是可读性文件格式，标签格式符合设计要求，统一样本标签格式；
- b) 数据按照要求进行组织；
- c) 数据应保证内容的完整性。

7 数据库设计

7.1 概念设计

数据库概念设计包括对不同类别卫星遥感解译样本的归类、综合、抽象等，可用数学模型的方法描述现实世界，建立的概念数据模型不依赖于数据库软硬件环境。卫星遥感解译样本数据库概念设计的具体流程为：

- a) 对各类卫星遥感解译样本数据进行归类、抽取；
- b) 确定各类卫星遥感解译样本组织形式、样本类别属性以及样本标识符结构。

7.2 逻辑设计

数据库逻辑设计需要确定不同类别的遥感解译样本数据与元数据信息的组织形式，建立符合样本数据存储内容的逻辑模型。

7.2.1 数据关联与组织

不同类别遥感解译样本应建立区域逻辑关联，同一类别与不同类别的遥感解译样本采用优化的数据结构和组织方法，减少数据冗余。

- a) 样本数据组织：按照样本类别分库存储，不同类别样本库中的样本影像数据和样本标签数据以非结构化文件形式存储在NonSQL非关系型数据库中；

- b) 元数据组织：将样本标识符信息和元数据文件进行映射解析，形成样本描述信息，采用数据库表组织方式存储，通过标识符实现元数据与样本数据的关联。

7.2.2 数据库逻辑结构

样本库按照存储对象的不同分为样本影像表、样本标签表和元数据表。其中样本影像表存储样本影像数据，样本标签表存储样本标签数据，样本影像表与样本标签表均通过样本标识符与元数据表进行关联。

7.3 物理设计

数据库物理设计需要确定数据库选型和库体结构。

7.3.1 数据库选型

数据库选型应充分考虑安全、稳定、兼容、可扩展,支持复杂数据类型、面向对象的数据建模、海量数据管理、数据备份、数据库恢复、安全管理、并行处理和并发控制，具有分布式的数据管理和动态存储空间管理以及良好的开发环境、性能价格比和服务支持能力。

7.3.2 库体结构设计

数据库库体结构设计应包括：

- a) 基于逻辑设计提出的逻辑模型，按照软硬件配置、数据量估算，分配数据库、软件、工作区的物理空间，确定各种样本数据的目录结构和存储位置；
- b) 遵循科学、实用、简洁和可扩展性的原则，依据数据库概念结构和逻辑结构，设计数据库物理表结构，并按照各需求分配不同的表空间，各类遥感解译样本数据应分别设计各自的数据表结构；
- c) 根据命名规范和规定要求，确定各类数据表的数据项名称、类型、宽度和值域范围，选定相应的索引关键数据项。

8 建库流程

样本建库是卫星遥感解译样本数据建库付诸实施的过程，包括样数据准备、样本处理、库体创建、样本入库前检查、数据入库等步骤。

8.1 数据准备

按照样本数据库数据内容，收集所需要的遥感影像和矢量数据等相关资料，对收集的资料进行整理，分析资料作为原始样本数据集或者标准样本数据集生产参考资料的可用性，并进行相关信息记录，形成待处理的源数据。

8.2 样本处理

基于待处理的源数据进行改造生产形成符合技术规定的原始样本数据集，将原始样本数据集中的样本尺寸、影像位深进行标准化处理，生成统一命名的样本标识符，形成符合技术规定的标准样本数据集。

8.2.1 原始样本采集

若待处理的源数据已经符合原始样本的数据内容和组织规范，则无需原始样本采集步骤。

根据数据质量要求,复核精修样本标签几何;根据样本内容分类体系,如果待入库样本类别与现有样本库样本分类体系存在差异,则进行相应的类别映射及合并拆分,录入样本的标签属性信息。

8.2.2 标准样本生产

根据数据内容规范与数据质量要求,标准化样本尺寸、样本影像位深,调整样本数据命名和标签字段结构等内容,赋予唯一标识,并进行规范化组织。具体内容主要包括:

- a) 样本命名规范:调整样本数据命名内容;
- b) 样本影像规范:裁剪样本至标准尺寸,统一样本影像位深和影像色彩模式。
- c) 样本标签规范:调整标签字段结构,处理标签字段分类属性信息。
- d) 样本数据组织:根据样本库数据内容组织要求,整理组织各项标准样本数据至样本数据组织规范;
- e) 元数据完善:根据元数据项内容和填写要求,整理待入库样本的各项元数据信息,完善已有元数据文件或新建元数据文件。

8.3 库体创建

根据数据库的逻辑设计和物理设计,对每类样本数据进行物理空间的分配和相关参数的设置,创建数据表、建立数据表关联等,物理空间分配时应考虑数据库的扩充性。

8.4 样本入库前检查

8.4.1 检查内容

样本入库前,需要对样本数据进行质量检查。检查项以6.3样本入库前数据质量要求保持一致。

8.4.2 检查方法

检查方法包括计算机自动检查和人机交互检查。对于样本数据入库前检查存在问题的数据部分,需根据样本数据内容进行规范修改,直到满足各项质量检查要求。

计算机自动检查按照建库技术要求,由质量检查软件进行自动检查,形成检查日志,其中错误信息会反映在检查日志中。

人机交互检查重点检查样本质量,若样本影像与样本标签不符合设计规范,形成检查结果。

8.5 数据入库

数据入库应根据所选择的样本数据组织方式进行,将各类样本逐一通过手动添加或者程序批量入库。入库时应保证样本影像与样本标签的一一对应性。入库过程中应将样本标识符信息和样本标签属性项信息自动映射至元数据表,保证样本标签、样本影像与样本元数据基于标识符能相互关联。

数据入库完成后应在数据库日志中追加入库记录操作,详细记录入库的各项操作信息,保障数据库安全运行。

9 质量检查

质量检查针对样本入库后检查,即对已入库的各类型样本数据进行内容检查,如果检查入库后的样本数据存在错误则需根据错误类型进行修正,若无法修正的则需将错误样本数据删除,重新检查入库。

9.1 检查制度

采取切实的保证措施，严格的检查验收制，确保遥感解译样本数据库成果的规范性。

9.2 检查方法

遥感解译样本数据库检查方法包括计算机自动检查和人机交互检查。

计算机自动检查按照建库技术要求，由质量检查软件进行自动检查，形成检查日志，其中错误信息会反映在检查日志中。

人机交互检查重点检查样本质量，若样本影像与样本标签不符合设计规范，形成检查结果。

9.3 检查内容

样本入库后质量检查内容包括数据完整性和数据正确性两类检查。数据完整性检查检查入库后样本数据数量、数据是否重复入库、入库参数是否正确等内容。数据正确性检查检查样本数据中影像、标签以及元数据信息匹配情况、入库前和入库后样本数据是否一致。

附 录 A
(规范性)
卫星遥感解译样本标识符

表A. 1规定了卫星遥感解译样本的标识符结构。

表 A. 1 卫星遥感解译样本标识符

序号	标识属性	符号长度	标识内容	标识描述
1	任务类型	1	A/B/C/D	地物分类/变化检测/目标检测/场景分类
2	样本类型	1	P/N	正样本/负样本
3	样本色彩模式	1	C/R	原始模式/拉伸渲染
4	样本尺寸	1	H/L/B/M/S/O	样本尺寸大小，H/L/B/M/S/O表征2048/1024/640/512/256
5	(前时相) 影像来源	3	XXX	卫星缩写，适用于变化检测样本，其余类型样本填写000
6	(后时相) 影像来源	3	XXX	卫星缩写
7	分辨率	3	XXX	单位厘米, 999表示大于10米分辨率
8	(前)影像时相	1	1/2/3/4/5/6/7/8/9/A/B/C	月份缩写，适用于变化检测样本，其余类型样本填写0
9	(后)影像时相	1	1/2/3/4/5/6/7/8/9/A/B/C	月份缩写
10	通道数	2	XX	通道数量
11	行政单元	6	XXXXXX	区县级行政区划代码
12	地形特征	1	F/P/H/B/M	平原/高原/丘陵/盆地/山地

表 A.1 卫星遥感解译样本标识符（续）

序号	标识属性	符号长度	标识内容	标识描述
13	区域特征	1	A/B/C/D	市区/城区/近郊/远郊
14	有效占比	3	000	样本标签几何面积占比，百分位表示
15	类别字段索引	2	00/01/02/03/04	用于表示类型编码字段所在的索引值，该索引值从0开始
16	样本制作时间	13	0000000000000	标记生产时间戳，精确到毫秒

附录 B
(资料性)

卫星遥感解译样本数据库数据表

数据表按照存储对象的不同分为样本数据集表、样本数据表和样本元数据表。其中样本数据集表存储样本数据集相关信息，样本数据表存储样本影像和样本标签数据，样本元数据表存储样本元数据，样本影像表与样本标签表均通过样本标识符与元数据表进行关联。

表B. 1规定了样本数据集的表结构。

表 B. 1 样本数据集表

表名称	SAMPLE_SET_INFO				
表中文名称	样本影像表				
序号	字段中文名	字段英文名	字段类型	有无空值	说明
1	序列号	ID	Int	无	
2	样本集名称	SetName	String	无	
3	样本集类型	SetType	String	无	
4	样本小片数量	SetNum	INT	无	
5	样本集创建时间	CreateTime	Datetime	无	
6	样本集创建单位	CreateUnit	String	无	
7	样本集创建人	CreatePerson	String	无	
8	样本集消亡时间	DeadTime	Datetime		

表B. 2规定了样本数据的表结构。

表 B. 2 样本数据表

表名称	SAMPLE_INFO				
表中文名称	样本影像表				
序号	字段名	标识符	字段类型	有无空值	说明
1	序列号	ID	Int	无	
2	样本集ID	SetID	INT	无	
3	样本标识符	SampleNo	String	无	
4	样本影像数据	Raster	String	无	
5	样本标签数据	Label	String	无	
6	样本创建时间	CreateTime	Datetime	无	
7	样本消亡时间	DeadTime	Datetime		

表B. 3规定了样本元数据的表结构。

表 B. 3 样本元数据表

表名称	SAMPLE_METADATA_INFO				
表中文名称	样本元数据表				
序号	字段名	标识符	字段类型	有无空值	说明
1	序列号	ID	Int	无	
2	样本标识符	SampleNo	String	无	
3	任务类型	TaskType	String	无	
4	样本类型	SampleType	String	无	
5	影像色彩模式	ColorType	String	无	
6	样本尺寸	SampleSize	String	无	
7	分类编码	ClassCode	String	无	
8	分类名称	ClassName	String	无	
9	空间参考	SpatialRef	String	无	
10	前影像来源	BSampleFrom	String	无	
11	后影像来源	ASampleFrom	String	无	
12	前影像时相	BMonth	String	无	
13	后影像时相	AMonth	String	无	
14	影像分辨率	SampleResolution	String	无	
15	通道数	BandCount	Int	无	
16	县级行政区划代码	PAC	String	无	
17	省份	Province	String	无	
18	市州名称	City	String	无	
19	区县名称	Town	String	无	
20	地形特征	Terrain	String	无	
21	区域特征	Region	String	无	
22	标注类别数	TypeCount	Int	无	
23	类别字段索引	TypeIndex	Int	无	
24	样本有效占比	RegionRatio	Double	无	