

气候数据拯救最佳做法指导原则

2016年版

水
气候
天气



世界气象组织

WMO-No. 1182

气候数据拯救最佳做法指导原则

2016年版



世界气象组织

WMO-No. 1182

编辑说明

《气象术语》是WMO术语数据库，可登录以下网址查询：<https://public.wmo.int/zh-hans/meteoterm>。

请通过选择文本中的超级链接进行复制的读者注意，紧跟：<http://>、<https://>、<ftp://>、<mailto:>后以及斜杠 (/)、破折号 (-)、句号 (。)和连贯字符序列（字母和数字）后会有额外空格。这些空格应从所粘贴的URL中删除。当光标悬停在链接上或点击链接时会显示正确的URL，然后再将其从浏览器复制。

WMO-No. 1182

© 世界气象组织, 2016

WMO对印刷、电子和任何其他格式的出版物，以及用各种语言出版的出版物拥有版权。短幅选摘WMO出版物无须授权，但须清晰完整地注明出处。涉及编辑及要求出版、重印或翻译本出版物全文或部分者，须联系：

Chair, Publications Board

World Meteorological Organization (WMO)

7 bis, avenue de la Paix

P.O. Box 2300

CH-1211 Geneva 2, Switzerland

电话: +41 (0) 22 730 84 03

传真: +41 (0) 22 730 81 17

电邮: publications@wmo.int

ISBN 978-92-63-51182-7

注：

WMO出版物中所用的称号和本出版物中的材料表示方式并不代表WMO秘书处对各国、领土、城市或地区、或其当局的法律地位、或对其边界划分的观点立场。

WMO出版物中的观点是作者的观点并不代表WMO。提及的具体商号或产品与未予提及或未刊登广告的同类相比并不表示前者得到了WMO的赞许或推荐。

目录

	页码
范畴.....	v
鸣谢.....	v
1. 引言 – 气候数据拯救的科学重要性.....	1
2. 气候数据拯救概述：存档、图像化和数字化.....	1
3. 存档纸质介质.....	2
3.1 搜寻和定位.....	2
3.2 保护和保存.....	3
3.3 创建纸质 / 微缩胶卷档案和图像的电子清单.....	4
4. 原始介质图像化.....	5
4.1 图像化过程的基本步骤.....	6
4.2 图像文件的验证和存储.....	6
5. 数字化数据值.....	7
5.1 将键入错误最小化.....	8
5.2 与国际数据库的比较.....	8
6. 存档数字介质（图像和数据值）.....	9
6.1 交叉检查印刷介质、图像和数字数据集.....	9
6.2 更新介质.....	9
6.3 数字介质存档建议.....	9
附录 1 全球活动.....	11
附录 2 数据拯救评估.....	12
附录 3 基础设施、设备、用品和人员.....	13
附录 4 纸质档案维护指南.....	14
附录 5 确定数据图像化和数字化的优先考虑事项.....	15
附录 6 电子成像技术.....	17

	页码
附录 7 其它数字化方法	20
附录 8 数字化的错误和质量保证过程	22
附录 9 经验教训.....	23
附录 10 显示数据价值的应用	25
附录 11 气候数据拯救核对清单	26
附录 12 术语表	28
附录 13 参考文献	30

范畴

本技术文件是对编号为 WMO/TD-1210、WCDMP-55 的《气候数据拯救指导原则》(2004) 的更新。它以最初的指导原则为基础,同时考虑到了 12 年来发生的技术变化以及世界各地从最近开展的气候数据拯救活动中汲取的经验教训。对数据拯救的概述包括其重要性、原始介质存档、图像化、数字化以及数字图像和数字数据的存档等章节。十二个附录提供了支持性信息。

本《气候数据拯救指导原则》旨在以推荐最佳做法的形式提供指导。由于国家气象水文部门 (NMHS) 技术开发的规模和阶段各不相同以及天气类型和气候的多变性,有些做法或许并不适用于每个 WMO 会员。话虽如此,但本指导原则涵盖了各类指导,可就如何组织和实施数据拯救提供援助,并为每个会员提供通用的技术解决方案。在[国际数据拯救 \(I-DARE\) 门户网站](#)上可查询到更具体的技术信息以及丰富的插图和照片,该网站由 WMO 负责维护,同时得到了荷兰皇家气象局以及 WMO 气候学委员会数据拯救专家组的协助。

虽然这些最佳做法特定于天气和气候数据,但也可用于 WMO 职权范围内外其它科学领域的数据拯救。尤其针对水文、海洋及其它环境数据的拯救可遵循类似的总体原则和做法,且基本上属于本指导原则的范畴。不过,此类数据的特殊性需要通过经与各界密切合作来确定和考虑,例如与包括 WMO 水文委员会和 WMO 政府间海洋委员会 (UNESCO) 的海洋与海洋气象联合技术委员会合作。

鸣谢

我们要感谢美国伊利诺伊大学中西部地区气候中心的 Nancy Westcott 女士 (现已退休),感谢她对本出版物的杰出贡献。

气候数据拯救最佳做法指导原则

1. 引言 – 气候数据拯救的科学重要性

气候数据拯救包括整理和保护所记录的仪器观测资料以及面临丢失风险的气候数据。数据拯救对于确保未来几代科学工作者及其他数据用户获取所有必要信息至关重要，以便评估气候变率和变化以及提供各类气候服务。这些数据有助于弥合古气候数据和当前观测资料之间的差距。全球许多气候数据集都包含上至 1940 年代和 1950 年代的数字化数据，但 19 世纪和 20 世纪初的许多观测资料目前仅是硬拷贝形式，需要对其加以拯救和数字化。世界有些地区的数据完全可追溯至 1900 年代，但还有许多地区没有这些数据。对气候模式和古气候数据（例如年轮、冰芯、花粉）观测的验证会极大受益于世界所有地区的长期仪器时间序列。

21 世纪仍然存在着数据稀疏地区。因此，重要的是获取和验证最新的记录。观测站位置和选址以及观测的质量控制 (QC) 会显著影响气象观测资料的可靠性。台站数据与附近地点的数据比较，以及查阅原始观测文件（和观测做法方面的原始信息）的能力将使气候工作者能够更好地验证过去和目前天气和气候状况。

回溯更长时间的气候记录并填补时间和空间差距将产生诸多效益：

- (a) 它有助于使农业气象、疾病向量化以及水文和气候数值模式表示法更加可信，并能够更好地预估未来的气候；
- (b) 拯救的数据与现有数据相结合，可更好地从历史角度了解目前的天气和气候；
- (c) 它可为评估自然和人为系统对气候变率和变化的历史敏感性提供依据，从而可更好地准确评估未来气候变率和变化的影响。

此类评估可为决策者减轻天气灾害造成的损失提供依据，而且可为经济发展提供更多信息。

为了切实方便气候服务提供方和研究人员获取数据，数据必须以数字形式（包括图像和数值）提供。此外，所有现有数据也必须以数字形式保存和提供，因为这些观测资料是未来的历史数据。原始文件对于验证数字化数值至关重要。

2. 气候数据拯救概述：存档、图像化和数字化

气候数据拯救包括整理和保护那些随时间逐步退化、损毁、疏忽、技术陈旧或简单分散而面临丢失风险的气候数据资产。非数字化的数据正面临着风险，因为原始纸质记录易损。数据拯救包括：纸质记录、微缩胶卷和微缩胶片记录的整理和图像化；键入数字和文字数据并将条形图数据数字化处理为可用格式；并将数据、元数据以及质量控制结果和程序存档。关于气候数据拯救活动各部分的概述参见表 1。

表 1 气候数据拯救活动的组成部分

数据拯救的组成部分	活动	关键词
存档纸质和微缩胶卷 / 微缩胶片介质 (第 3 节)	搜寻和定位	NMHS、观测点、大学、航空和海事机构、农业机构、国际图书馆和数据库、国家级档案馆
	保护和保存	清洁介质；置于无酸和有标签的档案盒中、避免灰尘和潮湿及害虫
	创建电子纸质 / 微缩胶卷档案清单	编目所有纸质介质；估算图像化和数字化工作范畴
图像介质 (第 4 节)	创建总图像清单	
	成像和验证 更新总图像清单 在每张 CD/DVD 或在每台计算机目录中创建图像文件清单	在验证图像可读后更新总图像清单，包括元数据在内 CD/DVD 文件清单与总图像清单交叉检查
数字化数据值 (第 5 节)	创建数字数据清单	
	录入条目、图表跟踪	数据输入到 CDMS
	质量控制数据	随着数据的数字化并通过各种质量控制测试，更新数字数据清单
存档数字介质 (第 6 节)	交叉检查印刷介质、图像和数字数据	将图像和数字数据清单与原始电子纸质 / 微缩胶卷档案清单进行比较
	备份电子介质	每天
	分散图像和数字数据档案的多个副本	至不同位置
	更新介质和转移技术	每 5 到 10 年

大部分国家都认真进行了天气观测，手动或自动记录结果，抄录到纸质表单或制作缩微胶片拷贝，并最终将其输入某种形式的电脑媒介（最好是输入到气候数据管理系统 (CDMS)），以便于获取和分析。遗憾的是，这些数据有时并未从纸质表单中抄录出来，有的是因为缺乏资金和 / 或人员不足，有的是因为数据保存在其它地方，有时是在海外。由于介质快速老化或是因为部分档案未记录迁移情况，许多纸质表单和微缩胶卷 / 微缩胶片拷贝面临着丢失的风险。

本指导原则旨在就整理、图像化、数字化和保护此类气候数据所需步骤提供建议。

3. 存档纸质介质

3.1 搜寻和定位

数据拯救的首项任务是定位数据。在开始保护程序之前，至关重要确定历史数据集并将其识别为可用气候数据。这些数据也许保存在国家档案馆或服务机构、观测站中或由私人收藏，不一定为气象界所知。有些数据甚至可能保存在海外档案馆，尤其是有殖民史的国

家。若不进行此类搜寻，大量或许有价值的数据集可能会遭遇封尘、无人问津。除了已在本组织或 NMHS 现有数据管理系统中经过整理、数字化和良好管理的数据之外，在以下地点和来源搜寻或许会发现已被遗忘的宝贵气候数据：

- (a) NMHS 及其它部门或机构通常保有数据档案；
- (b) 大学、学院、中学和个人（例如气象及相关组织的长期供职人员）；
- (c) 国际数据中心；
- (d) 图书馆或地方历史学家可能也有纸质、微缩胶卷 / 微缩胶片乃至数字化的数据拷贝；
- (e) 有数百年气象记录、在全球各地经营着种植园的大型食品企业等农业组织；
- (f) 军事机构可能也有可获取的记录，运输部门也可能有航空和海事天气记录；
- (g) 宗教组织，尤其是有观测或科学兴趣的宗教组织，例如教职修会和牧师；
- (h) 科学学会，特别是过去的组织（自然历史学会）；
- (i) 博物馆；
- (j) 船舶日志、历史报章和个人日记均是天气和气候信息的宝贵来源。

搜寻过程应交叉检查可用的数字记录，以确定记录是否已数字化、确定显著差距并为特定地区开展观测计划的日期提供线索。这包括查看国际数据库，例如：

- (a) 全球历史气候网络；
- (b) 国际地面气压数据库；
- (c) 综合地面数据库。

此外，国家图书馆藏有宝贵的数据集（这些图书馆的链接参见 [I-DARE 门户网站](#)）。

3.2 保护和保存

纸质和微缩胶卷 / 微缩胶片档案应按逻辑合理地进行整理，并保存在无酸档案盒中，放置在坚固的架子上或文件柜中：

- (a) 根据先前存放纸质图表或表格的位置和条件，纸质本可能必须在保存前除尘或吸尘；
- (b) 在处理可能被污染的记录时，需要遵循正确的工作、健康和安全程序；
- (c) 给易破损记录除尘需要专门的设备和培训；

- (d) 如果记录已发霉或经防霉剂或杀虫剂处理过，则需要格外小心，确保操作人员的安全；
- (e) 应佩戴手套处理易破损文件；
- (f) 根据灰尘量，或需要穿戴防护服和口罩。

专门的房间通常是为了保存归档的纸质和微缩胶卷数据：

- (a) 房间最好有温度和湿度控制；
- (b) 必须保护数据免遭虫咬、鼠咬、霉菌、火灾、洪水、灰尘、盗窃及所有其它危险；
- (c) 避免房间放置木质物品，以尽量减少火灾和虫害的危险；
- (d) 专业档案员可帮助指导这一过程，并应指定一位管理员定期检查房间。

数据通常按照类型（例如表格或图表类型）以及按照台站、年份和月份保存在无酸档案盒或文件柜中：

- (a) 数据的年数和台站的数量可确定是按年份还是按台站最合理地整理纸质本；
- (b) 为档案盒和 / 或文件柜贴上标签，并在纸质 / 缩微胶片档案的电子清单中注明保存位置（档案盒或文件抽屉的编号），以及台站 / 年份 / 月份、介质类型和表格类型，以便可轻松找到数据用于可能的图像化和数字化。

对于长期保存，应考虑以下事项：

- (a) 重要数据图像化之后，原始文件不可丢弃。如果资源状况不允许继续保存，则请联络国家档案馆、机构 / 部门的图书馆或大学图书馆加以保存。如果这些机构无法保存档案，且在征得国家档案机构的允许后，可联系国际图书馆或机构保存（进一步论述见附录 4）；
- (b) 微缩胶卷不适合于永久存档。微缩胶卷会老化，除非放置在有除湿机的空调房间；
- (c) 数字图像也不适合于永久存档（见第 6.2 节）。

附录 3 将进一步述及所需的基础设施、用品和人员。附录 4 将给出维护纸质档案的建议。

3.3 创建纸质 / 微缩胶卷档案和图像的电子清单

随着对纸质介质进行整理和保存，需要制定一系列清单以促进从纸质本转为电子图像再到数字数据。这一点十分必要，以便在发现和整理纸质数据后，可对它们进行图像化处理。在做成图像后，便可对其进行数字化，而完成数字化后，则可便捷地检索原始纸质数据和电子图

表 2 电子清单的制定和维护步骤

搜集要保存的所有记录和元数据	
根据逻辑存档计划整理记录	按台站 / 表格类型 / 年份 按表格类型 / 台站 / 年份 按年份 / 台站 如果是图表数据, 则还按气象要素 (风、降水等)
创建纸质 / 微缩胶卷档案的电子清单	包括台站、年份和档案盒 / 文件抽屉位置、介质类型、表格类型 包括变量类型 定义缺失的时期 估算介质类型、表格类型和图像量的数量, 以确定工作计划
从电子档案清单创建总图像化清单	确定台站、年份、月份、介质类型、表格类型或数量 定义记录时期 估算图像化的页数
在创建图像时, 为便于图像文件定位在各 CD/DVD 上或在每台计算机目录中创建图像文件清单	
包括仅含元数据的页面 (网络信息、观测做法等)	
在进行图像化过程中维护主成像清单	在验证图像时输入做成图像的页码 输入 CD/DVD 位置或计算机位置
对主成像清单与图像文件清单交叉检查 CD 或其它存储介质上的图像数量	

像, 用于对数字化值的质量控制验证。理想状态下, 每一步的命名方式都应尽可能多地包含信息。表 2 归纳了建立原始介质 (纸质和微缩胶卷 / 微缩胶片) 档案以及后续图像化、图像的计数和图像文件位置等所需步骤。

总之, 发现气候数据后, 应将其纳入纸质 / 微缩胶卷档案的电子清单, 来确定有哪些数据、已做出了哪些拯救努力以及还要开展的工作。电子清单档案包括台站、年份、月份、介质类型、表格编号或类型、要素类型、页数 (包括元数据页数), 尤其是档案盒 / 文件抽屉位置和用于表示文件的唯一标识符。变量类型也应被纳入。

创建了国家历史气候记录的电子清单后, 应加以维护, 接收到新数据后要进行更新, 进行图像化时, 新图像的位置应添加到总图像清单。总图像清单可用于估算图像化工作的进展。图像化工作的更多详情可参阅第 4 章。

4. 原始介质图像化

收集纸质和微缩胶卷副本并对其编目之后, 可开始规划, 以确定图像化和数字化的顺序。从何处开始以及如何开展这些工作均取决于各 NMHS 的需求。数字化的优先顺序有可能与图像化的有所不同。在对图像化和数字化排定优先顺序时要考虑的事项参见附录 5。

对纸质和 / 或微缩胶卷记录进行拍照或扫描, 创建图像或数字图片, 而后可保存以防进一步老化并可广泛提供。原始气象文件的所有记录都应通过数码相机或光学扫描仪转化成图像。这是为了确保拯救所有现有的元数据 (例如观测做法、地点和数据描述)。关于数码相机和扫描仪的相对优点以及理想的相机功能参见附录 6。

在所有情况下, 元数据 (对数据和观测做法、地点和站点特征的描述) 都必须与图像和数字值一并拯救并数字化。

4.1 图像化过程的基本步骤

图像化的做法和技术可能各不相同, 但均需文档记录和文件跟踪。强烈推荐的一些基本程序包括:

- (a) 如表 2 所列, 根据纸质 / 微缩胶卷资料电子清单创建总图像清单。包括待图像化的页数。在成像和验证成像时, 更新总成像清单, 添加上已生成图像和验证的页数, 并添加计算机目录或 CD/DVD 位置 (提供模板);
- (b) 在图像化过程中验证图像文件的可读性, 确保该图像可读且图像内容与文件名相符;
- (c) 数码相机或扫描仪将创建大量图像文件。每次生成一个图像并下载至计算机, 相机 / 扫描仪软件将为每个图像生成一个文件名。这些文件名可能与文件内容无关。如果同一天图像数次下载至计算机, 有些文件名或许会出现重复。建议使用软件生成与图像内容有关的文件名 (见下段 (d) 中示例)。有些相机和扫描仪有软件执行此项工作, 抑或可单独购买。见 [I-DARE 门户网站](#) 和附录 6;
- (d) 图像文件名应包括台站标识号 (SID)、表格类型的缩写 (ACR)、日期 (YYYYmmDD) 和页码 (PPP)。它们通常为 .png、.jpg 或 .tif 文件, 例如: SID_ACR_YYYYmmDD_PPP.png;
- (e) 以台站 / 年 / 月目录结构存储图像。维护主目录中的文件名主列表和各个目录或 CD/DVD 中的类似文件列表, 以便可以定位和检索图像;
- (f) 更新总成像清单, 添加图像的页数, 并将原计数的页数与实际制作并存储于 CD/DVD 或硬盘上的图像面数进行比较。计算和录入清单的页数应等于制作的图像数。录入图像的位置 (CD/DVD 或目录名);
- (g) 图像化过程的文档包括文件命名规范都至关重要, 因为从事图像化的人员可能会随着时间发生变动。

4.2 图像文件的验证和存储

先查验和核对图像文件, 而后再写入 CD/DVD 或硬盘或 CDMS:

- (a) 图像可能因相机位置失焦;

- (b) 图像可能会颜色太浅，可能看不到钢笔 / 铅笔所做的修正；
- (c) 因为相机 / 扫描仪的设置，图像可能太暗；
- (d) 强光可能会遮蔽部分图像资料。使用相机闪光灯或相机支架照明可将光线反射回相机镜头，从而遮蔽一些图像化的数据，介质为光面纸时尤其如此。

质量保证 (QA) 的有效方法是定期 (例如每 300 页) 检查一些图像，确保其可读。还要适当检查一些图像页面，将其与纸质档案比较，确保没有缺失任何一页数据。重要的是确认图像的文件名和图像内容至少每月 / 每年 / 每个台站匹配一次。

如果使用 CD/DVD，则它们必须贴上标签。

- (a) 使用时必须将图像文件的文件名以及各类识别元数据 (台站名和编号、年份、月份、表格类型、页数) 和 CD/DVD 标签输入文件清单；
- (b) 与各 CD/DVD 有关的文件副本应被包括在该 CD/DVD 以及总目录中；
- (c) CD 标签应被包括在总图像清单中。

至关重要是在 CD/DVD 和外部硬盘和 / 或 CDMS 中创建图像档案副本。如果将 CD 复制到硬盘或 CDMS 中，则每个 CD/DVD 应有一个目录，且还必须复制所有文件清单。建议使用 DVD 而不是 CD，因为 DVD 的存储容量更大。

良好数据管理的一个基本要求是异地存储数据和元数据备份，以防自然灾害或人为灾害损害建筑物和计算机而使数据损失。

- (a) 重要的是分发整个电子图像档案和总图像清单的副本；
- (b) 如果在区域或跨国中心处理数据，数据副本除了保存在区域中心外，还应分发给数据原产国。计划在区域中心处理数据时，各国之间必须事先达成协议，规定如果自然灾害或人为灾害损坏建筑物或计算机时应采取的措施。WMO 秘书处可协助达成此类协议。

5. 数字化数据值

自动记录带状图的映绘和文字数据转录为数字格式存储为计算机可读的数字称之为数字化。前者可参阅附录 7；本节讲述的是以键盘录入字母数字记录进行数字化。数字化通常在 NMHS 进行，但有时也由公司或通过公民科学倡议在区域中心进行，并且通常由气候学家等主题专家指导。如果书面数据是存放在国家档案馆，可能就需要在档案馆用键盘录入方式进行数字化。如果计划在 NMHS 以外甚至国外处理数据，各方之间必须事先达成协议。这也可在 WMO 秘书处的协助下达成。

以键盘录入方式进行数字化通常比图像化更耗时，20 个小时的数字化换算成图像化可能需要 1 个小时。由于数字化是一个这么长的过程，因此，有限的资源可能无法将所有数据进行

数字化处理。数据的数字化首先应当是基于科学、技术和社会经济考虑因素。如果无法以数字格式记录所有信息，则尤其重要的是保留原格式和原件副本。关于数字化优先顺序的进一步讨论见附录 5。

在某些情况下，如果没有资金录入大量的数据，可采用众包方法。志愿者将数据键入变为电子格式或将带状图的时间序列图映绘到专门设计的网站上。另一种方法是利用光学字符识别 (OCR) 软件来数字化印刷材料。该方法正在进行测试，可能将来会变得更为有效，并得到广泛使用。而现在 OCR 的价值有限，因为它需要专门格式才可读取。关于众包、带状图和 OCR 的进一步论述，可参见附录 7 和 [I-DARE 门户网站](#)。

数字化的第一步是设计数字数据清单，类似于为图像化所创建的清单，用以跟踪数字化过程的进展。数据应按台站、年份、月份和数据类型进行整理。

5.1 将键入错误最小化

有许多种技术能够将数字化误差降至最低。首先是两人或三人录入数据（两个或两个以上人员录入相同的数据）。这种做法是让多个录入员录入完全相同的数据，然后比较每个数据值。或者，录入员录入每天的数据，其他录入员录入每月数据，然后评估每月总值和平均值的一致性。两人录入可在 CDMS 之外进行，并在对比后导入数据。这样可将数据直接录入到 CDMS。

减少录入误差的第二种技术是创建录入表格模板。使用模板可使录入员确切知道数据在页面上的位置，从而减少误差数量。录入到模板的数据可直接导入数据库中。即使仅录入一部分表格，也建议为每个格式类型创建一个录入模板。

第三，重要的是完全按照印制格式转录数据（即 1 对 1 转录）。这在澳大利亚称之为“key as you see”，在法国称之为“saisir ce que l’ on voit”。避免所有形式的“即时”编码或单位转换。计算机在重新编码、变换单位、求平均乃至再重新编码方面都非常高效。按印制版录入可确保明确记录用于编码数据的公式。如果必须使用编码，则要注意它是否记录完整且可回推（即可返回原始数据）。数据必须按原始观测表格准确录入，即使观测员明显犯有错误。这些数据有时用于法律案件，而重要的是应维护原始数据。当然，也必须对错误值进行标记和修正。

双重 / 三重键入、创建模板以及 1 对 1 转录都可极大地减少失误。在录入数据或导入 CDMS 作为临时数据后，应进行质量控制（见附录 8）。这一点怎么强调也不过分。

最后，建议在给定页面上录入所有数据，主要是因为不可能再有机会返回到该页面并录入任何遗漏的数据（例如纬度、经度、观测员姓名、观测时间、其它气象变量以及所有数字和文字数据）。对所有变量数字化也有助于更好地对记录进行质量控制。

5.2 与国际数据库的比较

收集在国际数据库中存储的从 SYNOP 或 METAR 编码中获取的数字气候数据可能对 NMHS 具有吸引力。通常，这些数字数据在生成时并不在内部存储，而且由于下载海量数据文件所需的带宽问题，可能不易于获取。这些数据是可取的，因为已经数字化，且可能包含比存

储在内部气候数据库的数据有更多的时次和变量。不过，在许多情况下，它们可能会受到录入、编码和传输误差的困扰。它们还可能受到质量保证程序的影响。然而，如果原始纸质表格上的值不清晰，则可能要从 SYNOP 代码中恢复。原始介质的数字化以及与国际数据库的数据的比较将会带来更完整和更准确的气候数据。尽量使用国际数据库中的数据、新近数字化的数据和图像，以确保更多的数据来源。

6. 存档数字介质 (图像和数据值)

6.1 交叉检查印刷介质、图像和数字数据集

工作进行过程中，最好记录项目的进展。做法是将创建的图像和数字数据的月数与从已成像和待成像的页数中预期的数量以及每个站的记录期进行交叉检查。在数据拯救过程的初期，纸质 / 微缩胶卷档案的电子清单由原始纸质 / 微缩胶卷数据制作而成。工作结束时，将创建的数字数据的数量 (图像和数字化清单中所见) 与最初的电子档案清单相比较，以确定是否有数据缺失。

6.2 更新介质

对于数字数据而言，数据拯救的重要行与纸质数据的相同，因为其所存储的介质并不是永久的。例如磁带会随时间的推移而失去磁性，尤其是在温暖或潮湿条件下。纸张经过几百年后会破碎，而磁带介质及其它机读介质在几十年后会变得无法读取。磁带、光盘和 8 毫米盒式磁带通常已过时。如果数字数据仍存储在这些介质上，则应联系国家档案馆或大学，将这些介质上的数据移至当前的介质上，例如 DVD 或硬盘。否则，其结果可能是一串无法理解的数字和字符乱码。

随着计算机技术的发展，计算机、计算机操作系统、计算机语言和用于读取老介质的软件也已过时。这种过时问题将继续存在。我们强烈建议在未来将图像化和数字化的气候数据移至下一代新的但既有的介质 (例如 CD、硬盘) 上。目前，CD/DVD 被广泛使用，但在不久的将来可能不会再用。此外，建议保留多个备份，因为计算机 (以及 CD 和硬盘) 有时会发生灾难性的故障。

必须强调的是，用于记录、存储、读取和判读气候数据的软件必须从老计算机和计算机技术中迁移出来并加以更新。如果经常这样做，则在出现问题和数据丢失之前就可以拯救数据。迁移介质的决定或许应每 2-5 年评审一次，而更新计算机和计算机软件的决定或许应每 5-7 年评审一次。

最后，纯 ASCII 文本数据文件是首选格式，它便于读取数字数据，且可用于从一台计算机迁移至另一台计算机。如果使用数据库，则数据必须易于导出 ASCII 格式 (如 .csv 文件)。同样重要的是，为了迁移，存储文件的数字编码应有记录，而且要直观 (即无需解码)。

6.3 数字介质存档建议

- (a) 保护数字图像和数字数据文件。图像文件可视为官方记录，因此应予以保护；

- (b) 整理数据便于未来存取 (以便该数据可通过 CD 或通过目录结构获取: 台站、年份、月份)。在每张 CD 及总目录中保留一份文件清单;
 - (c) 确保每天对 CDMS 进行备份;
 - (d) 在适当介质上存储数字图像和数据, 以便在未来可继续使用数据, 并在需要时迁移数据集, 减轻过时和老化的风险;
 - (e) 异地存储数字介质以保护电子图像和数字数据。为了保护数据以防人为或自然灾害, 应在多个地点存储所有这些数据的副本, 包括存放在邻近地区之外的多国“区域气候中心”。
-

附录 1 全球活动

1. 谁应开展数据拯救?

拥有数据(纸质、微缩胶卷或数字)的任何团体或个人均应努力促进气候数据拯救。国家气候记录管理机构应当有数据拯救的特殊职责,因为它们更珍视待拯救的数据,而且知道哪些数据最重要。这些工作人员通常供职于 NMHS 的气候部门,作为气候数据管理者,数据拯救是其监管职责的一部分。然而,在许多公共机构(农业部门)和私营机构(大学、种植园、农业企业)中都有数据拯救的拥趸者。此外,为了促进数据拯救,已建立了志愿者组织,例如国际环境数据拯救组织(IEDRO)和国际地球大气环流重建计划(ACRE)。其它有意协助数据拯救活动的各方包括在职和退休的气候学者、图书管理员、历史学者、学生及配偶。这些人员在从事众包业务方面也很有价值。

2. 应在何处开展数据拯救活动?

有多种方法可选择数据拯救的地点。一个是在国内选择一个地方,存储待拯救的数据。这可以在 NMHS 总部和 / 或不同的省级办事处。最佳方法或许是在一个地方收集待拯救的数据,或如果每个地区均掌握或能够开发专业技术,则可在省级办事处开展图像化和保存工作。另一个可能性是在区域、跨国气候中心开展工作,例如在 WMO 区域气候中心。优势是在这些办事处可比个别国家更有实力搜集专业技术和设备。另一种方法是让第三方在该国或地区之外进行图像化或数字化,而所有文件归还来源国。最佳选项因各自情况而定。在将数据转至第三方之前,WMO 秘书处愿协助规划数据拯救活动的地点。

3. 国际数据拯救(I-DARE)门户网站

[I-DARE 门户网站](#)是从事数据保护、拯救和数字化的人员的网络资源。它为提供全球过去及当前数据拯救项目现状、需拯救的数据以及相关方法和技术等信息提供了单一入口。它是用于交换数据拯救各方面信息的网关,包括既有和新兴的拯救技术。此类信息同样适用于其它科学学科。

由于 [I-DARE 门户网站](#)旨在提升现有数据拯救活动的知名度、激励开展新活动并更好地协调国际数据拯救工作,因此, [I-DARE 门户网站](#)是一个实用的沟通工具。网站所列的每个项目可使具有类似兴趣的其它国家了解各国所做的出色工作。[I-DARE 门户网站](#)侧重于为社区提供有关过往和现有数据拯救活动的信息、还将协助确定差距和机遇、帮助最需要数据拯救的地区确定数据拯救的优先排序、并帮助项目筹资。

附录 2 数据拯救评估

表 3 为参与新的气候数据拯救计划的数据拯救评估人员提供了通用指南。

表 3. 数据拯救评估核对清单

1. 在 I-DARE 门户网站 查找与数据拯救相关的技术资源和最佳做法
2. 从国际数据库、图书馆、文献中下载所有已知的数据清单来评估国家已知的气象数据资产
3. 确定各国数据拯救的现状
4. 访问并评估现有的档案室
5. 与主管和员工交谈，了解数据拯救对 NMHS 至关重要的原因及其数据拯救的优先重点
6. 必要时，就制定原始来源、数据和元数据清单提供建议
7. 提供关于纸质数据存储、空间和供给需求及工作、健康和安全等考虑因素的建议
8. 提供图像化所需的方法和设备方面的建议
9. 评估与数据管理有关的 CDMS 能力及基础设施（年限、备份、内容）
10. 评审目前的数字化和 QA 过程来了解新数字化的数据如何适应工作流和 QC 过程或目前的数字化和 QC 过程是否需要修订
11. 提供数字化所需的方法和设备方面的建议
12. 提供图像化和数字化清单方面的建议
13. 提供元数据收集、存档和图像化方面 (WMO/TD-1186) 的建议
14. 提供键盘输入数据方面的建议
15. 提供关于拯救的数据可能用于向捐赠方 /NMHS 显示价值的建议
16. 提供关于图像化和数字化步骤所需人员的建议
17. 根据任务量和人员数量估算数据拯救每个部分的时间表
18. 根据 NMHS 的优先重点制定数据拯救计划
19. 将数据拯救任务信息添加到 I-DARE 门户网站

附录 3 基础设施、设备、用品和人员

在所有情况下，存储、图像化和数字化都应在安全、光线充足、无尘和气候受控的环境中进行（气候受控的最低要求包括空调和除湿机）。设备和用品采购根据各国的数据拯救需求和选用的方法而有所不同。表 4 可用于着手规划气候数据拯救项目的细节。

选用合适人员，这对于数据拯救项目的成功至关重要。气候学家最适合负责搜寻气候数据并决定整理档案。图像化程序最好是由关注保护的人建立，或许是懂摄影的人。细心周到、关注细节的人最适合操作成像仪和数字化仪，因为速度和准确性对于这些工作十分重要。有些人可能会交替进行图像化和键盘录入，以减少重复工作的枯燥感，并消除对个人的依赖性。如果工作人员熟悉计算机技术，而且两人一组进行，可更高效地进行图像化和数字化：一个人负责生成图像，另一个人负责检查图像质量，两人输入或跟踪同一页的数字化（双重录入）。我们强烈建议图像化和数字化工作在半天或四小时内完成，因为这两项工作均需要全神贯注。

所需的相机数量和翻拍支架的类型取决于数据是以装订还是未装订书籍提供。如果一次拍摄一张照片，建议使用两台相机，其中一台仅用作备份。翻拍支架用于放置相机和灯光。然而，要拍摄合订本中的数据，一个相机支架使用两台相机。这个系统使成像仪能够一次拍摄两页而不损坏装订。附录 6 和 [I-DARE 门户网站](#) 提供了关于相机支架的信息。

表 4 规划数据拯救任务的考虑要点

	整理	图像化	数字化
基础设施	光线充足的房间，配有气候受控和安全设备	房间有足够的电源插座和照明设备，以及足够的桌子用于分类纸质表格	房间有足够的电源插座和照明设备
设备	无酸档案盒 双格距书架（2 米宽，2.5 米高） 梯凳	两个或两个以上相机（功能和配件见附录 6） 翻拍支架 PC 书托架 桌子 技术手册	CDMS
用品	手套、工作服、口罩、防尘服	CD、DVD 和 / 或两个最大的外接硬盘 USB 电缆或 WIFI	
数据拯救人员	两人清理文件并整理档案及编制清单 进行工作、健康和安全管理培训，以防有污染的记录	两人拍照并进行质量检查、存储和更新清单	两人数字化和更新清单
后续工作人员	技术熟练人员更新清单并监测室内情况	气候工作者监督新数据和元数据的摄取	气候工作者负责质量控制及开发产品

附录 4 纸质档案维护指南

纸质档案应保存在专门的储存室，最好不是在地下室或阁楼。储存室应保持清洁，没有尘土、污垢、霉菌和虫害。应使用坚固的架子和无酸档案盒，不要把文件放在地板上。房间应光线充足，档案盒应贴有标签并便于取放。应有专人定期检查档案室的情况。

保存所有纸质文件至少直至 (a) 完成图像化；(b) 查看过所有图像的质量和完整性；(c) 复制完图像。重要的是制作多个清单和整个图像档案的电子副本，并将它们存放在几个不同的地点。

最好是一直保存纸质副本：

- (a) 纸质副本可能会有电子副本无法读取的测量或仪器误差标注（铅笔或钢笔斜线）；
- (b) 带修正标识和注释的原始副本对于数值的验证十分重要；
- (c) 计算机故障会导致电子副本丢失；
- (d) 因此纸质副本有比数字和电子介质更长的寿命。

将纸质档案保存为备份是谨慎的做法。

如果没有空间限制，纸质档案应保留在 NMHS。如果资源不允许继续保存纸质档案，则应联系国家档案馆或大学档案馆和图书馆。国家档案馆可制定数据记录（纸质和电子）保存时限的法律约束标准。如果这些国家机构对此无兴趣，则应联系国际档案馆。

附录 5 确定数据图像化和数字化的优先考虑事项

虽然立即对所有数据进行图像化和录入可能是最理想的，但现实情况是，存档过程需要时间，而且这一过程必须在某处开始。可以从填补现有时间序列的空白开始。根据国际数据拯救会议（2001 年 9 月 11-13 日，日内瓦），拯救历史和当前气候数据的最高优先重点应是：

- (a) 高质量数据；
- (b) 在国家、区域或全球范围至关重要且面临丢失风险的数据；
- (c) 可填补既有数据集空白的数据。

1. 图像化

在纸质文件图像化时，合乎逻辑的做法是从最老、最稀少或最易碎的文件开始，一次一个台站。然而，由于近年来纸张和墨水质量下降，可能需要尽快对较新的文件进行图像化。在确定图表图像化的优先顺序时，可从对 NMHS 最重要的图表开始（例如降雨、气压、风或日照图表）。工作簿（例如每月记录册）中的每一页都应当图像化，确保拯救所有现有的元数据（包括站点和数据描述及照片、仪器信息、校准、观测做法和质量控制程序）。

2. 数字化

鉴于数据数字化十分耗时，因此，优先顺序十分重要。在地面观测资料数字化时应考虑如下因素：

- (a) 要向潜在资助者展示概念证明，可能最好是选择有最长连续记录的台站或着重于显现特定热带气候特征（如干旱或洪水）的异常年份作为开始；
- (b) 首先要谨慎地录入已对某些数据数字化且仍在运行或有最长记录期的台站；
- (c) 全球范围内，20 世纪 50 年代之前的数据都很少。然而，在个别国家，1990 年之前的任何数据都可能是至关重要的；
- (d) 地点是另一个考虑因素，即选择可代表一个国家内不同气候区域的台站或具有社会经济意义的地点会比较有益；
- (e) 数据可用性可能是另一个考虑因素。要首先选择有最可靠数据和数据丢失最少的台站；
- (f) 然而，邻近台站 – 即使持续时间不长 – 对于证实测量结果十分重要；
- (g) 关于当前或历史数据，数据的时间分辨率（每月、每天、数小时）或对特定天气参数的需求可以确定数据数字化的优先顺序；

- (h) 如果需要温度、降水或风的日变化气候，则图表数据或每小时（数小时）观测十分重要；
- (i) 要检测基本气候变量，选择可换算为月值的气象变量（如温度和降水）（Bojinski 等，2014）；
- (j) 对于数据有空间空白的地区，可选择个别台站；
- (k) 对于雷暴、雹、龙卷、尘暴、雾和冰暴气候，则需要非仪器目视天气观测。

通常会考虑将地面观测资料进行图像化和数字化，不过高空探测资料、地面和高空图以及海洋观测资料同样至关重要。

- (a) 个别探测资料可提供关于温度、气压和湿度随高度变化的重要信息；
- (b) 鉴于个别探测数据可能会丢失，此类信息可能只通过高空图反演；
- (c) 地面图也十分重要，因为在绘制地面站数据后，原始形式有时会丢失。

这些是关于设定数据拯救优先顺序的一般指导原则和考虑因素。不过，每个国家或中心必须根据国家的主要需求以及气候数据记录状况，决定其优先顺序。在选定某个台站和年份后，最好检验图像化和数字化的程序。这将有助于确定人员配备需求以及如何简化过程。

关于表格数字化的最后一点提示：最好是录入给定页面上的所有数据，包括元数据（因为可能不会再有机会返回该页面并录入任何遗漏的数据）。最佳做法是创建一个模板用于各种表格类型的数字化。总之，元数据、数据和观测做法的描述（观测员姓名、观测时间）、地点（纬度、经度）以及站点特点必须与图像或数字值一并拯救和数字化。

附录 6 电子成像技术

特别需要创建电子成像，因为它可减少纸质或微缩胶卷介质的处理。它还可以对图像调整大小和增强，以使用键入数据时能够看清字迹。通常，JPEG 图像比较适合。某些来源显示 1600x1200 像素的图像就足够。

1. 扫描仪

通常可考虑使用扫描仪对纸质文件进行图像化。两种普通类型分别为平板扫描仪（常用于书籍）和馈电扫描仪（常用于单张纸页，例如图表）。扫描仪是相机的热门替代品，因为它们可减少图像化所需的设备量。不过扫描也有一些缺点：

- (a) 平板扫描仪会损坏书籍装订，或者有时会使最接近装订的数据失焦或扫描不到。这可通过设置相机加以避免；
- (b) 微缩胶卷和微缩胶片需要专用扫描仪；
- (c) 扫描仪有许多活动部件，通常需要维护和许可证；
- (d) 使用更复杂的扫描机器通常需要专业培训。由于这一培训问题，人员变动使一些项目的扫描工作中止。此外，培训不到位和许可证问题会妨碍使用扫描仪进行图像化；
- (e) 除了极贵的系统，扫描仪通常速度慢且笨重。

2. 相机

相机最常用于创建数字图像。它们适用于装订的和活页的纸质副本以及超大的文件，如地图。数码相机很常见，即使是自动相机也可提供可靠的图像，尤其是在光源稳定充足的情况下。使用固定式相机支架和照明系统、拍照遥控器和交流 (AC) 适配器，除翻页之外，几乎不用做任何事情，使得图像化过程快速和高效。此外，目前有便宜的软件用来重新命名数字图像，并将图像移至计算机。图像存档也十分快捷。

3. 相机的功能

表 5 归纳了设立照片成像工作站所需的设备。首先，必须提供一张终端桌和一台计算机以及相机装置（包括一个三脚架或相机支架）。足够数量的地面电源插座至关重要。虽然自然光或阳光对于图像化最理想 – 但请注意，光会被漫射而且要避免阴影 – 不一定总有自然光线的房间。因此，相机装置通常包括两个或四个灯。三脚架或相机支架与灯一并放置，使相机能

表 5 相机固定装置的基本要求

1	桌子
2.	电源插座
3.	相机支架 / 三脚架
4.	灯
5.	相机及配件
6.	计算机

够保持静止，以便可快速拍摄多张照片。重要的是频繁测试闪光灯，因为一些纸质介质是光面纸，会将光线反射回相机镜头，使成像数据的某些部分模糊不清。遥控功能可使摄影师离开相机拍摄照片，还可减少相机机身震动。保存数字图像的计算机十分重要。

相机电池续航时间通常相对较短。使用 AC 适配器可拍摄无数张照片，并可节省购买和更换电池的大量时间及费用。如果现有的相机无法使用 AC 适配器，则推荐备用的充电电池及充电器。重要的是，如果没有 AC 适配器而且必须使用电池，则相机的设计应当能够在更换电池时不用从相机支架上拆下相机。大部分数码相机都有塑料线将相机固定于相机支架。几十次拆卸相机更换电池后，这些线会被磨断。

图像从相机自动传输至 PC 需要有软件和 USB 连接线或无线通信 (WIFI) 接入。如果此类软件能够使鼠标接通相机快门，则或许不需要遥控线。如果没有 USB 端口或 WIFI，则应使用可获取的最大容量存储卡。数码相机通常包括能够为每张图像生成文件名并将照片直接传输至计算机的软件。这种软件可节省摄像师和档案员的大量时间。根据现有的相机，变焦镜头和更改相机设置的能力十分有用。表 6 归纳出可实现高效流程的理想相机配件。

考虑图像的大小，尤其是在使用存储卡时。根据图像的质量，可能并不需要最大的图像，中等大小的图像可能就足够了。人们发现，即使是 400 万像素的相机，也足以生成有足够分辨率的 JPEG 文件，从而产生可读的字母数字参数值。目前，就连手机相机都有约 1200 万像素。必须牢记，虽然字母数字数据恢复无需过高的分辨率，但使用自动计算机数字化程序时则或许需要。通常，照片都是彩色的，不是黑白的。彩色可使原笔记本上的订正及其它批注更易于取读。

表 6 促进高效流程的相机配件

1.	翻拍支架 - 可固定一个相机和 2-4 个灯
2.	为相机供电的 AC 适配器
3.	拍摄照片的遥控器或第 4 项中所列的相机软件
4.	拍摄照片并用 USB 线或 WIFI 直接将照片传输至计算机的相机软件
5.	或者，有最大容量的存储卡
6.	图像文件重命名的能力和软件
7.	现成的变焦镜头和用于特写的微距镜头
8.	改变相机设置的能力 (通过相机或计算机)
9.	防眩光玻璃用于轻压装订好的记录

表 7 确保照片质量的常见相机功能

1. 选择图像大小的能力
2. 取消闪光灯的能力
3. 微距特写模式
4. 调整感光度的能力
5. 调整白平衡的能力

其它要考虑的常见相机功能包括取消闪光灯的能力、微距特写模式、调整感光度的能力以及调整白平衡的能力。表 7 归纳了确保照片质量的常见相机功能。

附录 7 其它数字化方法

下文概要讨论了除键盘录入之外的数字化方法。所有方法均需要质量控制，因为任何转录都可能出错。

1. 光学字符识别、智能字符识别和智能字词识别

光学字符识别是一款常用软件，可用于打印文本的数字化，但无法处理手写观测资料。智能字符识别 (ICR) 针对的是打印的手写单字，并适用于不同的字体和打印风格。ICR 最适合读取结构性表单，准确率达 97%。智能字词识别可识别整个手写单词，无论是印刷体抑或草书体。然而，历史表单通常不符合 ICR/OCR 的要求，并且不适合使用当前的 ICR/OCR 技术。后两种方法使用机器学习技术更新手写笔体自动识别数据库。

在撰写本文时，这些技术非常昂贵，且如同所有数字化过程一样，需要质量控制以确保识别的准确性。这些技术仍处于开发阶段，随着时间的推移，有可能变得更加可靠。目前，最常用的数字化方式是录入，因为相信人脑能够最准确识别手写材料。

2. 众包

通过众包，信息由常被称为“公民科学工作者”的网上志愿者处理。一些既有的和成功的气候相关计划均采用了众包，尤其是 OldWeather 和雨霰雪社区协作网 (CoCoRAHS) 计划。OldWeather 项目的志愿者将船舶日志内容 (主要来自 20 世纪初) 录入到在线表格。每艘船的日志由三个人录入，并将其结果为形成最终产品而进行对比 (OldWeather 的链接参见 [I-DARE 门户网站](#))。CoCoRAHS 志愿者测量 10 厘米 (4 英寸) 雨量器的降水量，通常在其住处附近，并将这些每日降水量录入在线表格。这些将被放置到数据库并进行绘图。

虽然众包是利用志愿者来录入数据，但也需要成本。网上数据摄取基础设施的开发和维护、数据库开发、质量控制、为录入员 / 观测员编制培训指导材料以及吸引潜在志愿者的宣传活动等都需要时间、人力资源和融资渠道。与所有其它技术一样，众包也需要质量控制。

3. 图表数字化

为了数字化图表，观测员通常要将图表中的值转录成表格形式 - 这是一项繁琐的工作，尤其是当图表出现快速波动时，例如风速计或单个图表上的多条线。虽然能够从图表中读取精细的时间分辨率，但通常仅报告峰值或每日值。然而，利用图表数字化软件，个人可跟踪一条线，从图表中标示点位，并将这些点换算为数字。这可使数据值的时间分辨率从数分钟到数小时。在 20 世纪 90 年代，个体团体开发了一些将降水 (雨量计) 数据数字化的程序，但它们只能用于某些计算机和操作系统。该软件的源代码仍可用，但没有继续维护 (例如针对最新编程语言和计算机技术进行更新)，且必须由软件技术人员进行调整。此外，需要对软件的维护和使用人员进行培训。

近期，一些新的数字化软件程序已经开发并投入使用。这些程序的功能包括能够计算不同时间尺度(数小时至数分钟)的降雨量；读取各类图表类型；以及处理单个图表上的多条线。该软件的文档可参阅 [I-DARE 门户网站](#)。其中有些程序可免费使用，并规定将数据放在可供所有人使用的免费开放的数据库中。

附录 8 数字化的错误和质量保证过程

下文简要概述了对尤其与数据录入有关的数字化错误的质量控制。这些质量控制功能的部分或全部或许可在会员国的 CDMS 上获取。若是如此，当数据录入 CDMS 时，错误会被标记出来，并随后进行修正。如果 CDMS 上没有这些功能，则强烈建议对其加以考虑。新录入的数字数据的初始质量控制集中在录入、仪器、观测员和元数据错误，而不是在均一化过程中进行的更精细的分析，有时称之为“滞后模式”质量控制。录入的数据在经过质量控制检测验证之前应被视为（并标记为）临时数据。

采用双重录入、模板和 1 对 1 录入可将各个值的录入错误降至最低，但录入仍会出错。元数据问题（主要与台站命名和识别编号有关）、扫描的图像质量、仪器故障以及非标准观测做法等带来的问题仍是我们面临的挑战。为了确保最精确地以数字表示气象表格中的数据，应使用质量控制检验来标记元数据（例如台站名称或标识号、年份或月份、要素类型）和气象变量的可疑值。为人工检查标示出可疑值，同时添加标记显示错误的类型和验证结果。

常见录入错误包括颠倒数字、小数点缺失或错位、数字缺失（例如零）、错误的累积周期和缺失数值。其它常见错误包括：(a) 观测员错误（例如非常规单位）；(b) 仪器故障；(c) 元数据问题（台站标识或命名规范不一致、错误的年份或月份）；(d) 某个月份和台站的特定要素的颠倒或重复；(e) 表格不清晰。这些错误能够显示出格式问题和系统的数据质量问题。

利用标记来确定每个错误的原因。在完成对所标记错误值的验证后，评估人员可对数据选择适当的验证类型。数值可处置为：(a) 正确、与表格一致、合理的且保持不变；(b) 错误，已标记但不变更数据；(c) 错误，但使用指定的替换值；(d) 不匹配，修正为与原始表格一致；(e) 视为错误并设置为缺失；(f) 视为需录入，随后录入。通常，如果标记值不明确且可修正时，则无需更改。

质量控制检验和程序可向用户保证，数字化的数据可准确地表示原始文件中所记录的观测值。如果 CDMS 没有自动进行检验，则应制定质量控制清单电子表格来跟踪对各台站的每次检验。每次检验的离群值都要手工检查，并在每次检验结束时应用任何修正值。在一个台站进行每个质量控制步骤后，要更新质量控制核算电子表格。

质量控制检验包括范围检查（每日值和每月总值及平均值）、内部一致性检查（比较数据类型之间的测量值）和极值检查（关于个别值）。更多信息可参见 WMO《气候规范指南》（WMO-No.100 (2016) 第 3.4 节）和 Westcott 等 (2011) 以及 [I-DARE 门户网站](#) 上的其它来源。

附录 9 经验教训

1. 鼓励 NMHS 预报部门与气候部门加强合作

NMHS 气候部门通常负责监督气象站点实地操作员、国家气候数据（每小时、每日表）质量控制，以及部分数据输入至自己的 CDMS。通常，编码 SYNOP 数据由天气观测员利用 WMO 全球电信系统发送，但这些数字数据通常仅提供给 NMHS 预报部门，且通常为实时数据。通常，仅将手写的小时和每日数据表发送给气候部门，而不是发送编码数字数据。此外，如果预报部门将 SYNOP 台站或自动台站的解码天气数据存档，则通常就不会将其提供给气候部门。因此，气候部门必须从国际数据库中获取其国家的数据。最理想的办法是传输的数字数据存档在中央数据库，让预报员和气候工作者都可以访问，而且所有的产品也都出自这个数据库。这些团体之间更好的协调可使气候部门能够有更更新的且最重要的是有更完整的数据集用于开展气候分析。它还要能够与印刷版报告进行比较，以开展极值验证。

2. 审议当前数字化、元数据收集和质量控制的做法；培训需求

NMHS 气候部门通常会制定自己的数据录入和质量控制方法，并根据现有的人员配置、通信和计算机资源以及国内气象台站的数量 / 类型逐渐完善。建议定期评审当前的做法，并将相关的 WMO 指导原则纳入当前数据的清单编制、录入和质量控制结构中。可参阅 [I-DARE 门户网站](#) 的最佳做法部分。

建议对在实施数据拯救过程中获取的任何见解进行审议，这些见解可能会完善数据拯救或当前数据摄取过程，或可能会导致定位需要拯救的其它数据源。气候数据的优劣取决于观测员和分析员。

从事数据拯救工作的人会碰到新的并不熟悉的设备和软件。使用数码相机、扫描仪、数字化软件、访问数据库和实行质量控制都需要培训和经验。由于人们的工作变动，应对多人进行培训，在他们学习该过程时了解“操作”手册，确保不出现单点故障。这些手册必须随着程序和用户的变化而做出修改。

3. 证明数据拯救的效益

拯救数据的再整理、图像化、数字化和质量控制步骤既昂贵又耗时。这或许就是许多 NMHS 尚未实施数据拯救的原因之一。数据拯救计划的启动和维持从 NHMS 开始，它需要明确说明数据拯救的必要性。这可包括确定那些可由改进的气候数据解答的实际问题。例如历史气候数据可使 NMHS 能够准确快速地回答部长关于干旱或强降雨事件频率方面的水文气象问题、或提供过去 100 年最强风的风向，从而能够在正确的方向修建耗资 1000 万美元的机场新跑道。此类信息给 NMHS 管理者带来可收费、具“附加值”的产品，为 NMHS

带来额外收入。此外，整理气候数据可以在不丢弃宝贵历史记录的情况下，为 NMHS 腾出数百平方米的楼层空间。增加的楼层面积可变为额外的办公室或培训室。最后，如果 NMHS 愿意共享已拯救并数字化的数据，它将在国际上被视为该领域的领导者。

数据拯救活动通常需要多年的外部支持。附录 10 或许能协助提供数据拯救的进一步理由。已开展数据拯救的个人和团体可以成为倡导者，他们也许能为数据拯救工作的筹资提供建议和意见。

与其他有意于数据拯救的机构磋商，这可为数据拯救过程中的意外问题提供技术信息和可能的解决方案。[I-DARE 门户网站](#)将提供一些技术信息和联系方式。NMHS 可联络 WMO 秘书处，以寻求其它会员的帮助。

附录 10 显示数据价值的应用

1. 验证气象参数和代用数据的极值

目前认为，未来的气候将不断变化且更为多样，在温度、降水、气压和风等方面将产生更多极值。若不了解过去出现的极值，就不可能将这些未来极值确认为“破纪录”。保护原始记录并将过去的天气记录数字化，便于验证当前和未来的极值。利用可验证的天气参数值，可加大了解极值的总体分布情况，以及这些值的分布是否在发生变化。过去对罕有或没有气象数据的时期或地区进行的气候研究都是基于代用数据（年轮、冰芯）。新拯救的数据在验证这些方法的有效性方面极具价值。

2. 研究的整合、与研究项目联系

展现恢复历史气候数据价值的最佳方式是将其用于研究或让其他人用于研究。长时间序列有利于确定和检验过去事件（例如干旱和洪水、热浪和寒潮、龙卷风、热带气旋和雹暴及尘暴）的严重性和频率。气象工作者和气候工作者特别关注检验天气现象的成因，而其他人关注的是这些现象对社会乃至对国家历史方向产生了什么影响。这些影响以及人类对过去事件的反应有助于对未来极端天气事件做出应对规划。利用历史数据研究过去的气候事件可极大地激发潜在投资者的热情，特别是如果相同的现象正在发生或预计将要发生的话。能够显示这些特征的（不断变化的）气候也有助于预测其未来的概率。完整的数据时间序列还有助于调整和验证气候模式。

3. 向用户提供数据

重要的是为用户提供气候数据并使之便于获取。提供可靠的历史数据使科研人员能够参与科学交流，并可能获得学生支持。农业和水资源管理机构将是促进气候数据拯救需求的良好盟友，它们可能有自己的记录或资源可以提供。这些管理机构可能会发现拯救的数据对评估自然和人为系统对气候事件和极端事件的敏感性至关重要。此外，开发公众普遍关注的产品（例如关于国家节假日的天气气候）可能会引起人们对天气的兴趣，有助于培养“公民科学工作者”群体。

附录 11 气候数据拯救核对清单

1. 搜寻和定位历史数据集：地方、区域、国家、国际图书馆和气候中心的纸质、微缩胶卷 / 微缩胶片和数字数据
2. 整理、清洁和储存纸质、缩微胶卷 / 缩微胶片，并在安全、有温度和湿度控制的房间内将其放置在架子上或文件柜中的档案盒内并贴上标签
3. 为纸质 / 缩微胶卷档案创建电子清单，包括台站、年份、月份、介质类型、表格类型和档案盒或文件抽屉位置
4. 为已图像化和未图像化、经验证和未经验证的数据创建电子图像清单表格
5. 为已数字化和未数字化、经质量控制和未经质量控制的数据创建数字数据电子清单表格
6. 确定待图像化数据和待数字化数据的优先顺序
7. 审议当前的数字化和质量控制步骤，以了解新数字化数据流如何适应 NMHS 的工作流程，以及当前的数字化和质量控制步骤是否需要修改
8. 确定计算机 (CDMS)、人员、空间、图像化、数字化需求
9. 制定有关图像化和数字化的工作计划
10. 必要时，获取所需的设备并聘请人员
11. 制定图像化过程，包括利用清单电子表格进行数据图像化和存档，以及对图像和元数据进行质量控制
12. 数据和元数据所有副本图像化，并适当存档图像
13. 图像清单纳入目录和 CD 中
14. 备份图像档案并异地存储
15. 保留原始纸质 / 微缩胶卷副本
16. 制定数字化流程，包括模板和数据库开发、更新清单以及数据和元数据质量控制
17. 随着工作的推进，审议图像化和数字化流程，以确定是否有改进该流程的方法
18. 对图像化和数字化的数据与台站的记录周期进行交叉检查，确保过程的完整性
19. 每天用数字数据备份计算机 /CDMS
20. 利用当前的数据和元数据，定期更新介质、清单和数字数据清单

21. 创建多个副本数字数据并存放在多个地点
 22. 制定技术迁移计划，并确保将所有电子数据（已图像化和数字化数据）定期迁移至新的存储介质
 23. 根据经质量控制的数据制作气候产品
 24. 以便捷、可访问的形式向用户提供数据
-

附录 12 术语表

ASCII: 美国信息交换标准代码, 是计算机文本文件最常见的格式。

自记图: 图的记录主要由发条鼓自动生成气象测量值, 如温度、风向和风速以及大气压 (与带状图相同)。

CDMS: 气候数据管理系统

翻拍支架: 将相机和灯固定位置拍摄文件的结构

DARE: 数据拯救。用于 WMO 气候数据拯救活动。

数字数据: 已录入计算机并存储为数字而不是模拟或图表的数据。这些数字可以是纯文本形式, 也可以编入关系数据库, 例如微软 ACCESS 或 ORACLE 或 CDMS 所提供的数据库。

数字化: 模拟数据转录为数字形式供计算机处理的过程。通常是通过录入文本数据或从带状图中划出点位。

电子清单: 含数据档案信息的电子表格

图像: 数码相机或扫描仪记录的气候数据文件的图片形式

录入: 用键盘录入数据或将数据输入计算机

介质: 用于写入数据的纸张、微缩胶卷、微缩胶片、CD/DVD、磁带、硬盘或计算机

元数据: 描述资源所需的一组属性或要素。在数据拯救中, 元数据指台站信息和气候数据清单信息。

微缩胶卷: 将显著压缩的记录图像复制在摄影胶卷上的一种方法

微缩胶片: 微缩胶卷的矩形薄片形式

迁移: 克服硬件和软件技术过时的一种方法, 它是将数据从一台计算机介质移至另一台计算机介质, 使数字对象的知识内容得以保护

NMHS: 国家气象和水文部门

保护: 确保数据免受损害或丢失

重新格式化: 将信息从一种存储介质复制到另一种存储介质或从一种文件格式转换为另一种文件格式

刷新: 将信息从一种存储介质复制到同一个或另一个存储介质而不改变格式

扫描: 用扫描仪将模拟记录或数字记录的硬拷贝制作成图像文件

扫描仪：用于将文件扫描为数字文件格式的设备

电子表格：一种格式，尤其用于制作清单。这里是指微软 EXCEL 或 Lotus 1-2-3 的电子表格软件。

带状图：图的记录主要由发条鼓自动生成气象测量值，如温度、风向和风速以及大气压（与自记图相同）。

附录 13 参考文献

- Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons and M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Amer. Meteor. Soc.*, 95:1431–1443.
- Brunet, M. and P. Jones, 2011: Data rescue initiatives: bringing historical climate data into the 21st century, *Climate Research*, 47:20–40.
- Westcott, N.E., K. Andsager, L. Stoecker, M.L. Spinar, R. Smith, R. Obrecht M.C. Kruk, R.T. Truesdell and D. O'Connell, 2011: Quality control of 19th century weather data. Midwestern Regional Climate Center Contract Report 11–04, Illinois State Water Survey, 68 pp.
- World Meteorological Organization (WMO), 2003: *Guidelines on Metadata and Homogenization*. WMO/TD-1186, WMO, Geneva.
- , 2004: *Guidelines on Climate Data Rescue*. WMO/TD-1210, WMO, Geneva.
- , 2009: *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation*. WMO-TD 1500, WCDMP-No. 72, WMO, Geneva.
- , 2010: *Guide to Metrological Instruments and Methods of Observation*. (CIMO Guide), WMO–No. 8 (update of 2010 edition).
- , 2016: *Guide to Climatological Practices*. WMO–No. 100 (Update of 2011 edition), WMO, Geneva.
-

欲了解更多信息, 请联系:

世界气象组织

7 bis, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland

战略传播办公室

电话: +41 (0) 22 730 87 40/83 14 – 传真: +41 (0) 22 730 80 27

电邮: communications@wmo.int

public.wmo.int