

HJ

中华人民共和国国家环境保护标准

HJ 721-2014

环境数据集加工汇交流程

Environmental dataset's processing and collection workflow

(发布稿)

本电子版为发布稿。请以中国环境科学出版社出版的正式标准文本为准。

2014-12-25 发布

2015-03-01 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.....	1
4.1 环境数据集加工汇交流程及职责.....	1
4.2 环境数据集加工汇交立项.....	2
4.3 环境数据集实体采集制作.....	2
4.4 环境数据集审核.....	5
4.5 环境数据集汇总.....	6
4.6 环境数据集归档、保管及备案.....	6
4.7 环境数据集加工汇交安全要求.....	6
附录 A（规范性附录）环境数据集工作文件.....	7
附录 B（资料性附录）环境数据集加工汇交示例.....	10

前 言

为贯彻《中华人民共和国环境保护法》，促进环境信息化工作，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环境数据集加工、汇交的业务流程、内容、方法和要求。

本标准附录 A 为规范性附录，附录 B 为资料性附录。

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 2014 年 12 月 25 日批准。

本标准自 2015 年 3 月 1 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环境数据集加工、汇交的流程、内容、方法和要求。本标准适用于指导各级环境业务部门和机构进行环境数据集制作、存储、归档等工作。

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。

GB 17859	计算机信息系统安全保护等级划分准则
GB/T18894-2002	电子文件归档与管理规范
HJ 720	环境信息元数据规范
HJ 722	环境数据集说明文档格式
HJ 729	环境信息系统安全技术规范
HJ/T 417-2007	环境信息分类与代码
BMB17-2006	涉及国家秘密的计算机信息系统分级保护技术要求
环办[2012]92号	附件 11 环境身份认证技术规定
环办[2012]92号	附件 12 环境数据加密技术规定
环办[2012]92号	附件 13 环境数据访问技术规定
环办[2012]92号	附件 15 环境信息安全测试与评估技术规定

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。

3.2 数据集实体 dataset entities

由一系列静态数据文件组成的数据或图形的集合。

3.3 环境数据集实体 environmental dataset entities

由一系列静态环境数据文件组成的环境数据或图形的集合。

3.4 数据集说明文档 dataset's discription document

描述数据集的一组说明文件，组成内容必须包括概述、数据集标识、数据内容和结构、覆盖范围和元数据。

3.5 数据集制作过程文件 dataset's processing and collection process document

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。

3.6 加工汇交 processing and collection

制作环境数据集全部过程的总称，由环境数据集采集、制作、审核、汇总、归档、保管及备案等流程组成。

4 技术内容

4.1 环境数据集加工汇交流程及职责

4.1.1 环境数据集加工汇交流程

环境数据集加工汇交流程如图 1 所示。

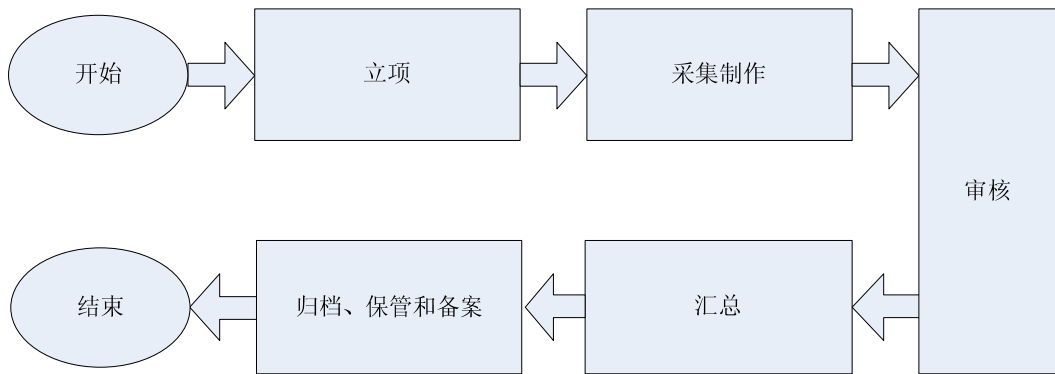


图 1 环境数据集加工汇交流程

环境数据集加工汇交详细工作流程见附录 A.1。

4.1.2 各方职责

4.1.2.1 环境数据集管理单位

负责立项、编制《环境数据集制作任务书》；审核环境数据集；备案数据集加工汇交的过程及结果。

4.1.2.2 采集加工制作单位

根据《环境数据集制作任务书》的要求落实数据来源，对数据来源的数据质量进行检查；负责制作环境数据集实体、元数据及说明文档；配合环境数据集管理单位完成数据审核及环境数据集归档。

4.1.2.3 支持单位

负责保管验收通过的环境数据集实体、元数据及说明文档，并开展环境数据集共享服务工作。

4.2 环境数据集加工汇交立项

4.2.1 环境数据集管理单位根据环境数据集共享工作要求以及使用需求确定环境数据集加工汇交任务，填写《环境数据集制作任务书》，见附录 A.2 中表 A.1。

4.2.2 环境数据集管理单位应明确环境数据集制作、追加或更新的目标，确定环境数据集的内容和完成时间，需要提交的环境数据集、相关的技术文档等。

4.2.3 《环境数据集制作任务书》命名规则为：环境数据集名称_TASK.扩展名。

4.3 环境数据集实体采集制作

4.3.1 确定数据源

为了保证环境数据集的完整性和正确性，采集加工制作单位应对数据源进行格式检查、质量检查和分析，并将检验方法、检验原则和检验结果在《环境数据集说明文档》中进行说明，格式应符合 HJ720 的要求；在有多种数据来源的情况下，应根据环境数据集制作需求选择合适的数据源，并详细记录数据源的选择理由。

4.3.2 数据采集

采集加工制作单位应根据《环境数据集制作任务书》进行数据采集。对数据采集情况进行检查，确认是否符合《环境数据集制作任务书》的要求，并对采集过程、结果及处理办法进行记录。

4.3.3 数据处理

4.3.3.1 数据处理工具和方法

采集加工制作单位应按照《环境数据集制作任务书》的要求，选取相应的工具和方法对数据源进行

加工，形成环境数据集实体数据文件；同时应在环境数据集说明文档中对数据处理工具和方法进行详细说明。

数据处理工具包括各类数据处理软件如 EXCEL，以及管理数据的文件系统和数据库系统如 ACCESS，数据处理方法包括数据查找、数据筛选、数据排序、分类汇总以及复杂的数据分析方法等。

4.3.3.2 数据文件的存储格式

数据文件的存储格式应选用国内、国际通用的数据格式，如文档文件格式、数据库文件格式以及通用图像文件格式等。

4.3.3.3 质量内控要求

- a) 质量管理：要求在元数据、环境数据集说明文档中对环境数据集质量控制过程和质量检查结果进行描述。
- b) 质量控制过程：包括自查、审核两个步骤。
- c) 质量检查结果描述：数据质量应当用质量定量元素和质量非定量元素两个组件来描述，质量定量元素包括完整性、可靠性、逻辑一致性、数据精度等；质量非定量元素包括目的、用途等。

4.3.4 环境数据集组织及命名

4.3.4.1 环境数据集分类编码

环境数据集分类编码应依据 HJ/T 417-2007 进行制订。

4.3.4.2 环境数据集实体文件命名

环境数据集实体文件命名：环境数据集名称“_”文件的补充说明部分。

- a) 每个环境数据集文件都隶属于某个环境数据集，文件名称的前半部分即为该数据集名称，后半部分为文件的补充说明部分，文件扩展名为文件类型。
- b) 文件补充说明部分用于补充说明该文件必须说明的一个或多个属性特征（如时间、空间位置、要素、格式等），用中横线“_”分割，名称不限长。

4.3.4.3 环境数据集组织

参照 HJ/T 417-2007，按不同分类信息的属性组成不同的环境数据集。

如环境质量数据可以按照监测及统计时段、监测区域、监测单位级别、监测对象组成不同的数据集，如：2012 中国环境状况公报、2012 年 XX 省环境状况公报、2013 年全国地表水水质月报（9 月份）等。

4.3.5 环境数据集元数据制作

4.3.5.1 环境数据集元数据信息必须包含 HJ720 规定的核心元数据信息，也可包含环境数据集特需的元数据信息。若为污染源自动监控数据集，可采用 HJ720 附录 F 的相关规定；若为污染源监督性监测数据集，可采用 HJ720 附录 G 的相关规定；若为环境统计数据集，可采用 HJ720 附录 H 的相关规定。

4.3.5.2 在制作环境数据集元数据时应采用 XML 和纯文本格式对元数据进行描述，XML 格式的 schema 应遵循 HJ720 的相关规定。

4.3.5.3 元数据文件中文名称命名规则为：环境数据集名称_META.XML 或环境数据集名称_META.TXT。

4.3.5.4 为保证元数据制作的规范性，应采用元数据制作工具完成元数据的制作和维护。

4.3.6 环境数据集说明文档制作

环境数据集说明文档编制应按照 HJ722 的有关规定执行。

4.3.7 环境数据集存储

4.3.7.1 环境数据集存储目录结构

环境数据集在介质中要制作四级目录，存储目录结构如图 2 所示。

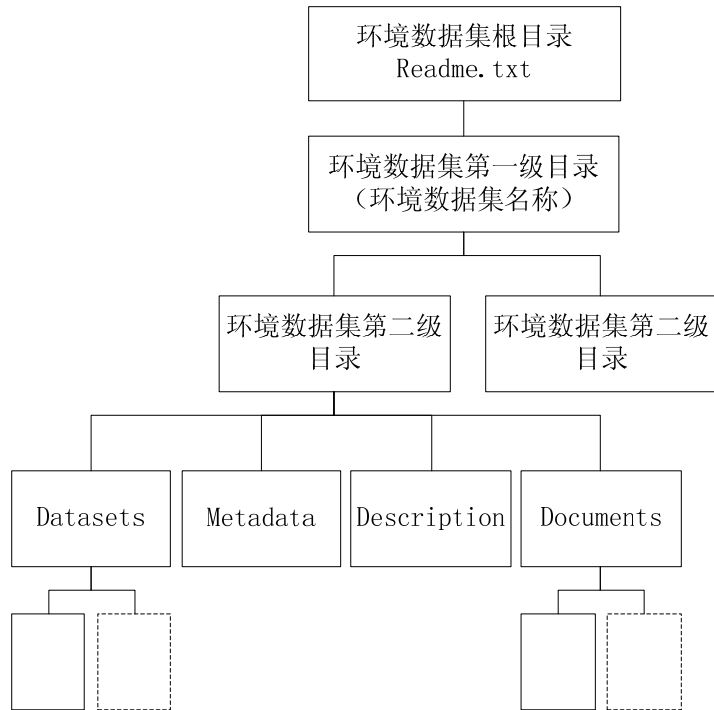


图 2 环境数据集存储目录结构

- a) 环境数据集根目录下应附加 readme.txt 文件，用来说明环境数据集存放介质序号、各级目录下存放文件内容，readme.txt 文件格式如下：
 - 1) 介质序号：介质序号/介质总数，并简述环境数据集的内容。介质序号书写格式为：介质序号/介质总数，如 2/3：表示环境数据集介质总数为 3，该介质在环境数据集中的序号是 2；
 - 2) Datasets 目录：环境数据集实体文件数量，并简述该目录（及子目录）下每个文件的内容；
 - 3) Metadata 目录：环境数据集元数据文档数量，并简述该目录（及子目录）下每个文件的内容；
 - 4) Description 目录：环境数据集说明文档数量，并简述该目录（及子目录）下每个文件的内容；
 - 5) Documents 目录：环境数据集附加文件档和环境数据集制作过程文件全部数量，并简述该目录（及子目录）下每个文件的内容。
- b) 环境数据集第一级目录为环境数据集名称；
- c) 第二级目录用以标识出相同“环境数据集名称”的不同数据的唯一性，使用“流水号”作为唯一性标识，由采集加工制作单位统一分配；
- d) 第三级目录为四个子目录，其中：第一个子目录的名称为“datasets”，该子目录用于放置环境数据集实体文件；
 - 1) 第二个子目录的名称为“metadata”，该子目录用于存放环境数据集元数据文档；
 - 2) 第三个子目录的名称为“description”，该子目录用于存放环境数据集说明文档；
 - 3) 第四个子目录的名称为“documents”，该子目录用于存放环境数据集附加文档及在环境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，包括环境数据集制作任务书、环境数据集说明文档的补充性文档、质量检查情况、环境数据集归档文档等材料。
- e) 第四级目录如果文件数目较多可以分多级子目录，命名规则可自定义，目录名应使用英文或数字等表示。

4.3.7.2 环境数据集在介质中的放置

如果环境数据集的数据量超过单个存储介质的容量，应分散在多个存储介质上存放。

示例：以光盘作为存储介质，具体存放方式如下。

- a) 当一张光盘可存放一个环境数据集时，全部环境数据集文件应统一按 4.3.7.1 的规定存放；
- b) 如果 datasets 目录容量较大，应分别放置在多张光盘上，则第一张光盘的目录结构应符合 4.3.7.1 的规定，自第二张起，每张光盘只保留子目录“datasets”和 readme.txt 文件，在 readme.txt 文件中应说明“介质序号”和“datasets”目录或子目录下放置的环境数据集实体文件数量。

4.3.7.3 环境数据集索引

为快速检索到所需的环境数据集，应建立多维索引，包括时间索引、内容索引、专题索引，并记录到元数据中去。

4.3.7.4 环境数据集示例

环境数据集应附带示例，用来帮助使用者使用环境数据集中的数据。示例应放入“documents”目录当中，参见附录 B。

4.3.7.5 建立“环境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，最终到环境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、以及该环节遇到的问题。环境数据集生命周期管理文档应放入“documents”目录当中。

4.3.7.6 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环境数据集制作的基本信息，包括：环境数据集名称、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、环境数据集索引等。

4.4 环境数据集审核

4.4.1 拟定审核方案

采集加工制作单位应根据《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。审核对象包括环境数据集实体、元数据和环境数据集说明文档。审核的内容包括以下方面：

- a) 按照《环境数据集制作任务书》的要求，是否按时进行加工汇交，数据的内容是否齐全；
- b) 必须包含的核心元数据信息填写是否完备；
- c) 加工汇交的环境数据集实体和元数据是否为电子版；若数据汇交形式为数据存储路径，路径是否真实、可访问；
- d) 数据是否为涉密数据，是否符合相关法律、法规的要求；
- e) 其他需要审核的内容。

4.4.2 环境数据集审核

- a) 环境数据集管理单位应根据审核方案对环境数据集进行审核。审核不通过的需退回采集加工制作单位修改；
- b) 环境数据集的审核可根据实际情况重复多次，直到环境数据集质量合格。在此过程中，采集加工制作单位应详细记录工作日志，为下一步环境数据集汇总工作做好准备。

4.4.3 环境数据集审核报告

- a) 环境数据集审核通过后，采集加工制作单位应撰写环境数据集审核报告；
- b) 环境数据集审核报告应详细记录审核工作过程，具体内容包括：环境数据集审核的时间、人员、内容、方法、环境数据集的正确性和完整性、文档内容与数据的一致性、审核结论等；

- c) 环境数据集审核报告文件应采用 .RTF 或 .DOC 的存储格式,命名规则为:环境数据集名称_CHECK.扩展名。

4.5 环境数据集汇总

采集加工制作单位负责对环境数据集实体、元数据文档、环境数据集说明文档和数据集制作过程文件进行汇总。

4.6 环境数据集归档、保管及备案

4.6.1 环境数据集归档

- a) 环境数据集的归档流程及档案命名规则应按照 GB/T18894-2002 的有关规定执行;
- b) 采集加工制作单位负责填写《机读载体记录档案归档登记表》,命名规则为:环境数据集名称_ARCHIVE.扩展名。按照附录 A 表 A.3《机读载体记录档案归档登记表》的格式要求;
- c) 为保证归档资料的完整性和有效性,环境数据集管理单位的档案管理人员应依据《机读载体记录档案归档登记表》对存储介质内的文件逐一检查,并按有关规定编写档案编号,办理归档。
- d) 归档工作完成后,交接双方需在《机读载体记录档案归档登记表》签字。

4.6.2 归档后,采集加工制作单位应将环境数据集实体、元数据文档、环境数据集说明文档交由支持单位统一集中保管和共享发布,并由支持单位负责保障共享数据安全。

4.6.3 环境数据集管理单位负责对每次数据集加工汇交的过程及结果进行备案,备案内容包括采集加工制作单位、加工汇交时间、加工汇交数据内容、加工汇交数据量、审核结果及其他需要说明的内容。

4.7 环境数据集加工汇交安全要求

在环境数据集加工汇交工作过程中应充分考虑数据安全问题。安全防护的等级评定应符合 GB17859 的规定,具体选用的安全技术应按照 HJ729 的相关要求,数据访问安全设计应符合环办[2012]92 号附件 11 和附件 13 的规定,数据加密设计应遵循环办[2012]92 号附件 12 的细则,并按照环办[2012]92 号附件 15 的要求对环境数据集加工汇交流程进行安全评测。

涉及国家秘密信息时,应符合 BMB17-2006 的相关规定对环境数据集加工汇交工作过程进行管理。

在环境数据集加工汇交过程中,应做好数据备份工作,避免由于存储介质故障导致数据丢失。

附录 A
 (规范性附录)
 环境数据集工作文件

A.1 环境数据集加工汇交详细工作流程

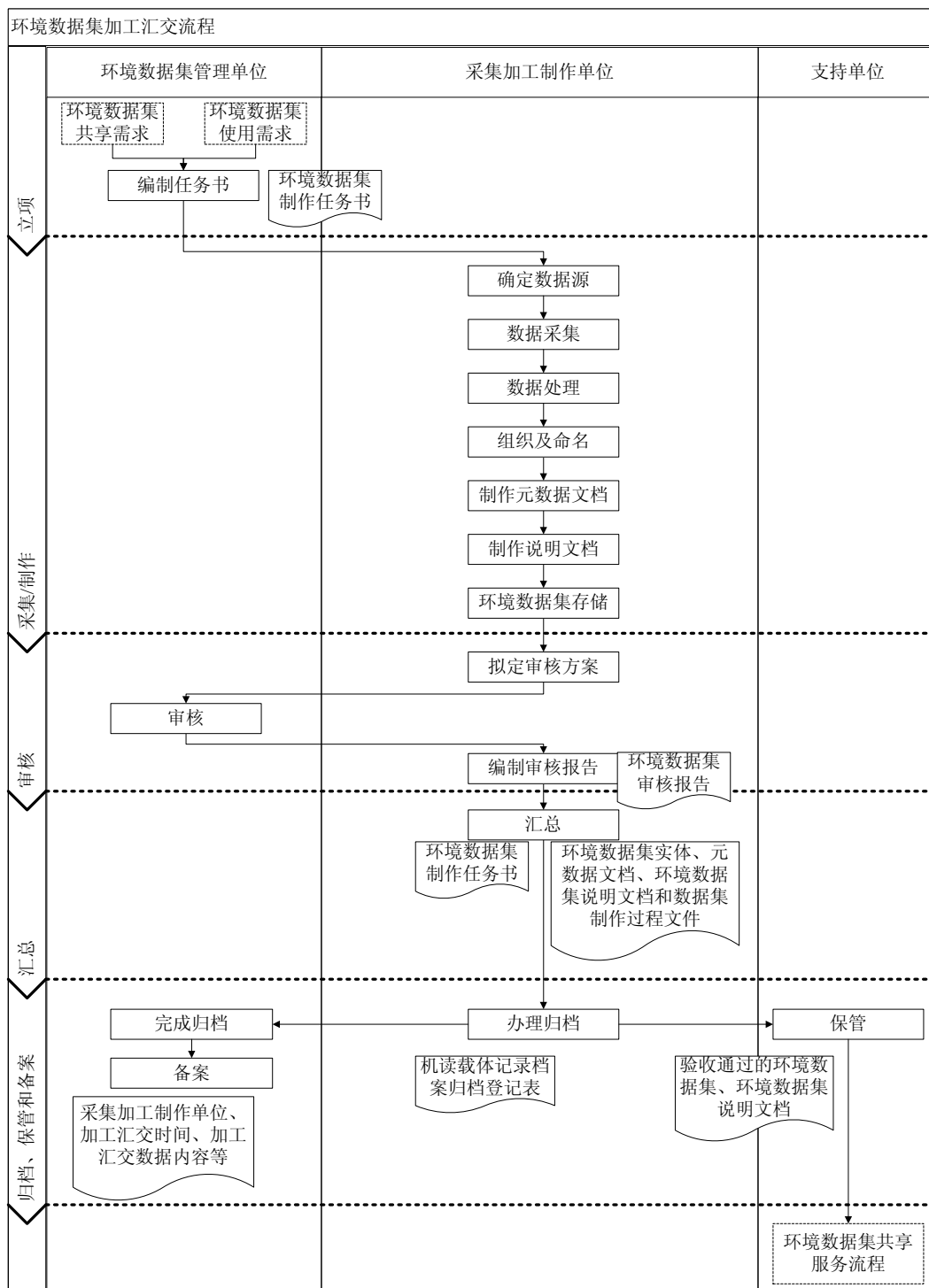


图 A.1 环境数据集加工汇交详细工作流程

A.2 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			
环境数据集管理单位		负责人	
环境数据集中文名称			
采集加工制作单位		规定完成时间	
环境数据集制作目标及要求			
需要提交的环境数据集及相关文档			

说明：本任务书由环境数据集管理单位填写。

A. 3 机读载体记录档案归档登记表

表 A. 2 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	病毒检测	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:

附录 B
(资料性附录)
环境数据集加工汇交示例

B.1 概述

根据第 4 章定义的环境数据集加工汇交流程, 给出《XX 省 9 个城市空气质量月报》数据集加工汇交流程涉及的单位职责及工作流程。

B.2 单位职责

《XX 省 9 个城市空气质量月报》数据集的管理单位为 XX 省环境监测中心站, 采集加工制作单位为 XX 省环境监测中心站大气室(简称大气室), 支持单位为 XX 省环境信息中心。

B.3 加工汇交流程

a) 立项

《XX 省 9 个城市空气质量月报》是 XX 省环境监测中心站确定的环境数据集共享工作内容之一, XX 省环境监测中心站编写了《XX 省 9 个城市空气质量月报数据集制作任务书》, 下发大气室负责制作。

b) 确定环境数据源

大气室确定《XX 省 9 个城市空气质量月报》数据来源为《XX 省城市空气质量日报》。

c) 数据采集

大气室根据《XX 省 9 个城市空气质量月报数据集制作任务书》进行数据采集, 采集完成后, 对数据采集情况进行检查, 确认是否符合《XX 省 9 个城市空气质量月报数据集制作任务书》的要求, 并对采集过程、结果及采集办法进行记录。

d) 数据处理

大气室按照《XX 省 9 个城市空气质量月报数据集制作任务书》的要求, 选取相应的工具和方法对指定月份的《XX 省城市空气质量日报》进行加工, 形成环境数据集实体数据文件; 在环境数据集说明文档中对数据处理工具和方法进行详细说明; 在元数据、环境数据集说明文档中对环境数据集质量控制过程和质量检查结果进行描述; 按照自查、审核两个步骤进行质量控制, 记录质量检查结果。数据文件的存储格式为文档文件格式。

e) 环境数据集组织及命名

数据集命名: XXXX 年 XX 月 XX 省 9 个城市空气质量状况月报

数据集分类:

类目名称: 环境质量报告

类目编码: 0103

分类标准: 环境信息分类与代码

f) 环境数据集元数据制作

按照 HJ720 进行元数据制作。

g) 环境数据集说明文档制作

按照 HJ722 进行说明文档制作。

h) 数据数据集存储

《XX 省 9 个城市空气质量月报》数据集存放在 XX 省环境监测中心站统一的磁盘阵列中。

i) 环境数据集审核

《XX 省 9 个城市空气质量月报》制作完成后，由 XX 省环境监测中心站对数据集进行审核。

j) 环境数据集汇总

审核通过后，大气室负责对环境数据集实体、元数据文档、环境数据集说明文档和数据集制作过程文件进行汇总。

k) 环境数据集归档、保管及备案

《XX 省 9 个城市空气质量月报》归档流程及档案命名规则按照 GB/T18894-2002《电子文件归档与管理规范》的有关规定执行。

归档后，大气室将环境数据集实体、元数据文档、环境数据集说明文档交由 XX 省环境信息中心统一集中保管和共享发布，并由 XX 省环境信息中心负责保障数据共享安全。

XX 省环境监测中心站负责对每次数据集加工汇交的过程及结果进行备案，备案内容包括采集加工制作单位、加工汇交时间、加工汇交数据内容、加工汇交数据量、审核结果及其他需要说明的问题。

1) 安全要求

《XX 省 9 个城市空气质量月报》为非涉密数据集，主要做好数据防篡改安全防护工作，并定期进行磁带备份。