

HJ

中华人民共和国国家环境保护标准

HJ□□□-20□□

环境数据集加工汇交流程

Process of environmental dataset's processing and collection

(征求意见稿)

20□□-□□-□□发布

20□□-□□-□□实施

环 境 保 护 部 发 布

目 次

前 言.....	II
1 适用范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 技术内容.....	2
附录 A（规范性附录） 环境数据集工作文件	13

前 言

为贯彻《中华人民共和国环境保护法》，构建环境信息的标准化体系，促进环境信息化工作，完善环境质量管理，规范环境数据集加工、汇交流程，制定本标准。

本标准规定了环境数据共享与服务建设中数据集加工、汇交的技术要求，规范了环境数据集申请、采集、制作、审核、汇总、验收和归档等业务流程。

本标准的附录 A 为规范性附录。

本标准为首次发布。

本标准由环境保护部科技标准司组织制订。

本标准主要起草单位：环境保护部信息中心、福建省环境信息中心。

本标准环境保护部 201□年□□月□□日批准。

本标准自 201□年□□月□□日起实施。

本标准由环境保护部解释。

环境数据集加工汇交流程

1 适用范围

本标准规定了环境数据共享与服务建设中数据集加工、汇交的技术要求，规范了环境数据集申请、采集、制作、审核、汇总、验收和归档等业务流程。

本标准适用于环境数据集加工汇交过程中对工作流程进行规范。

2 规范性引用文件

本标准内容引用了下列文件中的条款。凡是不注日期的引用文件，其有效版本适用于本标准。

HJ/T 416	环境信息术语
HJ/T 417	环境信息分类与代码
HJ □□□	环境信息元数据标准
HJ □□□	环境数据集说明文档格式标准
SDS/T 2111	元数据标准化基本原则和方法
SDS/T 2112	科学数据共享元数据内容

3 术语和定义

下列术语和定义适用于本标准。

3.1 环境数据集 Environmental Dataset

由环境保护业务相关的若干数据实例组成、遵循统一的数据模型、具有相对独立的业务语义的数据集合。

注：环境数据集的粒度应根据具体的应用需求来确定；可按照空间范围、时间尺度或要素划分环境数据集的分类；环境数据集所属分类可以是 HJ/T 417 规定的某一分类下的一部分或全部，也可以属于 HJ/T 417 规定的多个分类。

【HJ□□-20□□，术语和定义 3.2】（《环境数据集说明文档格式标准》发布后，补充编号，去掉此括号。）

3.2 数据集实体 Dataset Entities

由一系列数据文件组成，是数据或图形的集合，也是数据集的主体。数据文件应按照一定的格式排列，可选择 ASCII 字符、二进制数据（GRID、GRIB、BUF、netCDF 等）以及图像文件（GIF、JPEG 等）格式；数据文件的命名应符合一定的要求，使用能清晰反映该数据集内容的简短描述性字符定义文件名，以反映数据格式的通用字符做为后缀。

3.3 环境数据集实体 Environmental Dataset Entities

由一系列环境数据文件组成，是环境数据或图形的集合，也是环境数据集的主体。

3.4 数据集说明文档Dataset explanation document

环境数据集的创建单位应编写环境数据集说明文档，以准确描述环境数据集。环境数据集说明文档具有通用结构要求、内容要求和文档格式要求。

【HJ□□-20□□，4 概述】（《环境数据集说明文档格式标准》发布后，补充编号，去掉此括号。）

3.5 数据集制作过程文件Dataset making process document

数据集制作过程文件是数据集制作过程中形成的除数据集实体、元数据和数据集说明文档之外的各种文件，如记载数据集制作过程的工作单、数据集审核报告、数据集实体读取软件以及数据集制作流程的说明文件等。

4 技术内容

环境数据集加工汇交流程的规范化设计与制作，为各数据中心数据资源单位提供共享数据资源的规范化描述、加工、生产和表达的方法与措施，从而生产出内容丰富、形式统一、符合科学数据共享工程数据模式标准和数据元标准的数据集产品。

4.1 环境数据集加工汇交流程

环境数据集加工汇交流程包括：申请、采集、制作、审核、汇总、验收、数据集归档、发布，见图 1 所示。

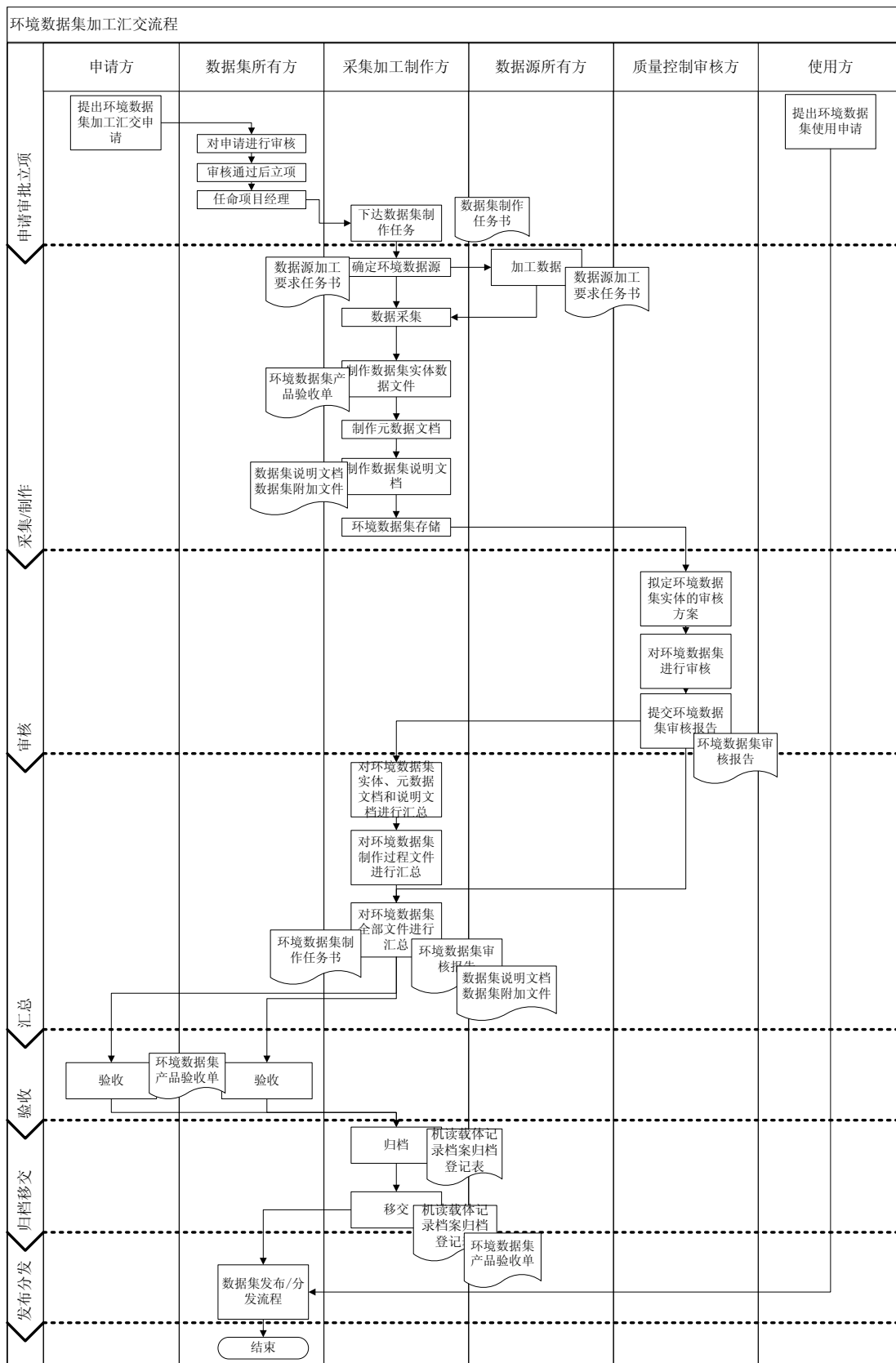


图 1 环境数据集加工汇交流程

各责任单位说明：

- 申请方，是提出数据集加工要求的各级环保部门或人员，申请方应准确、详细、全面描述数据集的数据要求和使用要求；申请方负责对数据集进行验收；
- 数据源所有方，是数据集数据来源的拥有单位，数据是一种资产，因此按照属地管理原则，数据源所有方负责向数据集加工制作方提供所需的准确的数据，负责进行数据集验收汇报；
- 数据集的所有方，是制作好的数据集的拥有单位，负责对数据集进行审批立项、验收，接收采集加工制作方移交的数据集，并负责数据集的发布和分发工作。本标准中数据集所有方为各级环境信息中心；
- 采集加工制作方，根据申请方的数据要求和使用要求，向数据源所有方提出数据加工要求，并负责完成从数据源所有方汇交加工后的数据，制做数据集实体、元数据及说明文档，协助质量控制审核方完成质量审核，完成数据集归档工作，最后向数据集所有方进行移交，负责提出数据集验收申请、进行验收汇报。本标准中采集加工制作方为各级环境信息中心；
- 质量控制审核方，根据申请人提出的数据要求和使用要求对采集加工制作方提交的数据集进行审核，进行验收汇报。本标准中质量控制审核方为各级环境信息中心；
- 使用方等，申请使用已制作好的数据集的各级环保部门或人员，向数据集所有方提出使用申请。

各责任单位主要职责分配见表 1：

表 1 责任单位职责分配

	申请	审批 立项	管 理	采 集	制 作	汇 总	审 核	验 收	归 档	移 交	发 布 分 发
申请方	●							●			
数据源所有方					●			●			
数据集所有方		●						●			●
采集加工制作方			●	●	●	●		●	●	●	
质量控制审核方							●	●			
使用方	●										

数据源所有方加工数据流程参考图 1 环境数据集加工汇交流程。

由多个子集组成的环境数据集，环境数据集及子集都应配备相应的管理人员、制作人员、审核人员、验收人员和归档人员等，见图 2。

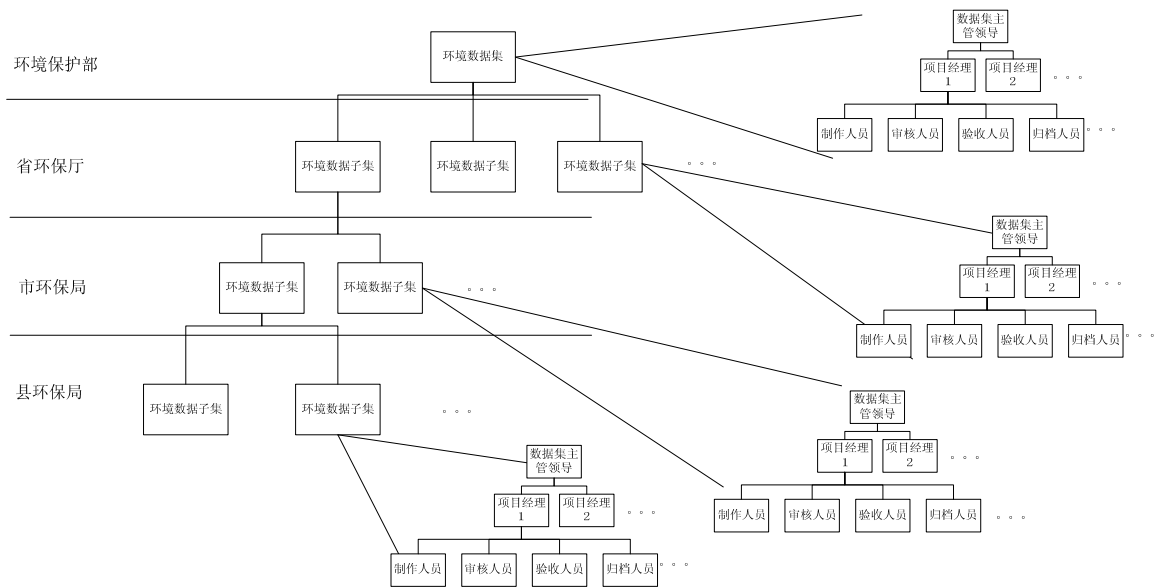


图 2 环境数据集多层制作机构示意图

环境数据集加工汇交流程按数据源来源不同分为环境数据集横向加工汇交和环境数据集纵向加工汇交。环境数据集横向加工汇交是在平级环境保护主管部门之间环境数据集汇交，环境数据集纵向加工汇交是不同级别环境保护主管部门之间环境数据集汇交。横向加工汇交和纵向加工汇交的数据源来源不同，但工作流程相同。

4.2 环境数据集加工汇交申请审批立项

当申请人需要的数据集涉及多个部门时，申请人向最高级别信息中心提出环境数据集加工汇交申请。审核通过后立项并任命项目经理。

例：申请全国范围内国家级、省级建设项目环评方面的数据，申请人应向环境保护部信息中心负责环境数据集加工汇交工作的主管领导提出环境数据集加工汇交申请。

4.3 下达环境数据集制作任务

环境数据集的制作应以申请人需求为出发点，在充分调研的基础上，确定数据集制作内容和要求。在调研的基础上，确定数据集制作、追加或更新的内容，填写《环境数据集制作任务书》，参见附录 A 表 A1.1，明确数据集制作（追加或更新）的目标、研究内容、质量控制要求、完成时间、需要提交的数据集及相关的技术文档等，并向数据集制作单位发《环境数据集制作任务书（下级单位）/函（同级单位）》，数据集制作单位应指定数据集制作人员。环境数据集制作任务接收单位根据数据来源在需要的情况下向数据提供部门发送《环境数据集制作任务书（下级单位）/函（同级单位）》，见图 2。

4.4 环境数据集实体采集制作

4.4.1 确定环境数据源

环境数据集制作人员调研制作数据集数据的来源、质量状况、完整性及相关技术文档

等，以保证数据集制作的可靠性。应按照以下步骤确定数据源并制作数据源加工要求：

- a) 数据源比对，针对要制作数据集的目标，如有多种数据来源，应对不同来源的数据进行比较分析，确定使用数据质量最好的数据源；
- b) 数据源校对，文件应经过输入、计算机校对或者人工校对两种方式之一，确定其是相应来源的数据文件；
- c) 数据源的检查，为了保证数据集中的数据完整、正确，应对数据源进行格式检查、质量检查和分析。对于时间跨度大的数据，尤其是早期计算机处理能力较低时录入的数据，存在一定数量的不规范和错误；对这类数据文件采用最新软件对其进行格式和质量检查处理；
- d) 应保留数据源质量控制码，且应将检验的方法及原则在数据集说明文档中说明；
- e) 确定数据源加工规则和提取规则，形成数据源加工要求任务书，作为采集加工的依据。

4.4.2 数据加工

数据源提供方根据加工要求任务书制作子数据集，子数据集的制作流程和管理要求与数据集相同。

4.4.3 数据采集

环境数据集制作人员根据数据源加工要求任务书进行数据采集。

4.4.4 数据处理

环境数据集制作人员按照《环境数据集制作任务书》的要求，选取合适的数据处理软件和编程语言，利用合理的数据处理方法，对采集的子数据集进行加工，形成数据集实体的数据文件。

- a) 数据处理方法及原则：
 - 1) 环境数据库数据处理方法及原则应严格按照国际或国内环境数据相关资料加工处理的相关规定或标准执行。如果没有相关的数据处理方法，则应在充分调研和分析国内外同类数据加工处理方法的基础上，结合数据源的基本特征，选择较为合适的处理方法和技术，保证数据集的质量。
 - 2) 追加的数据文件应与源数据集数据文件的内容、处理方法及存储格式等保持一致。
- b) 数据开发软件及编程语言，环境数据库应该采用通用的开发软件及编程语言。特殊数据或图形采用的开发软件、二次开发软件及开发环境（如数据库系统）应就软件运行的系统环境、安装和使用方法等在数据集说明文档中详细说明。
- c) 数据文件的存储格式，环境数据库数据文件格式应选用国内、国际通用的数据格式。如可选择 Word、Excel、Pdf 等文档格式、数据库文件（dbf、mdb 等）以及通用图像文件（GIF、JPEG 等）格式。数据文件中记录的顺序应按照一定规则排列。

4.4.5 质量检查

主要检查如下内容：

- a) 数据来源的可靠性：
 - 1) 数据集制作人员对数据集的数据来源应进行一定的质量分析，以保证数据集制作的可靠性。
 - 2) 对影响数据精度的观测仪器、手段及通讯等多种因素应做必要的说明；对于不同数据间的比较，要求采用统一标准。
 - 3) 针对要制作数据集的目标，如有多种数据来源，应对不同来源的数据进行比较分析，确定使用数据质量最好的数据源。
- b) 质量控制：
 - 1) 应对环境数据集进行质量控制。
 - 2) 如果对源数据已经具有一定标准质量控制或处理方法规定，则由此形成的数据集应严格按照规定的方法进行质量控制；如果源数据已经经过严格的质量控制，则保留其质量控制码，但对新加工形成的数据要进行必要的质量控制；如果源数据没有进行质量控制，则应对源数据进行必要的质量检验，检验的方法及原则应在数据集说明文档中说明。
 - 3) 环境数据集的质量控制一般使用环境科学或统计学等方法。经质量检验后的数据应标注质检码，质检码应采用国际或国内通用的标识和含义。
- c) 特征值的处理，环境数据集中出现的特征值应做出说明，尽量使用规范的特征码表示。
- d) 完整性检查，检查环境数据集中包含时间和空间信息的数据，保证数据的完整性。对不完整的数据应在数据集说明文档中注明缺失信息。

4.4.6 环境数据集组织及命名

- a) 环境数据集分类编码，依据环境数据所具有的属性如内容属性、时间属性、空间属性、来源属性、观测手段属性、格式属性、使用属性等，按照一定的原则和方法对环境数据集进行分类编码。
- b) 数据集组织，参照环境数据集分类编码及命名规范，对各类环境数据进行组合、加工，形成便于存档和使用的各种数据集。
- c) 数据集实体文件命名，环境数据集实体的数据文件名应该用英文或汉语拼音表示，文件名中不能包含下列任何半角字符：斜杠(\、/)、冒号(:)、星号(*)、问号(?)、引号(")、尖括号(<、>)和竖杠(|)。文件名长度不限。

4.5 环境数据集元数据制作

环境数据集元数据信息必须包含《环境信息元数据标准》规定的核心元数据，也可包含该数据集特需的元数据信息。若为污染源在线监控数据集，可参考《环境信息元数据规范应用方案 1 污染源在线监控元数据》；若为污染源监督性监测数据集，可参考《环境信息

元数据规范应用方案 2 污染源监督性监测元数据》；若为环境统计数据集，可参考《环境信息元数据规范应用方案 3 环境统计信息元数据》。

在制作环境数据集元数据时应采用 XML 语言对元数据进行描述，其 schema 应遵循《环境信息元数据标准》的附录 A。

中文元数据文件名规定为：“数据集代码”_META_C.DOC 或“数据集代码”_META_C.TXT；英文元数据文件名规定为：“数据集代码”_META_E.DOC 或“数据集代码”_META_E.TXT。

4.6 环境数据集说明文档制作

环境数据集说明文档制作需遵循【HJ□□-20□□】（《环境数据集说明文档格式标准》发布后，补充编号，去掉此括号。）

4.7 环境数据集建立

- a) 存储介质。存储介质可使用大型的盘阵、磁带机和光盘，对于其它类型高密度介质，经资料管理部门批准后，也可使用。
- b) 数据集存储目录结构。每个环境数据集在介质中要制作多级目录，其存储目录结构见图 3。

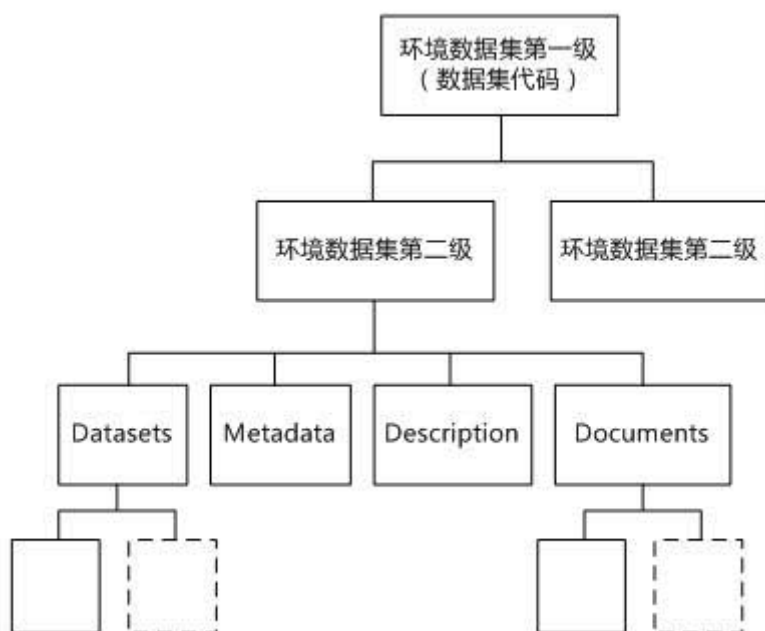


图 3 环境数据集存储目录结构

环境数据集第一级目录为环境数据集代码。

环境数据集第二级目录用以标识出相同“环境数据集代码”的不同数据的唯一性。使用“流水号”作为唯一性标识，由采集加工制作方统一分配。

环境数据集第三级要编制四个子目录：

第一个目录名为“datasets”，放置数据集实体文件。如果文件数目较多，根据需要可以分多级子目录，子目录名依据一定规则命名，若没有相应规定和要求可以自定义，目录名应使用英文、汉语拼音或数字等表示。

第二个目录名为“metadata”，存放环境数据集元数据文档。

第三个目录名为“description”，存放环境数据集说明文档。

第四个目录名为“documents”，存放数据集附加文档，在数据集的制作过程中形成的具有保存和使用参考价值的其它文档。包括数据集制作任务书、数据集说明文档的补充性文档、质量检查情况、数据集归档文档等材料。

每个环境数据集根目录下都要编制一个 readme.txt，说明数据集存放介质序号、各级目录下存放文件内容。格式如图 4（以光盘为例）：

- 1、光盘序号：光盘序号/总数
（介质序号书写格式为：介质序号/介质总数，如 2/3：表示数据集介质总数为 3，该介质在数据集中的序号是 2。）简述数据集的内容。
- 2、Datasets 目录：数据集实体文件数量。
（说明该目录/子目录下数据集实体文件数据量）
- 3、Metadata 目录：元数据文档。
- 4、Description 目录：数据集说明文档。
- 5、Documents 目录：数据集附加文件名称/数据集制作过程文件名称。

图 4 readme.txt 文件内容

c) 数据集在介质中的放置

根据环境数据集量的大小，每个数据集在介质（以光盘为例）中有两种存放方式，具体如下：

- 1) 一张光盘只存放一个环境数据集，则按数据集规定的存储方式放置。
 - 2) 如果环境数据集的容量较大，需分别放置在多张光盘上，则环境数据集的第一张光盘的目录结构应按上述规定方法放置，自第二张后，每张光盘的数据集名目录下可以只保留一个子目录“datasets”和一个 readme.txt，在 readme.txt 文件中只要说明“介质序号”和“datasets”目录或子目录下放置的数据集实体文件数量即可。
- d) 环境数据集索引。随着环境数据汇交的时间越长，积累的数据会越来越多，最终将达到海量的水平。对于这些海量级的数据，往往因为无法被用户有效检索定位而无法体现出其应有的价值。为此，需要建立多维索引，包括：时间索引、空间索引、内容索引、专题索引，记录到元数据中去。
- e) 环境数据集示例，在每个环境数据集中都附带示例，让使用者获取到数据集后可以知道怎么能比较好的使用数据集中的数据，同时也能通过示例从更直观的角度

去了解数据。放入“documents”目录当中。

- f) 建立“数据集生命周期管理文档”，通过 xml 文件格式，记录文件从采集/制作到归档，最终到数据集不能再被使用的全过程信息，其中包括：每个环节的办理时间，办理人以及该环节碰到的问题。放入“documents”目录当中。
- g) 环境数据集存储介质外标签内容。环境数据集存储介质外标签内容应能反映数据集制作基本信息，包括：环境数据集中文名称、环境数据集代码、环境数据集版本号、环境数据集介质序号、环境数据集时间范围、环境数据集制作时间和制作单位、环境数据集索引。

4.8 环境数据集审核

4.8.1 拟定环境数据集实体的审核方案

数据审核与制作人员应由不同的技术人员负责。

根据《环境数据集制作任务书》的要求，拟定环境数据集实体的审核方案，包括审核的技术路线和方法，审核方案包括并不限于 4.8.2 环境数据集审核的内容。经数据集项目经理确认后，按此方案对环境数据集实体进行审核。

4.8.2 环境数据集审核

环境数据集审核包括环境数据集实体、元数据和环境数据集说明文档的审核。

环境数据集实体的审核指标包括：数据实体的代表性、准确性、精密性、可比性、完整性。不满足以上条件的，需要退回制作部门修改。

以环境监测数据实体为例：

- a) 代表性是指具有代表性的时间、地点，并按规定的采样要求采集的有效样品的特性。
- b) 为了保证监测数据的完整性应审核不同监测对象的采样频次、采样时间、监测项目设置是否达到监测技术规范的要求，分析方法、检出限、测试精度、评价标准是否符合监测目的的要求。
- c) 数据的可比性（合理性）分析包含时空分布合理性分析、污染物排放规律合理性分析、监测指标之间相互关系合理性分析，以及采样、监测、数据处理等全过程可比性分析，还包含标准物质准确度、各行业之间、污染物之间、实验室之间监测数据的可比性和统一性。

环境数据集说明文档的审核应按照 4.6 节环境数据集说明文档制作要求进行审核。

元数据的审核按照 4.5 环境元数据制作要求进行审核。

环境数据集审核过程根据实际情况可重复多次，直到数据集的正确性和唯一性得到保证。在此过程中，审核人员和制作人员应注意随时在工作日志上详细记录数据集加工处理和审核过程中的有关技术上的特殊处理，为相关文档的整理做准备。

4.8.3 提交环境数据集审核报告

审核人员应根据审核结果，认真整理撰写环境数据集审核报告。报告应详细记录审核工作的全过程，包括：环境数据集审核的时间、人员、内容、方法、环境数据集正确性、环境数据集完整性、文档内容与数据的一致性、审核结果（即关于环境数据集质量的认定）及环境数据集审核人员与制作人员的交互过程等。

审核结束，应向环境数据集制作人员提交环境数据集审核报告。

审核报告文件的命名规定为：“数据集代码”_CHECK.文件扩展名，文件扩展名通常采用如.RTF 或.DOC 等。

4.9 环境数据集汇总

4.9.1 环境数据集实体、元数据文档和说明文档汇总

环境数据集制作人员负责环境数据集实体、元数据、环境数据集说明文档及其附加文件的汇总。

4.9.2 环境数据集制作过程文件汇总

a) 制作过程工作单

环境数据集制作过程工作单包含《环境数据集制作任务书》和《机读载体记录档案归档登记表》，由核心制作人员负责制作。其中《环境数据集制作任务书》由项目经理填写，是环境数据集制作、审核和验收汇交人员进行环境数据集制作、审核和验收汇交的重要依据，环境数据集完成时作为环境数据集制作过程文件应在环境数据集中保存。在环境数据集制作与归档过程中形成的《环境数据集产品验收单》和《机读载体记录档案归档登记表》也须保存，便于今后管理和使用。

《环境数据集制作任务书》命名规定为：“数据集代码”_TASK.扩展名；

《机读载体记录档案归档登记表》命名规定为：“数据集代码”_ARCHIVE.文件扩展名，文件扩展名通常采用如.RTF 或.DOC 等。

- b) 审核报告，环境数据集审核报告由审核人员提供，是对环境数据集质量状况的附加说明。
- c) 制作流程说明文件，环境数据集制作流程说明文件，包括数据集制作流程、源程序清单及其功能说明等。
- d) 此文件可根据要求和环境数据集制作的具体情况决定是否制作。
- e) 其它说明文件，环境数据集制作过程中制作人员认为尚需补充的其它重要信息的说明或解释。

4.9.3 制作流程和其他说明文件命名规定

文件的命名规定为：“数据集代码”_文件标识.文件扩展名。

文件标识用能反映文件基本内容的英文或汉语拼音标识；文件扩展名采用文件格式的通用标识方法，如.RTF、.DOC 等。

4.10 环境数据集验收

a) 提出验收申请

环境数据集审核通过并完成汇总工作后，由项目经理向负责人提出验收申请。

b) 验收申请审核

负责人根据验收申请材料决定是否进行验收，制定验收计划，包括验收时间、验收人员、验收内容。

c) 召开验收会

由采集加工制作方、数据源所有方、质量控制审核方做验收汇报，负责人、申请方及相关专家（如果需要）进行验收。

验收人员就数据集的正确性和完整性进行验收，验收中发现问题时应及时与制作人员联系，并责成制作人员改正。验收合格后，验收人员须在《环境数据集产品验收单》的验收人栏签字，参见附录 A 表 A2.1。验收合格的环境数据集，由归档人员进行归档。

《环境数据集产品验收单》命名规定为：“数据集代码”_ACCEPT.文件扩展名。

4.11 环境数据集归档

4.11.1 归档原则

环境数据集归档的目的是为了服务应用，归档数据集及相关文档应按规定统一管理。环境数据集归档应按照《电子文件归档与电子档案管理办法》执行。

4.11.2 归档方式和移交

a) 归档人员应在环境数据集验收后及时将所有数据集向总体组移交归档。归档应该包括全部数据集产品，保证数据集产品完整安全归入档案库房。

b) 为保证归档资料的完整、有效，档案管理人员应依据《环境数据集产品验收单》和《机读载体记录档案归档登记表》对存储介质内放置的各个目录的文件逐一检查，并按规定编写档案编号，以入库归档。

c) 归档数据集的移交，应完成《机读载体记录档案归档登记表》的填写，由移交双方签字，参见附录 A 表 A3.1。

4.12 环境数据集发布/分发

环境数据集发布由负责人进行审核。环境数据集发布需要按照《环境信息共享资源发布和使用技术规定》（在编）中相关要求进行。

附录A
 (规范性附录)
 环境数据集工作文件

A.1 环境数据集制作任务书

表 A.1 环境数据集制作任务书

任务名称			
承担单位		负责人	
数据集中文名称			
制作人员		完成时间	
数据集制作目标及要求			
需要提交的数据集及相关文档			

说明：

- 1、本登记表由下达数据集制作任务的项目经理填写。
- 2、表中各单元格的高度是固定的，不要自行调整，如果预留的位置不够，可以通过调整字体的大小和行距来解决。

A.2 环境数据集产品验收单

表 A.2 环境数据集产品验收单

数据集代码：_____

数 据 集 实 体	数据集中文名称			
	数据来源			
	时间范围：			经纬度或地域名：
	文件数：	数据总量：	介质及数量：	
	资料内容（要素/项目）：			
	文件格式：			
元 数 据 和 说 明 文 档	元数据名称			
	说明文档名称			
数 据 集 附 加 文 件 和 制 作 过 程 文 件	文件名称：			
数 据 集 制 作 责 任 人	制作人		制作时间	
	审核人		审核时间	
	审核结果：			
数 据 集 验 收 人	验收人		验收时间	
	验收意见：			

说明：

- 1、表中有关数据集的内容由数据集制作人员填写，其它各栏由相关人员填写。
- 2、表中各单元格的高度是固定的，不要自行调整，如果预留的位置不够，可以通过调整字体的大小和行距来解决。

A.3 机读载体记录档案归档登记表

表 A.3 机读载体记录档案归档登记表

移交单位：

经办人：

接收单位：

经办人：

档号或文书登记号	载体检验	病毒检验	真实性检验	完整性检验	有效性检验	技术方法及相关软件说明、资料检验

移交日期：