

Руководящие принципы по передовой практике спасения данных

Издание 2024 г.

ПОГОДА КЛИМАТ ВОДА



ВСЕМИРНАЯ
МЕТЕОРОЛОГИЧЕСКАЯ
ОРГАНИЗАЦИЯ

ВМО-№ 1182

Руководящие принципы по передовой практике спасения данных

Издание 2024 г.



ВСЕМИРНАЯ
МЕТЕОРОЛОГИЧЕСКАЯ
ОРГАНИЗАЦИЯ

ВМО-№ 1182

ВМО-№ 1182

© Всемирная метеорологическая организация, 2024

Право на опубликование в печатной, электронной или какой-либо иной форме на каком-либо языке сохраняется за ВМО. Небольшие выдержки из публикаций ВМО могут воспроизводиться без разрешения при условии четкого указания источника в полном объеме. Корреспонденцию редакционного характера и запросы в отношении частичного или полного опубликования, воспроизведения или перевода настоящей публикации следует направлять по адресу:

Chair, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box 2300
CH-1211 Geneva 2, Switzerland

Тел.: +41 (0) 22 730 84 03
Электронная почта: publications@wmo.int

ISBN 978-92-63-41182-0

ПРИМЕЧАНИЕ

Обозначения, употребляемые в публикациях ВМО, а также изложение материала в настоящей публикации не означают выражения со стороны ВМО какого бы то ни было мнения в отношении правового статуса какой-либо страны, территории, города или района, или их властей, а также в отношении делимитации их границ.

Упоминание отдельных компаний или какой-либо продукции не означает, что они одобрены или рекомендованы ВМО и что им отдается предпочтение перед другими аналогичными, но не упомянутыми или не прорекламированными компаниями или продукцией.

СОДЕРЖАНИЕ

	Стр.
ВЫРАЖЕНИЕ ПРИЗНАТЕЛЬНОСТИ	v
ОБЛАСТЬ ПРИМЕНЕНИЯ	vi
1. ВВЕДЕНИЕ — НАУЧНОЕ ЗНАЧЕНИЕ ДЕЯТЕЛЬНОСТИ ПО СПАСЕНИЮ ДАННЫХ	1
2. ОБЗОР ДЕЯТЕЛЬНОСТИ ПО СПАСЕНИЮ ДАННЫХ: АРХИВАЦИЯ, ОБРАБОТКА ИЗОБРАЖЕНИЙ И ОЦИФРОВКА	2
3. АРХИВАЦИЯ БУМАЖНЫХ НОСИТЕЛЕЙ	4
3.1 Поиск и определение местонахождения	4
3.2 Обеспечение сохранности и хранение	5
3.3 Создание электронных реестров массивов данных на бумажных носителях/микрофильмах и других устаревших носителях и изображений	6
4. ОБРАБОТКА ИЗОБРАЖЕНИЙ НА ИСХОДНЫХ НОСИТЕЛЯХ	7
4.1 Основные этапы процесса обработки изображений	8
4.2 Проверка и хранение графических файлов	9
5. ПРЕОБРАЗОВАНИЕ В ЦИФРОВУЮ ФОРМУ ЗНАЧЕНИЙ ДАННЫХ	10
5.1 Минимизация ошибок при вводе данных	11
5.2 Сравнение с международными базами данных	12
5.3 Метаданные	12
5.3.1 Сбор метеорологических метаданных	12
5.3.2 Специфика метаданных, относящихся к судовым наблюдениям	13
5.3.3 Сбор гидрологических метаданных	14
5.4 Оцифровка информации о зависимости между расходом и уровнем воды и измерении стока	16
6. АРХИВАЦИЯ ЦИФРОВЫХ НОСИТЕЛЕЙ (ИЗОБРАЖЕНИЙ И ЗНАЧЕНИЙ ДАННЫХ)	17
6.1 Перекрестная проверка печатных носителей, изображений и цифровых наборов данных	17
6.2 Обновление носителей	17
6.3 Рекомендации по архивации цифровых носителей	18
Приложение 1. Глобальная деятельность	19
Приложение 2. Наиболее распространенные примеры типов документов, которые могут содержать метеорологические, морские и гидрологические данные (и метаданные)	21
Приложение 3. Оценка спасения данных	23
Приложение 4. Инфраструктура, оборудование, принадлежности и персонал	24
Приложение 5. Руководящие указания по содержанию бумажных архивов	26
Приложение 6. Определение приоритетов при обработке изображений и оцифровке данных	27
Приложение 7. Методы электронной обработки изображений	31
Приложение 8. Другие методы оцифровки	35
Приложение 9. Ошибки при оцифровке и процесс обеспечения качества	40
Приложение 10. Накопленный опыт	45
Приложение 11. Применения для демонстрации ценности данных	48
Приложение 12. Контрольный перечень для спасения данных об окружающей среде ..	50
Приложение 13. Глоссарий	52

Приложение 14. Ссылки	54
-----------------------------	----

ВЫРАЖЕНИЕ ПРИЗНАТЕЛЬНОСТИ

Мы хотели бы поблагодарить следующих лиц за их выдающийся вклад в подготовку этой публикации:

г-на Эндрю М. Лоррея, Национальный институт водных и атмосферных исследований (НИВА), Новая Зеландия

г-на Акселя Андерсона, Метеорологическая служба Германии, Германия

г-на Денниса Штубера, МетеоФранс, Франция

г-жу Эмелин Рукот, МетеоФранс, Франция

г-жу Кейт де Смет, Университетский колледж Дублина, Ирландия

г-жу Мариэлу Васкес Гузман, Спасение климатических данных (CLIMDRE), Чили

г-на Мэттея Фрая, Центр экологии и гидрологии Соединенного Королевства, Соединенное Королевство Великобритании и Северной Ирландии

г-жу Нэнси Уэсткотт, Региональный климатический центр Среднего Запада, Служба водных ресурсов штата Иллинойс Университета Иллиноса (в отставке), Соединенные Штаты Америки

г-на Пола Поли, Европейский центр среднесрочных прогнозов погоды (ЕЦСПП), Соединенное Королевство

г-на Филипа Д. Джонса, Университет Восточной Англии, Соединенное Королевство

г-на Стефана Брённимана, Университет Берна, Швейцария

г-на Стефана Граба, Университет Витватерсранда, Южная Африка

Г-н Джерард ван дер Шриер, Королевский Нидерландский метеорологический институт (КНМИ), Королевство Нидерландов; г-н Рик Крутамель, Международная организация по спасению данных об окружающей среде (МОСДО), Соединенные Штаты Америки; г-н Роб Аллан, Метеобюро Соединенного Королевства (в отставке), Соединенное Королевство; а также г-н Вашингтон Отиено, г-жа Джоанна Корхонен и г-н Томмасо Абрате, Секретариат ВМО, оказали помощь в координации пересмотра этой публикации Службой по вопросам изменения климата в рамках программы «Коперник» (C3S), МОСДО, в рамках Моделей циркуляции атмосферы Земли (АКРЕ) и при участии гидрологических сообществ ВМО.

Многие другие коллеги внесли свой вклад в разработку этих руководящих принципов. Мы хотели бы выразить особую благодарность г-же Ричайху Ву, Секретариат ВМО, за анализ всех существующих руководящих принципов ВМО и C3S по спасению данных и за ее существенный вклад в подготовку первоначального проекта этой публикации.

Г-н Омар Баддур и г-н Пир Хехлер из Секретариата ВМО осуществляли общую координацию работы над публикацией.

ОБЛАСТЬ ПРИМЕНЕНИЯ

Данный технический документ представляет собой обновленный вариант *Руководящих принципов по передовой практике спасения данных* издания 2016 года (ВМО-№ 1182). В нем отражены руководящие принципы спасения данных Службы по вопросам изменения климата в рамках программы «Коперник» (C3S) Европейского союза (Wilkinson et al., 2019; Brunet et al., 2020; Wilkinson and Freemann, 2021), а также руководящие принципы из публикации ВМО *Guidelines for Hydrological Data Rescue* (Руководящие принципы по спасению гидрологических данных) (WMO-No. 1146).

Данные Руководящие принципы призваны обеспечить руководство в форме рекомендуемой передовой практики. Они включают обзор спасения данных и последующие разделы, посвященные его значению, архивации исходных носителей, обработке изображений, оцифровке и архивации цифровых изображений и цифровых данных. В четырнадцати приложениях представлена вспомогательная информация.

Ввиду разнообразия национальных метеорологических, морских и гидрологических служб с точки зрения их размера и уровня технического развития, а также изменчивости условий окружающей среды некоторые практики могут оказаться полезными не для каждого Члена ВМО. Вместе с тем Руководящие принципы охватывают широкий спектр руководящих указаний, которые должны обеспечить помощь в организации и осуществлении спасения данных и предоставить обобщенные технологические решения каждому Члену. Более конкретную технологическую информацию, а также содержательные примеры и фотографии можно найти на портале C3S по спасению данных, который поддерживается C3S в тесном сотрудничестве с ВМО и другими глобальными участниками процесса спасения данных.

Эта передовая практика может применяться для спасения научных данных, в частности, спасения метеорологических, гидрологических, морских и других данных об окружающей среде. Специфику данных об окружающей среде необходимо выявлять и учитывать в тесном сотрудничестве с соответствующими сообществами.

1. **ВВЕДЕНИЕ — НАУЧНОЕ ЗНАЧЕНИЕ ДЕЯТЕЛЬНОСТИ ПО СПАСЕНИЮ ДАННЫХ**

Спасение данных подразумевает обеспечение доступа и организацию деятельности по сохранению зарегистрированных данных наблюдений приборов и данных об окружающей среде, которые могут быть утрачены. Спасение данных имеет важнейшее значение для обеспечения доступа будущим поколениям ученых и других пользователей данных ко всей информации, необходимой для оценки изменчивости и изменения окружающей среды, а также для предоставления широкого спектра обслуживания в области окружающей среды. Эти данные помогают ликвидировать пробел между историческими данными и текущими наблюдениями. Длина доступных цифровых данных об окружающей среде варьируется в зависимости от страны и учреждения. В то время как многие данные наблюдений двадцатого века доступны в цифровом формате, имеются случаи, когда эти данные, наряду с более старыми записями девятнадцатого и восемнадцатого веков, существуют только в печатном виде и нуждаются в спасении и оцифровке. Достоверность климатических моделей и палеоклиматических данных (дендрохронологических, данных по керну льда, наблюдений и анализа пыльцы и т. д.) значительно повысится при наличии длинных временных рядов инструментальных наблюдений во всех регионах мира.

Регионы с недостаточным покрытием данными сохраняются и в двадцать первом веке. Поэтому получение и проверка более поздних данных имеют важнейшее значение. На надежность наблюдений за окружающей средой могут сильно повлиять местоположение и размещение станции, а также контроль качества (КК) процедур, связанных с проведением наблюдений. Сравнение данных наблюдений с одного объекта с данными с близлежащих объектов и изучение исходной документации наблюдений (и исходной информации о методах наблюдений) позволят ученым обеспечить более качественную проверку прошлых и настоящих режимов окружающей среды.

Расширение рядов данных об окружающей среде за счет прошлых периодов и заполнение временных и пространственных пробелов дают следующие преимущества:

- a) помогают повысить надежность представления агрометеорологических данных, данных о векторе распространения заболеваний, а также гидрологических и климатологических численных моделей и обеспечивают более качественные перспективные оценки будущего климата;
- b) объединение спасенных данных с уже имеющимися способствует лучшему представлению текущей погоды и климата в историческом контексте;
- c) обеспечивают основу для оценки исторической чувствительности природных и антропогенных систем к изменчивости и изменению окружающей среды, тем самым позволяя более точно оценить воздействие будущих изменчивости и изменения окружающей среды.

Такие оценки могут послужить в качестве исходных данных для лиц, занимающихся выработкой политики, с целью уменьшения потерь от погодных бедствий и предоставления более полной информации для целей экономического развития.

Для обеспечения практической доступности поставщикам обслуживания и исследователям в области окружающей среды данные (включая метаданные) (как изображения, так и численные значения) должны быть предоставлены в цифровой форме. Кроме того, все текущие данные также должны сохраняться и предоставляться в цифровой форме, поскольку в будущем эти наблюдения будут рассматриваться как исторические данные. Исходная документация играет решающую роль в проверке цифровых значений.

2. **ОБЗОР ДЕЯТЕЛЬНОСТИ ПО СПАСЕНИЮ ДАННЫХ: АРХИВАЦИЯ, ОБРАБОТКА ИЗОБРАЖЕНИЙ И ОЦИФРОВКА**

Спасение климатических данных подразумевает обеспечение доступа и организацию деятельности по сохранению данных об окружающей среде, которые могут быть потеряны в результате повреждения, разрушения, запущенности, технического износа или простого рассредоточения массивов данных об окружающей среде с течением времени. Неоцифрованные данные подвергаются опасности в силу уязвимости исходных бумажных/медианосителей данных. Спасение данных предусматривает организацию и сканирование данных на бумажной основе, микрофильмах, микрофишах и других устаревших носителях; работу на клавиатуре с численными и текстовыми данными и оцифровку ленточных диаграмм в соответствующем формате, а также архивацию данных, метаданных и результатов и процедур проверок качества. Обзор компонентов деятельности по спасению данных приводится в таблице 1. Общими для каждого из этих компонентов являются этапы планирования, осуществления и контроля качества, обеспечивающие успешное и качественное спасение данных. Выделение времени в начале проекта/компонента для определения необходимых этапов, требований к данным и оборудованию, а также для разработки необходимых мер контроля качества позволяет специалистам-практикам выявить потенциальные проблемы и установить соответствующие процедуры.

Большинство стран вели систематические наблюдения за состоянием окружающей среды, регистрировали свои наблюдения вручную или автоматически, заполняли бумажные формы или переносили их на микрофильмы или другие носители и в конечном итоге сохраняли их на некоем типе компьютерных носителей для облегчения доступа и анализа. К сожалению, в некоторых случаях эти данные не переносились с бумажных форм, будь то в силу отсутствия средств и/или нехватки персонала или по причине того, что данные хранились в другом месте, иногда за рубежом. Кроме того, многие бумажные формы, копии микрофильмов/микрофиш и другие устаревшие носители могут быть утрачены в результате быстрого износа носителя или не оформленного документально перемещения частей архивов.

Цель настоящих Руководящих принципов — представить рекомендации в отношении шагов, которые необходимо предпринять для организации, обработки изображений, оцифровки и обеспечения сохранности таких данных об окружающей среде. В приложениях 1-3 представлена дополнительная информация о спасении данных как глобальной деятельности (приложение 1), о типах документов, содержащих метеорологические и климатические, гидрологические и морские данные (приложение 2), а также контрольный перечень для организации новых мероприятий по спасению данных (приложение 3).

Таблица 1. Компоненты деятельности по спасению данных

<i>Компоненты спасения данных</i>	<i>Деятельность</i>	<i>Ключевые слова</i>
Архивация информации на бумажных носителях, в виде микрофильмов/микрофиш и на других устаревших носителях (раздел 3)	Поиск и определение местонахождения	Национальные метеорологические и гидрологические службы (НМГС), пункты наблюдений, университеты, авиационные и морские учреждения, гидрологические и сельскохозяйственные организации, международные библиотеки и базы данных, национальные архивы
	Поддержание в сохранности и хранение	Чистые среды; размещение в бескислотных маркированных архивных коробках, защищенных от пыли, влаги и насекомых
	Создание электронных реестров массива данных на бумажных носителях/микрофильмах или устаревших носителях	Каталогизация всех носителей информации на бумажной основе; оценка масштабов деятельности по обработке изображений и оцифровке данных
Средства обработки изображений (раздел 4)	Создание основного реестра изображений	
	Обработка и проверка изображений Обновление основного реестра изображений Создание реестров графических файлов на каждом цифровом устройстве хранения/в каждом каталоге данных	Обновление основного реестра изображений после их проверки на читаемость, включая метаданные Перекрестная проверка реестров файлов на внешних носителях с основным реестром изображений
Оцифровка значений данных (раздел 5)	Создание реестра цифровых данных	
	Ввод с клавиатуры, перенос данных диаграмм	Ввод данных в систему управления данными
	Контроль качества данных	Обновление реестра цифровых данных по мере оцифровки данных и прохождения различных проверок контроля качества
Архивация цифровых носителей информации (раздел 6)	Перекрестная проверка печатных носителей информации, изображений и оцифрованных данных	Сравнение реестров изображений и цифровых данных с исходными электронными реестрами массива данных на бумажных носителях/микрофильмах или устаревших носителях
	Резервное копирование электронных носителей информации	Ежедневно
	Распределение дубликатов архивов изображений и цифровых данных	Разные места хранения
	Обновление носителей и миграция технологий	Каждые 5—10 лет

3. **АРХИВАЦИЯ БУМАЖНЫХ НОСИТЕЛЕЙ¹**

3.1 **Поиск и определение местонахождения**

Первоочередной задачей спасения данных является определение их местонахождения. Прежде чем приступить к процедуре обеспечения сохранности, крайне важно идентифицировать исторические наборы данных и признать их в качестве значимых данных об окружающей среде. Они могут храниться в национальных архивах и службах, на станциях наблюдений или в частных коллекциях и других местах, что часто бывает неизвестно экспертному сообществу. Некоторые данные могут храниться в зарубежных архивах, особенно в архивах стран с колониальной историей. Без проведения такого поиска многие потенциально ценные наборы данных могут остаться невыявленными, и доступ к ним может быть невозможен. Поиск в следующих местах и источниках может выявить ценные данные об окружающей среде, которые еще не были упорядочены и оцифрованы и управление которыми еще не было реализовано в рамках существующей системы управления данными и которые могли быть так или иначе упущены из виду по другим причинам:

- a) НМГС и другие правительственные департаменты или агентства, которые часто располагают собственными массивами данных;
- b) океанографические и полярные институты, морские рыболовные промыслы, морские обсерватории, учреждения береговой охраны и агентства по охране морской среды;
- c) астрономические, магнитные и геофизические обсерватории;
- d) архивы служб национальных парков, служб электросвязи и почтовых служб, служб аэропортов и авиационных служб, коммерческих компаний;
- e) гидрологические институты и учреждения, а также энергосбытовые компании с проектами в области гидроэнергетики и другие компании, связанные с реками и озерами;
- f) государственные учреждения, национальные, региональные и местные архивы, дипломатические и консульские службы;
- g) университеты, колледжи, средние школы и частные лица, например, давно работающие сотрудники метеорологических и связанных с ними организаций;
- h) международные центры данных;
- i) библиотеки или местные историки, архивы которых могут располагать копиями данных на бумажных носителях, микрофильмах/микрофишах или даже цифровыми копиями данных;
- j) такие сельскохозяйственные организации, как крупные продовольственные корпорации, содержащие плантации по всему миру и располагающие метеорологическими данными наблюдений за сотни лет;
- k) военные ведомства и министерства транспорта, которые могут располагать данными, к которым может быть предоставлен доступ, особенно авиационными и морскими метеорологическими данными;
- l) религиозные организации, особенно заинтересованные в осуществлении наблюдений или имеющие научные интересы, например, монашеские ордены, занимающиеся обучением и воспитанием, и отдельные представители духовенства;

¹ Архивация цифровых носителей рассматривается в разделе 6.

- m) научные общества, особенно организации, специализирующиеся на исторических данных (общества естествознания);
- n) музеи;
- o) судовые журналы, исторические газеты и личные дневники, которые также могут быть ценными источниками информации о погоде и климате.

В процессе поиска необходимо провести перекрестную проверку имеющихся цифровых данных, чтобы определить, были ли они уже оцифрованы, выявить существенные пробелы и пролить свет на возможные даты начала программ наблюдений в конкретных регионах. Она предусматривает проверку по международным базам данных, включая:

- a) Глобальную сеть исторических климатологических данных;
- b) Международный банк данных о приземном давлении;
- c) Интегрированную базу данных приземных наблюдений;
- d) Глобальный центр данных по стоку;
- e) Глобальную систему информации о подземных водах.

3.2 Обеспечение сохранности и хранение

Массивы данных на бумажных носителях и микрофильмах/микрофишах и массивы данных на других устаревших носителях должны быть организованы логически и храниться в бескислотных архивных коробках на прочных полках или в картотечных шкафах, при этом следует учитывать следующее:

- a) в зависимости от предыдущего местонахождения и состояния бумажных диаграмм или форм, перед помещением на хранение бумажные копии, возможно, потребуется очистить от пыли или пропылесосить;
- b) при работе с возможно загрязненными данными необходимо соблюдать надлежащие процедуры в области безопасности и гигиены труда;
- c) для удаления пыли с хрупких документов требуется специальное оборудование и подготовка;
- d) необходимо соблюдать особую осторожность для обеспечения безопасности оператора, если документы заплесневели или были обработаны ингибиторами плесени или пестицидами;
- e) с хрупкими документами следует работать сухими чистыми руками, держась за края, если это необходимо; для защиты здоровья оператора можно использовать чистые нитриловые или хлопчатобумажные перчатки;
- f) в зависимости от количества пыли могут потребоваться защитная одежда и маски.

Под хранение архивных данных на бумажных носителях и микрофильмах или данных, хранящихся иным образом, часто выделяются специальные помещения. В этом случае необходимо учитывать следующие моменты:

- a) в идеальном варианте в помещениях должны регулироваться температура и влажность;
- b) данные должны быть защищены от насекомых, грызунов, плесени, пожара, наводнения, пыли, хищения и прочих опасностей;

- c) во избежание возникновения пожара и нашествия насекомых необходимо свести к минимуму использование в помещении деревянных конструкций;
- d) в организации процесса может помочь профессиональный архивариус, а периодический осмотр помещения следует поручить менеджеру.

Данные обычно хранятся в бескислотных архивных коробках или в картотечных шкафах с разбивкой по типам (например, тип формы или диаграммы), а также по станциям, годам и месяцам. При хранении данных таким образом:

- a) от количества лет, за которые имеются данные, и от количества станций может зависеть наиболее логичный способ, которым должны быть организованы бумажные копии — по годам или станциям;
- b) промаркируйте архивные коробки и/или картотечные шкафы и добавьте место хранения (номер коробки или ящика картотечного шкафа), а также станцию/год/месяц, тип носителя и тип формы в электронный реестр массива данных на бумажных носителях/микрофильмах/других носителях, чтобы данные, если потребуется, можно было легко найти для обработки изображений и оцифровки.

При долгосрочном хранении необходимо учитывать следующее:

- a) после обработки изображений критически важных данных не следует выбрасывать исходные документы. Если ресурсы не позволяют обеспечить длительное хранение, по этому вопросу следует обратиться в национальные архивы, ведомственные/служебные или университетские библиотеки. Если эти институты не могут хранить архивы, и если это допускается национальным архивным агентством, по данному вопросу следует обратиться в международные библиотеки или институты, занимающиеся хранением (дополнительные сведения можно найти в приложении 5);
- b) микрофильм непригоден для постоянного архивного хранения. При несоблюдении условий хранения в специальных помещениях с кондиционером и осушителем воздуха микрофильм приходит в негодность;
- c) цифровые изображения также не являются способом постоянной архивации (см. раздел 6.2).

Дальнейшие сведения о необходимой инфраструктуре, материалах и персонале приводятся в приложении 4. Предложения по содержанию бумажных архивов изложены в приложении 5.

3.3 **Создание электронных реестров массивов данных на бумажных носителях/микрофильмах и других устаревших носителях и изображений**

В процессе организации и хранения бумажных носителей требуется сформировать ряд реестров для облегчения перехода от бумажных копий к электронным изображениям и цифровым данным. Это необходимо для того, чтобы после обнаружения и организации данных на бумажных носителях к ним можно было получить доступ для обработки изображений. В дальнейшем, когда изображения будут созданы, они могут быть оцифрованы, после чего исходные данные на бумажных носителях и электронные изображения могут быть легко извлечены для контроля качества цифровых значений. В идеальном варианте протокол именования должен сохранять как можно больше информации на каждом этапе. В таблице 2 кратко описаны шаги, необходимые для создания архива исходных носителей (бумажных и устаревших носителей), а также для последующей обработки, учета изображений и размещения графических файлов.

В целом, по мере обнаружения данных об окружающей среде следует включать их в электронный реестр массивов данных на бумажных/старых носителях, чтобы определить,

какие имеются данные, какие усилия по их спасению уже были предприняты и что еще предстоит сделать. Данные электронного реестра должны включать станцию, год, месяц, тип носителя, номер или тип формы, тип элемента, количество страниц (включая страницы метаданных) и, в частности, местоположение коробки/ящика картотечного шкафа и уникальные идентификаторы, которые используются для обозначения документов. Также должен быть включен тип переменной.

После того, как будет создан электронный реестр исторических данных об окружающей среде, его надлежит поддерживать и обновлять по мере получения новых данных; также по мере обработки изображений в основной реестр изображений необходимо добавлять сведения об их новом местонахождении. Основной реестр изображений может использоваться для оценки хода выполнения задачи по их обработке. Более подробно задача по обработке изображений рассматривается в разделе 4.

4. ОБРАБОТКА ИЗОБРАЖЕНИЙ НА ИСХОДНЫХ НОСИТЕЛЯХ

По завершении сбора и каталогизации бумажных копий и микрофильмов или копий, хранящихся иным образом, можно приступить к планированию, чтобы определить порядок обработки изображений и оцифровки данных. С чего начать и как выполнить эти задачи будет зависеть от потребностей конкретных НМГС. Приоритеты при обработке изображений и оцифровке данных, вероятно, будут отличаться. В приложении 6 рассматриваются вопросы, которые необходимо учитывать при определении приоритетов в порядке обработки и оцифровки.

Бумажные и/или микрофильмированные либо хранящиеся иным образом страницы фотографируются или сканируются для создания изображения или цифрового снимка, которые затем можно уберечь от дальнейшего разрушения и сделать широкодоступными. Все страницы исходных метеорологических документов (включая пустые страницы и обложки) должны быть обработаны с помощью цифровой фотокамеры или оптического сканера. Это необходимо для того, чтобы обеспечить сохранность всех имеющихся метаданных (например, в отношении практики наблюдений, описания местности и данных). В приложении 7 рассматриваются относительные преимущества цифровых фотокамер и сканеров, желательные характеристики фотокамер, а также правила и тонкости обработки изображений.

Во всех случаях метаданные (описание данных и практики наблюдений, местонахождение и характеристики местности) подлежат спасению и оцифровке вместе с изображениями и цифровыми значениями.

Таблица 2. Этапы создания и ведения электронных реестров данных

Собрать все данные и метаданные для помещения на хранение	
Упорядочить данные в соответствии с логически организованным архивным планом	По станциям/типу формы/году По типу формы/станции/году По году/станциям В отношении данных диаграмм также по метеорологическим элементам (ветер, осадки и т. д.)
Создать электронный реестр массивов данных на бумажных/устаревших носителях	Включая станцию, год и местоположение коробки/ящика картотечного шкафа, тип носителя, тип формы Указать типы переменных Определить отсутствующие периоды Оценить количество типов носителей, типов форм и общий объем изображений для определения плана работы

Создать основной реестр изображений на основе электронного реестра массивов данных	Идентифицировать станцию, год, месяц, тип носителя, тип или номер формы Определить периоды наблюдений Оценить количество страниц для обработки изображений
Создавать реестры изображений по мере создания изображений на каждом цифровом устройстве хранения/в каждом каталоге для удобства определения местонахождения графических файлов или в каждом компьютерном каталоге	
Включить страницы, содержащие только метаданные (информацию о сетях, практике наблюдений и т. д.)	
Вести основной реестр изображений по мере выполнения процесса обработки изображений	Указывать количество обработанных страниц по мере проверки изображений Указать местонахождение цифрового устройства хранения/каталога или компьютера
Выполнить перекрестную проверку основного реестра изображений и реестров графических файлов на предмет количества изображений на цифровых устройствах хранения/в каталогах	

4.1 Основные этапы процесса обработки изображений

Практика и методы обработки изображений могут быть разными, но все они требуют отслеживания документации и отслеживания файлов. При работе с хрупкими документами необходима особая осторожность во избежание их дальнейшего разрушения в процессе обработки изображений (включая транспортировку из архива к месту сканирования и обратно); при наличии сомнений можно проконсультироваться с профессиональным архивариусом (это касается и задачи получения изображений страниц из книг или дневников таким образом, чтобы не повредить переплет). К числу настоятельно рекомендуемых основных процедур относятся следующие:

- a) создать основной реестр изображений на основе электронного реестра массива данных на бумажных/старых носителях, как указано в таблице 2. Указать количество страниц для обработки изображений. Обновлять основной реестр изображений по мере их обработки и проверки путем указания количества обработанных и проверенных страниц и добавления сведений о местонахождении компьютерного каталога или цифрового устройства хранения/каталога (предоставить шаблон);
- b) по мере обработки изображений регулярно проверять обработанные файлы на полноту и читаемость, чтобы убедиться, что изображения считываются, а их информационное наполнение соответствует имени файла, а также регулярно организовывать обработанные файлы и осуществлять их резервное копирование в соответствии с установленным протоколом во избежание путаницы между файлами или их потери (см. также раздел 4.2 ниже);
- c) большое количество изображений будет создано с помощью цифровой фотокамеры или сканера. Для каждого изображения при его создании и загрузке на компьютер программное обеспечение фотокамеры/сканера генерирует имя файла. Имена файлов могут не иметь отношения к их информационному наполнению. Некоторые из них могут повторяться, если изображения многократно загружаются на компьютер в один и тот же день. Рекомендуется использовать программное обеспечение для создания имени файла, которое связано с информационным наполнением изображения (см. пример в пункте d) ниже). Программное обеспечение для выполнения этой задачи поставляется в комплекте с некоторыми фотокамерами и сканерами или может быть приобретено отдельно. См. портал C3S по спасению данных и приложение 7;

- d) имя графического файла должно включать идентификационный номер станции (SID), акроним для обозначения типа формы (ACR), дату (YYYYmmDD) и номер страницы (PPP). Обычно это файлы в формате .png, .jpg или .tif (например, SID_ACR_YYYYmmDD_PPP.png);
- e) хранить изображения в соответствующей структуре, например, в каталогах «станция/год/месяц». Для простоты определения местонахождения и извлечения изображений следует вести основной список имен файлов, включенных в главный каталог, и аналогичные списки файлов, расположенных в отдельных каталогах или на цифровых устройствах хранения;
- f) обновлять основной реестр изображений с учетом количества обработанных страниц и сопоставлять исходное количество страниц с количеством изображений, фактически созданных и сохраненных на цифровых устройствах хранения/в каталогах. Количество страниц, подсчитанных и внесенных в реестры, должно быть равно количеству созданных изображений. Указать местонахождение изображений (цифровое устройство хранения/облачное хранилище или наименование каталога);
- g) документирование процесса обработки изображений, включая соглашение о порядке именования файлов и любые конкретные проблемы, возникающие в процессе обработки изображений, имеет решающее значение, поскольку персонал, участвующий в обработке изображений, со временем может меняться.

4.2 Проверка и хранение графических файлов

Просмотрите и проверьте обработанные графические файлы, прежде чем записывать их на цифровое устройство хранения/в облачное хранилище или в систему управления климатическими данными (СУКД) (см. [Climate Data Management System Specifications](#) (Спецификации Системы управления климатическими данными) (WMO-No. 1131)), и имейте в виду следующие моменты:

- a) изображения могут быть не в фокусе из-за положения фотокамеры;
- b) изображения могут быть слишком бледными, а исправления, сделанные ручкой или карандашом, могут быть не видны;
- c) изображения могут быть слишком темными из-за настроек фотокамеры/сканера;
- d) части обработанных данных могут быть засвечены. Использование прямого освещения с помощью вспышки фотоаппарата или со штатива может привести к обратному отражению света в объектив фотоаппарата и засвечиванию части изображения, особенно на глянцевой бумаге.

Эффективный подход к обеспечению качества (ОК) заключается в том, чтобы периодически (например, через каждые 300 страниц) проверять несколько изображений на предмет читаемости. Также рекомендуется проверка полноты изображений и их сравнение с бумажным архивом, чтобы убедиться, что ни одна страница с данными не была пропущена. Важно хотя бы раз в месяц/в год/в отношении станции убедиться в том, что имя графического файла соответствует его информационному наполнению.

В случае использования цифровых устройств хранения они должны быть промаркированы.

- a) При использовании имен графических файлов их следует вносить в реестр файлов вместе с различными идентифицирующими метаданными (название и номер станции, год, месяц, тип формы, количество страниц) и маркировкой цифрового устройства хранения/местонахождением облачного хранилища;

- b) копия имен файлов, связанных с каждым цифровым устройством хранения, должна быть указана на таком цифровом устройстве хранения и внесена в основной каталог;
- c) маркировка цифрового устройства хранения должна быть включена в основной реестр изображений.

Чрезвычайно важно создать дубликаты (резервные копии) архива изображений на цифровых устройствах хранения/в облачных хранилищах и/или в СУКД/системе управления гидрометрическими данными (СУГД). Если данные с цифровых устройств хранения/из облачных хранилищ копируются в СУКД/СУГД, каждое цифровое устройство хранения/облачное хранилище должно иметь каталог, и все реестры файлов также подлежат копированию.

Фундаментальным требованием в отношении надлежащего управления данными является удаленное хранение резервных копий данных и метаданных в качестве защиты на случай их утраты при повреждении зданий и компьютеров в результате стихийных бедствий или антропогенных катастроф.

- a) Важно обеспечить распределенное хранение копий всего электронного архива изображений и основного реестра изображений;
- b) если обработка данных производится в региональном или многонациональном центре, копии данных должны распределяться по странам происхождения данных и храниться в региональном центре. При планировании обработки данных в региональном центре необходимо наличие предварительного соглашения между участвующими странами, в котором оговаривается, какие действия необходимо предпринять в случае повреждения зданий или компьютеров в результате стихийных бедствий или антропогенных катастроф. Секретариат ВМО может оказать помощь в заключении такого рода соглашений.

5. ПРЕОБРАЗОВАНИЕ В ЦИФРОВУЮ ФОРМУ ЗНАЧЕНИЙ ДАННЫХ

Оцифровкой называют как перенос данных с диаграмм (включая ленточные диаграммы самописца), так и преобразование текстовых данных в цифровой формат, пригодный для записи и чтения на электронных носителях. Первый случай рассматривается в приложении 8; оцифровка буквенно-цифровых данных посредством ввода данных рассматривается в настоящем разделе. Оцифровка часто производится в НМГС. Иногда она выполняется и в региональных центрах силами компаний или в рамках гражданских научных инициатив, обычно под руководством экспертов в соответствующей области, например, климатологов, гидрологов и т. д. Если печатные данные находятся в национальном архиве, то, скорее всего, оцифровка потребует на месте. При планировании обработки данных вне НМГС или даже за пределами страны необходимо предварительно утвердить соглашение между сторонами. Опять же, Секретариат ВМО может оказать помощь с соглашениями такого рода.

Оцифровка обычно занимает гораздо больше времени, чем обработка изображений, и из-за длительности процесса оцифровки ограниченные ресурсы могут помешать оцифровать все данные. Степень приоритетности подлежащих оцифровке данных должна определяться с учетом научных, технических и социально-экономических факторов. В случае невозможности внести всю информацию в форму в цифровом формате особенно важно обеспечить сохранность исходных форм и их копий. Более подробно вопрос о приоритетах оцифровки рассматривается в приложении 6.

В некоторых случаях в отсутствие средств для оцифровки больших объемов данных используется краудсорсинг. Добровольцы могут вводить данные в электронную форму или переносить временные ряды с диаграмм на специально созданные веб-сайты. Другой подход заключается в использовании при оцифровке печатных материалов программного обеспечения для оптического распознавания символов (ОРС). Этот подход находится на

стадии тестирования и может более эффективно и широко использоваться в будущем. В настоящее время ОРС имеет ограниченное значение, поскольку требует использования специализированных форм. Более подробно вопросы краудсорсинга, диаграмм и ОРС рассматриваются в приложении 8 и на портале СЗС по спасению данных.

Первым шагом в оцифровке данных является разработка реестра цифровых данных, по аналогии с реестром изображений, для отслеживания процесса оцифровки. Данные должны быть организованы по станциям, годам, месяцам и типу данных.

Независимо от того, какой метод будет применяться (ввод данных или перенос данных с диаграмм), в начале проекта необходимо разработать процедуры, описывающие шаги по оцифровке данных и контролю качества выходной продукции, используя подмножество данных. Внимательно изучив эту выборку, специалисты-практики смогут выявить общие черты в своих данных (например, выцветшие или нечеткие записи, пробелы в данных, исторические аннотации) и принять решение о подходе к работе с проблемными данными и внесению соответствующей информации, если это необходимо. Предоставление специалистам по оцифровке хороших инструкций также помогает избежать проблем с качеством данных, особенно если оцифровщики не являются экспертами. Прежде чем составлять инструкции для современных специалистов по оцифровке, следует собрать и изучить инструкции, использовавшиеся в прошлом для сбора и обработки данных, а также учесть языковые проблемы и специфические термины, характерные для данной области. Инструкции должны быть как можно более подробными (см. приложение 9).

5.1 Минимизация ошибок при вводе данных

Существует ряд приемов, позволяющих минимизировать ошибки при вводе данных. Первый касается двойного или тройного ввода данных (когда одни и те же данные вводят двое или более лиц). Это можно сделать, поручив нескольким операторам ввести одинаковые данные, а затем сопоставив каждое значение данных. В качестве альтернативы один оператор или группа операторов могут вносить суточные данные, в то время как другой вносит данные за месяц; затем можно оценить оба набора данных на предмет согласованности месячных общих и усредненных значений. Двойной ввод данных может осуществляться вне СУКД, а данные импортироваться после их сопоставления.

Второй прием, позволяющий сократить ошибки при вводе, заключается в создании шаблона формы, в который вводятся данные. Благодаря использованию шаблона оператор точно знает, где найти данные на странице, что уменьшает количество ошибок. Шаблон также может включать проверку правильности введенных в форму данных, что еще больше снижает количество ошибок в процессе преобразования. Помимо этого, данные, внесенные в шаблон, могут напрямую импортироваться в базу данных. Даже при частичном заполнении формы рекомендуется создать шаблон для ввода данных по каждому типу формы.

В-третьих, очень важно воспроизводить данные именно так, как они напечатаны (т. е. один в один). Этот прием известен как «набираю, что вижу» в Австралии и как "saisir ce que l'on voit" во Франции. Следует избегать всех форм кодирования или пересчета единиц измерения в реальном времени. Перекодирование, изменение единиц измерения, усреднение и даже переперекодирование весьма эффективно выполняются с помощью компьютеров. Ввод данных в том виде, в каком они напечатаны, может обеспечить наличие точной записи формул, использованных для перекодировки данных. Если необходимо использовать кодирование, следует позаботиться о том, чтобы оно было хорошо задокументировано и обратимо (обеспечить возможность вернуться к исходным данным). Данные должны вводиться в точности так, как они указаны в исходной форме наблюдений, даже в том случае, если наблюдатель очевидно допустил ошибку. Эти данные иногда используются в судебных делах, поэтому важно сохранить их в исходном виде. Конечно, ошибочные значения также должны быть отмечены и, возможно, исправлены при дальнейшей обработке данных. Копия данных в исходном виде и цифровые данные с исправленными ошибками должны храниться как два разных набора данных.

Двойной/тройной ввод данных, создание шаблона и воспроизведение один в один значительно сократят количество ошибок ввода данных. После ввода или импорта данных в СУКД в качестве предварительных данных необходимо произвести контроль качества (см. приложение 9). Этот этап невозможно переоценить.

Наконец, рекомендуется сразу вводить все данные на странице (например, широту, долготу, имя наблюдателя, время наблюдения, другие метеорологические переменные и все численные и текстовые данные). Главным образом это связано с тем, что вряд ли появится другая возможность вернуться к этой странице и ввести пропущенные данные. Оцифровка всех переменных также способствует более эффективному контролю качества записи.

5.2 Сравнение с международными базами данных

НМГС могут быть заинтересованы в приобретении цифровых климатических данных, хранящихся в международных базах данных, полученных из кодов SYNOP или METAR. Нередко эти данные не помещались на хранение внутри организации в момент их появления, и доступ к ним может быть затруднен из-за пропускной способности, необходимой для загрузки очень больших файлов данных. Получить такие данные представляется желательным, поскольку они уже оцифрованы и могут содержать более частые наблюдения и больше переменных параметров, чем хранится в собственной климатической базе данных. Несмотря на то, что эти наборы данных, скорее всего, прошли процедуры обеспечения качества, они все еще могут содержать ошибки, допущенные при вводе, кодировании и передаче. Однако если значение на исходной бумажной форме было неясно, возможно, что его можно восстановить по коду SYNOP. Оцифровка исходных носителей и сопоставление с данными из международных баз данных приведет к созданию более полной и точной версии климатических данных. Дальнейшее подтверждение происхождения данных обеспечивается за счет включения данных из международных баз данных и вновь оцифрованных данных наблюдений и изображений во всех возможных случаях.

5.3 Метаданные

Метаданные содержат информацию и предоставляют подробные сведения о данных, например, о том, как, где, когда и кем производились регистрация, сбор, передача данных и управление ими (Aguilar, et al., 2003).

5.3.1 Сбор метеорологических метаданных

Минимальные требования к метеорологическим метаданным включают: информацию об идентификаторах станций/платформ, географические координаты и высоту над уровнем моря площадок для производства наблюдений, особенности обработки данных (составные единицы, специальные коды, расчеты, алгоритмы и т. д.).

Передовая практика требует более полных метаданных, таких как:

- идентификаторы станций/платформ (например, дата начала/окончания работы станции, тип станции, организация, ответственная за данные);
- географические данные (например, топографическая информация, местоположение);
- местная обстановка (например, местное землепользование/земной покров, препятствия, тип почвы);
- техническое обслуживание приборов (например, тип приборов, их укрытие и монтаж, результаты калибровки и инспекции приборов);

- практика наблюдений (например, сроки и графики производства наблюдений, внесение поправок);
- обработка данных (например, методы и алгоритмы);
- исторические события и передача данных (например, форматы, период, за который сообщаются данные).

5.3.2 **Специфика метаданных, относящихся к судовым наблюдениям**

Дата и срок наблюдений

В разных странах используются разные форматы даты, и это будет отражено в исходных документах. Например, 4 июля 1921 года может быть записано как 4/7/1921 в британских документах и 7/4/1921 в документах из США. В более старых документах месяц может быть написан полностью или использована римская цифра, например, 4 VII 1921. Даты должны быть перенесены из исходного документа в согласованном формате, исключающем неоднозначность. Шаблон для ввода данных должен предусматривать отдельные ячейки для указания года, месяца и дня в электронной таблице или в веб-интерфейсе. Формат (ГГГГ, ММ, ДД) в отдельных колонках согласуется с форматом, используемым для наземных данных, и позволяет избежать неоднозначности.

Проблема с вводом дат из судовых журналов или метеорологических форм касается определения даты, когда морское судно пересекает Международную линию перемены даты в любом направлении. При вводе данных важно учитывать, что при пересечении линии дата будет меняться. Однако изменение даты могло быть произведено автором судового журнала не сразу, и день может быть либо пропущен, либо продублирован в зависимости от направления движения. Это необходимо исправить с помощью проверки обеспечения качества.

Кроме того, определение дня и времени суток по бортовым журналам девятнадцатого века может быть малопонятным, если персонал, ответственный за спасение данных, не знаком со структурой корабельных, или навигационных, суток. Вплоть до девятнадцатого века корабельные сутки были привязаны к полудню, поскольку именно в это время определялось местоположение судна и рассчитывался уточненный курс. Фактически корабельные сутки заканчивались в полдень, и начинался новый день, то есть корабельные сутки опережали гражданские на 12 часов. Если упустить это из виду в процессе ввода данных, то наблюдения за периоды, составляющие менее суток, будут отнесены к неправильному дню и дате. Рекомендуется внести изменения в формат ввода данных, чтобы более полно учитывать эту структуру.

В некоторых гораздо более старых журналах и реестрах сутки могут быть разделены на три части (первую, среднюю и третью/последнюю), обозначающие время с полудня до 20:00, с 20:00 до 4:00 и с 4:00 до полудня. Необходимо установить время проведения этих наблюдений. Наблюдения, проводимые в полдень, которые, с точки зрения корабельных офицеров, являются самыми важными за сутки, очевидно приходятся на конец третьей части суток, поэтому два других эпизода наблюдений должны быть отнесены к 20:00 и 4:00, что соответствует конечным точкам первой и второй частей суток.

При вводе данных метеорологических наблюдений из палубных журналов двадцатого века ответственные лица должны подумать о том, как быть с судами, которые переводят свои часы в соответствии с часовым поясом территории, через которую они проходят. Этот сложный вопрос решается на этапе обработки данных; тем не менее, операторам необходимо знать, как вносить информацию там, где указывается часовой пояс, чтобы обеспечить эффективность обработки данных.

Если в судовом журнале отмечаются корректировки судовых часов или на судне фиксируются данные о применении часовых поясов, в шаблоне для ввода данных должно быть предусмотрено положение для записи этой информации. Операторы

должны записывать часовой пояс, указанный в судовом журнале (например, GMT, +3, -8 и т. д.), независимо от фактического положения судна, и должны убедиться, что каждый часовой пояс, отличный от GMT, указывается со знаком +/-, для того чтобы избежать неопределенности. Операторы ввода данных также должны записывать (в том же столбце часовых поясов или наборе ячеек), когда судовые часы переведены вперед или назад и на сколько. Имея эту информацию, можно внести все необходимые коррективы на этапе обработки данных.

Положение, из которого производятся наблюдения

Второй важный параметр, который необходимо указывать, — это положение судна. Это самый сложный и потенциально проблемный параметр, данные по которому необходимо ввести. Положение судна в море выражается в координатах широты и долготы в градусах и минутах, а также в виде буквы, обозначающей полушарие и указываемой перед ними или после них. Обратите внимание, что иногда широта и долгота также определялись посредством счисления. В прошлом стандартной практикой было ежедневно оценивать и фиксировать полуденные координаты, называемые «счислением пути», когда не было возможности определить более точное положение, а место судна, определяемое по наблюдениям, указывалось тогда, когда наблюдение за Солнцем в полдень было возможно. В отсутствие указания места судна, определяемого по наблюдениям, следует ввести данные счисления. При наличии обоих показателей вводятся данные по обоим. В других случаях положение судна между полуденными координатами можно определить на основании курса по компасу и пройденного расстояния согласно записям в судовых журналах.

Широта всегда указывает на расположение к северу или югу от экватора и обозначается буквами N или S перед координатами либо после них. Долгота не имела такого простого выражения до конца девятнадцатого — начала двадцатого века. В наши дни долгота выражается как восточная или западная от Гринвичского меридиана, а в научном плане — с помощью обозначения, согласно которому западная долгота представляет собой отрицательное число, а восточная — положительное. В исторических данных долгота всегда выражается как восточная или западная от меридиана, но этот меридиан может быть определен для любой точки Земли. Применительно к данным до 1850-х годов, а в некоторых случаях и гораздо позже, не следует полагать, что зафиксированная в них долгота привязана к Гринвичскому меридиану. Необходимо предусмотреть положение о внесении операторами данных об используемом меридиане. При вводе данных оператор должен обратить внимание на изменение меридиана, и в электронной таблице следует предусмотреть возможность указать название используемого меридиана. Поэтому предназначенные для ввода данные должны включать название меридиана (или название ориентира, если ориентир использовался в качестве точки отсчета), а не его фактическую долготу. Иногда в судовом журнале отсутствуют сведения о положении судна; так бывает во всех случаях, когда судно стоит в порту. В пределах видимости суши и, как правило, в прибрежных водах, знакомых капитану судна, широта и долгота также не записывались в судовую журнал просто потому, что в этом не было необходимости, что еще раз напоминает о том, что судовые журналы и реестры велись по соображениям практического характера, а не для записи научных данных.

5.3.3 Сбор гидрологических метаданных

Принимая решение о том, какие метаданные необходимо фиксировать, подумайте, кому может понадобиться доступ к данным, как в настоящее время, так и в будущем, и с какой целью (например, техническим специалистам может потребоваться информация о том, где именно установлены определенные приборы, гидрологи могут захотеть найти станции с качественными рядами наблюдений за паводками и т. д.).

Структурируйте данные иерархически, используя такие понятия, как река, водосборный бассейн, станция и временной ряд (измерение), обеспечивая возможность проследить все значения данных до станций и мест производства наблюдения.

Обеспечьте сохранение надежной пространственной информации, например, о местоположениях станций, включая вертикальные высоты относительно исходного уровня и подробную информацию о том, что это за уровень. Площади водосборных бассейнов часто необходимы для получения таких стандартных результатов, как сток.

По возможности сводите текстовые описания к ряду простых общих элементов. Например, вместо того чтобы хранить текстовые описания способов измерения и обработки данных с каждой станции, в качестве альтернативы можно перечислить общие для всех станций элементы — приборы, частоту измерений, измеряемые параметры, метод оцифровки, метод последующей обработки и т. д. Они могут быть описаны в ряде полей, что облегчит поиск и использование информации. При этом всегда желательно сохранять текстовое описание, если без него полезная информация будет потеряна.

Где это возможно, разработайте и используйте контролируемые перечни. Например, составьте перечень используемых приборов или конструкций гидрометрических станций. Сохраняйте любые полезные метаданные об элементах перечня или более подробное описание.

Подробно укажите измеряемые параметры (например, речного потока, температуры воды, температуры воздуха), точные единицы измерения, статистику измерений (мгновенные, средние, общие, максимальные значения и т. д.), дату начала и дату окончания записи данных.

Храните информацию о качестве измерений, где это возможно, и информацию о любых процессах, применяемых для получения результата, например, данные о стоке, полученные путем установления зависимости между расходом и уровнем на основе данных об уровне воды, измеренном вручную с помощью щита контроля уровня воды.

Соберите информацию о самом процессе оцифровки или спасения данных. В будущем может оказаться важным выяснить, вводились ли временные ряды вручную или оцифровывались с диаграммы и производилась ли оцифровка непосредственно с диаграммы с помощью планшета или с отсканированного изображения.

При использовании гидрологических данных особенно важно собрать информацию о качестве показаний приборов в рамках гидрологического режима. Это может быть просто классификация станций или фрагменты данных наблюдений как благоприятных или неблагоприятных для высокого или низкого стока, но может быть и более подробная информация, если она доступна.

По возможности давайте ссылки на другие источники. Например, при описании прибора, используемого для измерения данных, можно включить простую ссылку на отсканированную копию руководства к прибору.

Поскольку метаданные, вероятно, будут храниться отдельно от данных, например, в случае временных рядов речного стока, в метаданных должно быть указано место хранения данных. Если данные не были оцифрованы, должно быть описано физическое местонахождение бумажных носителей данных.

Основные важнейшие метаданные представлены в следующем перечне:

- широта и долгота по месту проведения измерений, а также географический исходный уровень для этих значений;
- наименование организации, которая собирала данные;
- время начала и окончания записи гидрологических данных;
- измеряемые параметры и единицы измерения;

- метод, использованный для измерения параметра (например, сток, рассчитанный по уровню воды путем установления зависимости между расходом и уровнем), и его точность, где это возможно (например, уровень воды, измеренный с помощью водомерной рейки с точностью до 0,5 см);
- временной интервал данных и применимый часовой пояс;
- информация о том, что представляет собой значение данных в каждом временном интервале (например, среднее значение в предыдущем интервале, общее значение в последующем интервале или мгновенное значение);
- способ оцифровки данных наблюдений;
- место хранения данных наблюдений в случае невозможности включить такие метаданные в данные наблюдений.

5.4 **Оцифровка информации о зависимости между расходом и уровнем воды и измерении стока**

Речной сток часто измеряется на открытых участках реки методом скорость-площадь, в рамках которого измерения проводятся на регулярной основе с целью определения фактического стока и его регистрации относительно уровня воды. Эти измерения могут быть использованы для установления зависимости между расходом и уровнем воды на данной станции. Часто могут возникать ситуации, когда зависимости уже были таким образом установлены и что данные об этих зависимостях и измерениях, на основе которых они были получены, также были сохранены. В таких случаях важно оцифровать всю историю измерений вместе с любыми комментариями полевых технических специалистов и гидрометристов, которые занимались их проведением и анализом. Важно также оцифровать кривые расхода и уравнения и хранить эту информацию вместе с оцифрованными данными диаграмм.

Измерения часто состоят из записанных вручную данных по каждому замеру расхода воды. Это могут быть просто дата и время измерения, измеренный расход и уровень воды, но нередко они включают также информацию о скорости и площади поперечного сечения, используемую для расчета общего расхода, а также комментарии относительно условий на участке во время измерения. Информация об уровне и расходе воды представляет собой наиболее важные данные для установления и анализа зависимостей между расходом и уровнем и должна быть достаточной для этой цели. Более подробная информация должна оставаться доступной, так как она может пригодиться в случае необходимости дополнительной оценки каких-либо отдельных измерений. Для некоторых измерений, обычно тех, которые фиксируют очень высокий или очень низкий сток, подробная информация может быть весьма полезной при оценке качества измерения расхода и необходимости его учета при установлении зависимости между расходом и уровнем.

При наличии измерений, и зависимостей между расходом и уровнем, возможно, имеет смысл пересмотреть оценку качества зависимости. Если доступны только зависимости между расходом и уровнем, вероятно, придется принять их на веру, но принимая решение о целесообразности этого следует использовать любой имеющийся местный опыт. Ряды стока, полученные на основе зависимостей между расходом и уровнем, необходимо тщательно проверять, чтобы убедиться в отсутствии явных сдвигов, вызванных изменениями в зависимостях.

6. **АРХИВАЦИЯ ЦИФРОВЫХ НОСИТЕЛЕЙ (ИЗОБРАЖЕНИЙ И ЗНАЧЕНИЙ ДАННЫХ)**

6.1 **Перекрестная проверка печатных носителей, изображений и цифровых наборов данных**

В процессе работы полезно фиксировать ход осуществления проекта. Этого можно достигнуть путем перекрестной проверки количества месяцев, за которые имеются изображения и цифровые данные, относительно ожидаемого количества, определяемого исходя из количества обработанных страниц и страниц, подлежащих обработке, и относительно периода наблюдений по каждой станции. В начале процесса спасения данных рекомендуется составить электронный реестр массивов данных на основе исходных данных на бумажных носителях/микрофильмах/других старых носителях. В дальнейшем, в конце процесса, количество созданных цифровых данных (согласно реестрам изображений и оцифрованных данных) может быть сопоставлено с исходным электронным реестром массивов данных, чтобы выявить любые недостающие данные.

6.2 **Обновление носителей**

Спасение данных имеет такое же значение для цифровых данных, как и для данных на бумажных носителях, поскольку никакие носители, используемые для хранения обоих типов данных, не могут храниться вечно. Например, магнитные ленты со временем теряют свои свойства, особенно в условиях высокой температуры или повышенной влажности. Бумага ветшает столетиями, а магнитные ленты и другие машиносчитываемые носители становятся нечитаемыми за считанные десятилетия. Магнитные ленты, оптические диски и кассеты 8 мм в общем случае морально устарели, и устройства для чтения этих носителей могут быть недоступны. Если цифровые данные все еще хранятся на этих носителях, следует обратиться в национальные архивы или университеты по месту их нахождения и рекомендовать им перенести эти данные на современные цифровые носители, обеспечивающие возможность долгосрочного хранения. Не менее важно сохранить описания форматов и программное обеспечение для чтения/декодирования данных. В противном случае вместо данных можно получить непонятный набор цифр и символов.

По мере развития вычислительной техники компьютеры, компьютерные операционные системы, языки программирования и программное обеспечение, используемые для чтения данных со старых носителей, также устаревают. Проблема морального износа будет сохраняться. В будущем настоятельно рекомендуется обеспечить миграцию обработанных графических и цифровых климатических данных на последние зарекомендовавшие себя носители (включая облачные хранилища, жесткие диски и т. д.). До недавнего времени широко использовались CD и DVD, но и они могут выйти из употребления в ближайшем будущем. Поскольку компьютеры (а также цифровые устройства хранения) периодически выходят из строя без возможности восстановления, рекомендуется также наладить процесс резервного копирования данных.

Следует подчеркнуть необходимость обеспечить миграцию программного обеспечения для записи, хранения, чтения и интерпретации рассматриваемых климатических данных со старых компьютеров и компьютерных технологий и его обновление. Если делать это достаточно регулярно, то спасение данных будет обеспечиваться до появления проблемы и их утраты. Решение о миграции носителей следует пересматривать с примерной периодичностью от двух до пяти лет, а решение об обновлении компьютеров и программного обеспечения — каждые пять—семь лет.

Наконец, предпочтительными для хранения изображений считаются форматы JPEG, TIFF и PNG, а формат ASCII может рассматриваться как предпочтительный для текстовых файлов данных с точки зрения удобства чтения цифровых данных и переноса данных с одного компьютера на другой. При использовании баз данных важно, чтобы данные могли легко экспортироваться в формат ASCII (например, файл .csv). Для целей миграции также важно обеспечить документальную основу цифрового кода, в котором хранятся файлы, и простоту его использования (т. е. чтобы он не требовал расшифровки).

6.3 Рекомендации по архивации цифровых носителей

- a) Защитите цифровые графические файлы и цифровые файлы данных. Цифровые графические файлы могут рассматриваться как официальные документы и поэтому должны быть защищены.
 - b) Упорядочите данные для удобства доступа к ним в будущем, организовав поиск по цифровому устройству хранения или по структуре каталога (станция, год, месяц). Следует сохранять реестр файлов на каждом цифровом устройстве хранения и в основном каталоге.
 - c) Обеспечьте ежедневное резервное копирование СУКД.
 - d) Храните цифровые изображения и данные на соответствующих носителях, чтобы данные оставались доступными в течение неопределенно долгого времени, обеспечивая миграцию наборов данных по мере необходимости для снижения рисков морального износа и разрушения.
 - e) Организуйте удаленное хранение цифровых носителей для обеспечения сохранности электронных изображений и цифровых данных. Для защиты данных в случае антропогенной катастрофы или стихийного бедствия следует хранить копии изображений и данных в нескольких местах, в том числе в многонациональном Региональном климатическом центре за пределами собственного региона.
 - f) Следует четко описать метод, с помощью которого будущие пользователи смогут получить доступ к сохраненным данным.
 - g) Разработайте долгосрочный институциональный план управления данными и архивации и обеспечьте поддержку его осуществления.
-

ПРИЛОЖЕНИЕ 1. ГЛОБАЛЬНАЯ ДЕЯТЕЛЬНОСТЬ

1. КТО ДОЛЖЕН ЗАНИМАТЬСЯ СПАСЕНИЕМ ДАННЫХ?

Любая группа или отдельное лицо, в распоряжении которых находятся данные на бумажных носителях, микрофильмах/старых носителях или цифровые данные, должны пытаться содействовать спасению данных об окружающей среде. Лица, которые несут ответственность за управление регистрацией национальных данных об окружающей среде, должны играть особую роль в области спасения данных, поскольку они в большей мере способны понимать значение и оценивать данные, которые они спасают, и понимать, какие из них являются наиболее важными. Такие специалисты во многих случаях состоят в государственных учреждениях, включая НМГС, для которых спасение данных является одним из элементов их обязанностей по хранению, а также в университетах и научно-исследовательских институтах. При этом сторонников спасения данных можно обнаружить во многих других учреждениях, как в государственных (сельскохозяйственные департаменты и пр.), так и в частных (сельские муниципалитеты, агробизнес и пр.). Кроме того, несколько добровольческих организаций, таких как Международная организация по спасению данных об окружающей среде (МОСДО) и инициатива «Модели циркуляции атмосферы Земли» (АКРЕ) были созданы для содействия спасению данных. Другие лица также могут быть заинтересованы в оказании помощи деятельности по спасению данных и могут сыграть важную роль, в том числе в привлечении широкой общественности к решению этой проблемы.

2. ГДЕ ДОЛЖНЫ ПРОВОДИТЬСЯ МЕРОПРИЯТИЯ ПО СПАСЕНИЮ ДАННЫХ?

Существует несколько подходов к выбору места для спасения данных. Одним из них может быть выбор места в стране, где хранятся данные, требующие спасения. Это может быть центральное учреждение НМГС и/или различные провинциальные отделения. Наилучшим подходом может быть сбор данных, требующих спасения, в одном месте или, если в каждом регионе имеется или может быть накоплен соответствующий опыт, обработка изображений и обеспечение их сохранности в местных отделениях. Другой возможностью является работа в региональных, многонациональных климатических центрах, таких как региональные климатические центры ВМО. Преимущество в этом случае заключается в том, что эти отделения могут располагать опытом и оборудованием в объемах, возможно, превышающих имеющиеся в отдельных странах. Альтернативный подход заключается в том, что третья сторона выполняет обработку изображений или оцифровку за пределами страны или региона, а все документы возвращаются в страну, из которой поступают данные. Оптимальный вариант будет зависеть от конкретных обстоятельств. Прежде чем данные будут переданы третьей стороне, Секретариат ВМО будет рад внести свой вклад в планирование места проведения мероприятий по спасению данных.

Решение о том, использовать ли собственные силы или внешний подряд, будет зависеть от проекта: его продолжительности, количества сканируемых изображений, необходимого и имеющегося персонала, наличия финансовых ресурсов и т. д. Помимо более общих и стратегических плюсов и минусов внешнего подряда важно учитывать риски, связанные с перемещением бумажных документов (изъятием из архива, перемещением через здание и в транспортное средство, условиями хранения в транспортном средстве и на территории компании/организации, осуществляющей обработку изображений, возвращением в архив).

3. ПОРТАЛ C3S ПО СПАСЕНИЮ ДАННЫХ

Портал C3S по спасению данных является сетевым ресурсом для всех интересующихся вопросами сохранности, спасения и оцифровки данных. Он обеспечивает единую точку входа для получения информации о состоянии прошлых и текущих проектов по спасению данных по всему миру; о данных, которые нуждаются в спасении, и об используемых методах и технологиях. Он представляет собой канал для обмена информацией по всем аспектам спасения данных, включая применяемые и перспективные технологии спасения данных. Такая информация также применяется в других научных дисциплинах.

Поскольку его целями является повышение уровня общественного восприятия существующих видов деятельности по спасению данных и стимулирование развития ее новых направлений, а также улучшение координации международных усилий по спасению данных, портал C3S по спасению данных представляет собой полезное средство коммуникации. Каждый представленный на портале проект позволяет лицам, имеющим схожие интересы, узнать об отличной работе, проделанной каждой страной. Портал C3S по спасению данных уделяет особое внимание информированию сообщества о прошлых и существующих мероприятиях по спасению данных, что также поможет выявить пробелы и возможности, определить приоритеты в области спасения данных в регионах, где эта деятельность наиболее востребована, и привлекать финансовые ресурсы для реализации проектов.

ПРИЛОЖЕНИЕ 2. НАИБОЛЕЕ РАСПРОСТРАНЕННЫЕ ПРИМЕРЫ ТИПОВ ДОКУМЕНТОВ, КОТОРЫЕ МОГУТ СОДЕРЖАТЬ МЕТЕОРОЛОГИЧЕСКИЕ, МОРСКИЕ И ГИДРОЛОГИЧЕСКИЕ ДАННЫЕ (И МЕТАДААННЫЕ)

Рукописные источники — наземные данные

- Метеорологические реестры (в переплете)
- Метеорологические формы, иногда называемые формами данных рельефа (отдельные листы)
- Формы станций метаданных
- Записные книжки с черновиками и набросками
- Барограммы
- Термограммы
- Тефиграммы
- Плювиограммы (диаграммы регистрации осадков)
- Автографические диаграммы уровня воды
- Суточные ленты регистрации показаний водомерного поста
- Архивы данных гидрометрических наблюдений
- Гигрограммы
- Ленты данных о продолжительности светового дня
- Синоптические карты
- Данные метеозонда
- Реестры показаний анемометров
- Гидрологические реестры
- Данные наблюдений за льдом и высотой снежного покрова
- Реестры данных маяков
- Отчеты об экспедициях
- Дневники наблюдений за погодой
- Официальные отчеты о стихийных бедствиях
- Газеты (рассматриваются как основной источник)
- Церковные записи
- Миссионерские записи
- Консульские записи
- Колониальные записи и правительственные бюллетени
- Страховые записи
- Сельскохозяйственные записи
- Данные фенологических наблюдений
- Данные показаний пиргелиометра
- Метеорологические и гидрологические сборники (например, ежемесячные бюллетени, ежегодные отчеты, монографии, отчеты обсерваторий)

Рукописные источники — морские данные

- Судовые журналы
- Морские метеорологические судовые журналы (также называемые книгами погоды, метеорологическими реестрами или метеорологическими журналами и часто включающие форму метаданных)
- Морские метеорологические формы
- Выдержки из записей судовых журналов
- Реестры синхронизированных метеорологических наблюдений
- Барограммы
- Термограммы
- Тефиграммы
- Синоптические карты
- Записные книжки
- Отчеты о морских экспедициях
- Отчеты о кампаниях на исследовательских судах

- Отчеты и журналы по прокладке морских кабелей
- Книги учета китобойного и рыболовного промыслов
- Дневные отчеты по китобойному и рыболовному промыслам
- Дневники, полевые заметки и отчеты инспекторов по китобойному и рыболовному промыслам
- Другие различные записи, связанные с регулированием китобойного и рыболовного промыслов (Международная китобойная комиссия и ее предшественники)
- Данные морского страхования
- Гидрографические и геодезические отчеты
- Отчеты о дноуглубительных работах
- Записи о приливах и отливах
- Доклады об океанических течениях
- Сообщения об ураганах и штормах
- Сообщения о льдах
- Ледовые карты
- Журналы и сводки погоды по морскому радио
- Основные регистрационные записи порта
- Судовые газеты

Прочие источники

Множество данных об окружающей среде можно найти в печатных изданиях. Источники варьируются от отчетов об экспедициях до ежегодных сводок об окружающей среде и статей в научных и географических журналах и других изданиях. Содержание может варьироваться от табличных данных до текстовых записей. Печатные источники, такие как отчет о путешествии или экспедиции, иногда могут содержать полезные метаданные об инструментах или методах наблюдения. Печатные книги и статьи часто могут служить свидетельством наличия наблюдений, а иногда и указывать, где их можно найти; сноски и библиографии в этих источниках часто могут быть весьма информативными. Однако печатные источники не могут заменить исходные документы, поскольку часто представляют собой отредактированные версии исходного материала.

ПРИЛОЖЕНИЕ 3. ОЦЕНКА СПАСЕНИЯ ДАННЫХ

Таблица 3 содержит общее руководство для лиц, которые занимаются оценкой спасения данных в рамках новых программ спасения данных.

Таблица 3. Контрольный перечень для оценки спасения данных

1. Изучить портал C3S по спасению данных на предмет технических ресурсов и передовой практики, связанных со спасением данных
2. Загрузить любые известные реестры данных из международных баз данных, библиотек, литературы для оценки известных массивов данных страны об окружающей среде
3. Определить текущее положение дел в области спасения данных в стране
4. Посетить и оценить существующие архивные помещения
5. Поговорить с директором и сотрудниками, чтобы узнать, в чем заключается значение деятельности по спасению данных для организации и каковы ее приоритеты в области спасения данных
6. Провести консультации по вопросам разработки реестра первоисточников, данных и метаданных, если это необходимо
7. Провести консультации по вопросам хранения бумажных документов/старых носителей, необходимого пространства и принадлежностей, а также по вопросам труда, охраны здоровья и безопасности
8. Провести консультации по методам и оборудованию, необходимому для обработки изображений
9. Оценить возможности и инфраструктуру СУКД/СУГД, связанные с управлением данными (период, резервное копирование, содержание)
10. Изучить текущие процессы оцифровки и ОК, чтобы понять, как новые оцифрованные данные могут вписаться в рабочий процесс и процесс ОК и не нуждается ли текущий процесс оцифровки и ОК в пересмотре
11. Провести консультации по методам и оборудованию, необходимому для оцифровки
12. Провести консультации относительно реестра изображений и цифровых данных
13. Провести консультации по вопросам сбора метаданных, архивации и обработки изображений (см. Aguilar et al., 2003)
14. Провести консультации по вопросам ввода ключевых данных или переноса данных с диаграмм
15. Провести консультации по вопросам потенциального использования спасенных данных для демонстрации их ценности для доноров/НМГС
16. Провести консультации по персоналу, необходимому для выполнения этапов обработки изображений и оцифровки данных
17. Оценить сроки выполнения каждого компонента работ по спасению данных, исходя из масштаба задачи и числа сотрудников
18. Разработать план спасения данных на основе приоритетов НМГС
19. Добавить информацию о программе спасения данных на портал C3S по спасению данных

ПРИЛОЖЕНИЕ 4. ИНФРАСТРУКТУРА, ОБОРУДОВАНИЕ, ПРИНАДЛЕЖНОСТИ И ПЕРСОНАЛ

Во всех случаях хранение, обработка изображений и оцифровка должны осуществляться в безопасном, хорошо освещенном, беспыльном помещении, оборудованном системой климат-контроля (минимальные требования к климат-контролю включают систему кондиционирования и осушитель воздуха). Оборудование и принадлежности будут приобретаться в зависимости от потребностей каждой страны в области спасения данных и избранных для этого методов. Для составления подробного плана проекта по спасению данных можно воспользоваться таблицей 4.

Решающее значение для успешного осуществления проекта по спасению данных имеет надлежащий подбор персонала. Идеальным специалистом для поиска климатических данных и принятия решений по организации архива является климатолог, в то время как гидрологи и океанографы идеально подходят для оказания помощи с гидрологическими и морскими данными соответственно. Разрабатывать процедуры обработки изображений лучше всего тому, кто интересуется вопросами сохранности, возможно, лицу, увлекающемуся фотографией. Лучшими специалистами по обработке изображений и оцифровке становятся люди, которые отличаются аккуратностью и вниманием к деталям, поскольку для выполнения этих задач важны скорость и точность. Несколько человек могут обрабатывать изображения и вводить данные попеременно, чтобы разнообразить монотонную деятельность и устранить зависимость от одного человека. Для спасения данных диаграмм важна последовательность редактирования, и, возможно, лучше всего, если этим будет заниматься один человек на каждой метеорологической или гидрологической станции. По завершении редактирования задача оцифровки диаграмм может быть выполнена группой специалистов-практиков.

Как обработка изображений, так и оцифровка проходят более эффективно, если персонал знаком с компьютерными технологиями, и если работа выполняется в паре: один человек обрабатывает изображение, другой проверяет качество обработки, и двое вводят или переносят данные с одной и той же страницы при оцифровке (двойной ввод данных). Настоятельно рекомендуется ограничивать время выполнения работ по обработке изображений и оцифровке половиной дня или четырьмя часами, поскольку обе задачи требуют безраздельной концентрации.

Количество фотокамер и тип требуемой копировальной установки зависят от того, представлены ли данные в переплетенном виде или без переплета. Постраничную съемку рекомендуется выполнять двумя фотокамерами, одна из которых будет использоваться только в качестве запасной. Для размещения фотокамеры и освещения используется репроустановка. Однако для фотографирования данных в переплетенном томе используется штатив с двумя фотокамерами. Эта система позволяет специалисту по обработке изображений фотографировать две страницы одновременно, не повреждая переплет. Информация о штативах для фотокамер представлена в приложении 7 и на портале C3S по спасению данных.

Таблица 4. Соображения по планированию программ спасения данных

	<i>Организация</i>	<i>Обработка изображений</i>	<i>Оцифровка</i>
Инфраструктура	Хорошо освещенное помещение с системой климат-контроля и охраной	Помещение с достаточным количеством электрических розеток и освещением, а также с достаточным количеством столов для сортировки бумажных форм	Помещение с достаточным количеством электрических розеток и освещением

	<i>Организация</i>	<i>Обработка изображений</i>	<i>Оцифровка</i>
Оборудование	Бескислотные архивные коробки Крупногабаритные полки для хранения материалов большого формата (шириной 2 м, высотой 2,5 м) Стремянка	Две фотокамеры или более (характеристики и аксессуары см. в приложении 7) Репроустановка Компьютер Подставка для книг Стол Технические руководства	СУКД/СУГД, планшет для оцифровки (в соответствии с программным обеспечением для оцифровки диаграмм)
Принадлежности	Перчатки, спецодежда, защитные маски, тряпки для пыли	Цифровые устройства хранения/облачные хранилища USB-кабель или беспроводная связь (WIFI)	
Персонал, ответственный за спасение данных	Два человека для очистки документов, организации и разработки архива и создания реестров	Два человека для производства фотосъемки, проверки качества, хранения и обновления реестра	Два человека для оцифровки и обновления реестра
	Подготовка по вопросам обеспечения безопасности и гигиены труда при работе с загрязненными данными		
Персонал для последующей деятельности	Квалифицированный специалист для обновления реестров и контроля состояния помещений	Климатолог/гидролог/морской эксперт/эксперт по вопросам окружающей среды для контроля за поступлением новых данных и метаданных	Климатолог/гидролог/морской эксперт/эксперт по вопросам окружающей среды для контроля качества и разработки продукции

ПРИЛОЖЕНИЕ 5. РУКОВОДЯЩИЕ УКАЗАНИЯ ПО СОДЕРЖАНИЮ БУМАЖНЫХ АРХИВОВ

Бумажные архивы должны храниться в специально отведенном помещении, желательно не в подвале и не на чердаке. Помещение следует содержать в чистоте, не допуская появления пыли, грязи, плесени и насекомых. Чтобы документы не лежали на полу, следует использовать прочные стеллажи (предпочтительно металлические полки) и бескислотные архивные коробки. Помещение должно быть хорошо освещено, а коробки промаркированы и легко доступны. Влажность и температура в идеале должны регулироваться для поддержания благоприятного микроклимата и обеспечения надлежащих условий хранения. Специально назначенный сотрудник должен регулярно проверять состояние архивного помещения.

Бумажные документы необходимо хранить хотя бы до тех пор, пока а) не будет закончена обработка изображений; б) не будет проведена проверка качества и полноты всех изображений и с) не будут сделаны дубликаты всех изображений. Важно сделать несколько электронных копий реестров и всего архива изображений и обеспечить их удаленное хранение в нескольких разных местах.

Бумажные копии всегда предпочтительно сохранять по ряду причин:

- а) бумажные копии могут содержать указания на ошибки измерений или приборов (пометки карандашом или ручкой), которые не читаются на электронных копиях;
- б) оригиналы с исправлениями и примечаниями важны для проверки значений;
- с) электронные копии могут быть утеряны из-за сбоя в работе компьютера;
- д) до настоящего времени бумажные копии сохранялись дольше, чем цифровые и электронные носители.

Представляется разумным содержание в качестве запасного варианта бумажного архива.

В отсутствие нехватки помещений бумажный архив должен оставаться в НМГС. Если ресурсы не позволяют обеспечить длительное хранение бумажных архивов, следует обратиться в национальные или университетские архивы и библиотеки. Национальные архивы страны могут устанавливать юридически обязательные стандарты в отношении длительности хранения данных (как в бумажном, так и в электронном виде). При отсутствии интереса к хранению бумажного архива со стороны этих национальных учреждений следует обратиться в международные архивы.

ПРИЛОЖЕНИЕ 6. ОПРЕДЕЛЕНИЕ ПРИОРИТЕТОВ ПРИ ОБРАБОТКЕ ИЗОБРАЖЕНИЙ И ОЦИФРОВКЕ ДАННЫХ

Хотя идеально было бы сразу обработать все изображения и ввести все данные, реальность такова, что процесс архивации занимает время и должен с чего-то начинаться. Начать можно с заполнения пробелов в существующих временных рядах. На Международном совещании по вопросам спасения данных (Женева, 11—13 сентября 2001 г.) было установлено, что первостепенную важность при спасении как исторических, так и текущих климатических данных должны иметь следующие аспекты данных:

- a) высокое качество;
- b) большое национальное, региональное или международное значение и угроза утраты;
- c) способность заполнить пробелы в существующих наборах данных.

1. ОБРАБОТКА ИЗОБРАЖЕНИЙ

При обработке изображений, полученных с бумажных документов, представляется логичным начинать с самых старых, редких или хрупких документов по каждой станции в отдельности. Однако поскольку качество бумаги и чернил в последние годы снизилось, более новые документы могут потребовать более срочной обработки. При определении приоритетов обработки диаграмм можно начать с диаграмм, имеющих наибольшее значение для НМГС (возможно, диаграмм осадков, давления, ветра или солнечной активности). Чтобы удостовериться в спасении всех имеющихся метаданных (включая описание и фотографии местности и данных, информацию о приборах, калибровки, практику наблюдений и процедуры контроля качества), необходимо обработать изображения всех страниц книги (например, ежемесячного реестра).

2. ОЦИФРОВКА

Поскольку оцифровка данных занимает много времени, важно правильно расставить приоритеты. При оцифровке данных приземных наблюдений следует учитывать следующие факторы:

- a) чтобы представить потенциальным спонсорам доказательство концепции, на начальном этапе может оказаться полезным выбрать станцию с наиболее продолжительным непрерывным учетом данных или сосредоточиться на необычном годе, продемонстрировавшем особенно актуальные явления окружающей среды (такие как засуха или наводнение);
- b) возможно, было бы разумно сначала ввести данные тех станций, по которым уже имеются частично оцифрованные данные, которые все еще действуют или имеют наиболее продолжительный период учета данных;
- c) в мире имеется меньше данных за период ранее 1950-х годов. При этом в отдельных странах любые данные до 1990 года могут иметь решающее значение;
- d) также необходимо учитывать фактор местонахождения. Может быть полезно выбрать станции, представляющие различные климатические регионы страны или районы, имеющие большое социально-экономическое значение;

- e) еще одним соображением может быть доступность данных. Первой может быть выбрана станция с наиболее достоверными данными и наименьшим объемом отсутствующих данных;
- f) однако соседние станции, даже с небольшим периодом наблюдений, имеют значение для подтверждения измерений;
- g) для текущих или исторических данных временное разрешение данных (ежемесячное, ежедневное, менее суток) или потребность в конкретных параметрах могут определять приоритеты в отношении оцифровки данных;
- h) например, если желательно получить климатологию суточных изменений температуры, осадков или ветра, то решающее значение имеют данные диаграмм или данные почасовых наблюдений (наблюдений за периоды, составляющие менее суток);
- i) для изучения в рамках спасения климатических данных важнейших климатических переменных выберите метеорологические переменные, которые можно перевести в месячные значения, такие как температура и осадки (Bojinski et al., 2014);
- j) для районов, где существует пространственный пробел в данных, могут быть выбраны отдельные станции;
- k) для климатологии гроз, града, торнадо, пыльных бурь, тумана и ледяных дождей требуются неинструментальные визуальные метеорологические наблюдения.

На предмет обработки изображений и оцифровки обычно рассматриваются приземные наблюдения, но аэрологическое зондирование, приземные и аэрологические диаграммы и морские наблюдения также имеют большое значение.

- a) Отдельные результаты зондирования обеспечивают ценную информацию об изменениях температуры, давления и влажности по мере изменения высоты;
- b) поскольку отдельные данные зондирования могут быть утеряны, такая информация может быть получена только из аэрологических диаграмм;
- c) приземные диаграммы также могут иметь весьма важное значение, поскольку после нанесения на карту данных наземных станций исходные формы иногда терялись.

Таковы общие руководящие принципы и соображения по определению приоритетов в области спасения данных. Каждая страна или центр должны определить свои собственные приоритеты исходя из основных потребностей страны и состояния климатических данных. После выбора станции и года нужно опробовать процедуры обработки изображений и оцифровки данных. Это поможет определить кадровые потребности и способы рационализации процесса.

Последнее, о чем следует напомнить в отношении оцифровки форм: рекомендуется оцифровывать сразу все имеющиеся на странице данные, включая метаданные (поскольку другая возможность вернуться к этой странице и внести пропущенные данные едва ли представится). Лучше всего это можно сделать, создав шаблон для оцифровки различных типов форм. Во всех случаях метаданные, описание данных и практики наблюдений (имя наблюдателя, время наблюдения), положение (широта, долгота) и характеристики местности подлежат спасению и оцифровке вместе с изображениями и цифровыми значениями.

3. СПЕЦИФИКА СПАСЕНИЯ ГИДРОЛОГИЧЕСКИХ ДАННЫХ

Определение приоритетности речных гидрометрических станций

При спасении данных гидрометрических станций, на которых производятся замеры параметров поверхностных вод, очень важно обеспечить отбор данных по наиболее значимым станциям. Определение приоритетов и выбор гидрологических станций связаны с определенными трудностями, поскольку полученные данные будут использоваться в различных целях, для достижения которых может потребоваться информация с различных типов станций. Приоритеты, вероятно, будут значительно различаться в зависимости от местных условий и будут определяться на основе понимания текущей доступности данных и наиболее насущных потребностей в дополнительной информации.

Ниже приведен список критериев, которые следует учитывать при выборе станций для спасения данных, в порядке их приоритетности (начиная с наиболее важных):

- 1) продолжительность регистрации данных наблюдений;
- 2) национальное оперативное значение реки/водотока (с точки зрения водных ресурсов, риска паводков и т. д.);
- 3) способность предоставлять информацию о гидрологических изменениях (например, о режиме естественного стока);
- 4) способность предоставлять информацию для улучшения понимания гидрологии, например, в отношении репрезентативности гидрологического режима в стране, включая способы оценки этой репрезентативности;
- 5) воспринимаемое качество (данных) гидрометрической станции;
- 6) воспринимаемый будущий потенциал в отношении пп. 2, 3 и 4 выше.

Очевидно, что эти критерии не являются полностью независимыми друг от друга; например, данные очень низкого качества в пункте 5 могут не иметь большого значения, каковы бы ни были продолжительность наблюдений или значение реки.

Определение приоритетности данных наблюдений за поверхностными водами

На той или иной конкретной гидрометрической станции могут храниться разнотипные данные наблюдений, относящиеся к речному стоку, и в условиях ограниченности ресурсов снова возникнет необходимость определить приоритетность того, что нужно спасать.

В приведенном ниже списке представлен приоритетный порядок отбора данных наблюдений (начиная с наиболее важных):

- 1) обработанные данные о речном стоке (например, среднесуточный сток);
- 2) обобщенные данные о речном стоке (например, среднемесечный сток и годовые максимумы) и предыдущие сводки и анализы гидрологических данных;
- 3) ключевые метаданные гидрометрических станций и дополнительные метаданные, касающиеся качества данных, такие как описание существующих гидротехнических сооружений, если таковые имеются, и основных параметров водосборного бассейна, которые облегчают интерпретацию данных;
- 4) рекордные показатели уровня воды;
- 5) измерения мгновенного расхода;

- 6) зависимость между уровнем и расходом;
- 7) поперечные сечения реки;
- 8) другие метаданные гидрометрической станции, например, обследования объекта и фотографии.

Эти приоритеты могут меняться в зависимости от конкретных местных проблем, типа станции и качества документов.

Определение приоритетности прочих данных наблюдений

Спасение других наборов данных будет зависеть от их доступности и от других аспектов, таких как качество данных наблюдений. При этом следует принять во внимание следующие моменты:

- 1) осадки: нередко данные широко доступны, поэтому спасение данных о количестве осадков за длительный период наблюдений имеет большое значение. В случае нехватки данных наблюдений о количестве осадков спасение данных становится приоритетной задачей. Подтверждающие данные наблюдений в районах или периодах со слабым охватом данными также имеют большую ценность;
 - 2) снег: данные наблюдений о количестве выпавшего снега и высоте снежного покрова имеют решающее значение в некоторых регионах, и зачастую их очень трудно найти;
 - 3) испарение: измерения эффективного испарения могут оказаться чрезвычайно полезными для будущих гидрологических исследований;
 - 4) автоматическая метеорологическая станция (АМС): полный и качественный набор метеорологических данных с АМС может позволить оценить потенциальное испарение и будет считаться полезным, если подобные наборы данных не будут широко доступны;
 - 5) метаданные метеорологических станций: дополнительная информация о широко используемых метеорологических станциях позволяет получить более четкое представление о качестве данных. Эти метаданные могут включать точное местоположение и любые изменения с течением времени, период наблюдений и измеряемые переменные, а также информацию о размещении объекта, типах используемых приборов, процедурах считывания показаний приборов и т. д.;
 - 6) тематические данные: карты или пространственные координаты позволяют лучше понять гидрологические процессы и поэтому важны для производства гидрологических исследований. Тематические данные могут включать карты почв, геологические карты, информацию о земном покрове или даже картографические карты с указанием высоты над уровнем моря и расположением рек. В НМГС могут храниться такие карты, которые уже недоступны в других местах, и они должны иметь высокий приоритет при сохранении данных.
-

ПРИЛОЖЕНИЕ 7. МЕТОДЫ ЭЛЕКТРОННОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

Создание электронного изображения весьма желательно, поскольку сокращает объем работ с бумажными или иными старыми носителями. Это также позволяет изменить размер и улучшить качество изображения, чтобы сделать надписи более разборчивыми при вводе данных. Как правило, подходят изображения в формате JPEG, а форматы изображения без сжатия (например, несжатые изображения TIFF) имеют то преимущество, что их можно хотя бы частично прочесть, даже если некоторые данные повреждены. В зависимости от размера исходного документа согласно некоторым источникам размер изображения 1600 на 1200 пикселей является достаточным.

1. СКАНЕРЫ

Сканеры часто рассматриваются как средство обработки изображений с бумажных документов. Два основных вида сканеров включают планшетные (часто используются для книг) и протяжные сканеры (часто используются для отдельных страниц, например, диаграмм). Сканеры представляют собой привлекательную альтернативу фотокамерам, поскольку они могут уменьшить количество оборудования, необходимого для получения изображений. Однако метод сканирования имеет ряд недостатков:

- a) планшетный сканер может повреждать книжные переплеты, а иногда стать причиной того, что данные, расположенные ближе к переплету, окажутся не в фокусе или не будут отображаться. Этого можно избежать при использовании фотокамеры. Следует отметить, что существуют сканеры, специально предназначенные для книг;
- b) при подаче документа на сканер через устройство автоматической подачи существует риск того, что отдельные листы будут смяты или даже уничтожены при попадании в сканер (замятие бумаги);
- c) микрофильмы и микрофиши требуют применения специальных сканеров;
- d) сканеры имеют много подвижных деталей и обычно требуют обслуживания и наличия лицензии;
- e) для работы с более сложными сканирующими устройствами часто требуется специальная подготовка. Кадровые изменения, вызванные вопросом подготовки, привели к остановке сканирования в рамках некоторых проектов. В других случаях вопросы недостаточной подготовки и лицензирования стали препятствием для использования сканеров при обработке изображений;
- f) за исключением очень дорогих систем, сканеры, как правило, работают медленно и занимают много места.

2. ФОТОКАМЕРЫ

Наиболее популярным средством создания цифровых изображений являются фотокамеры. Они хорошо справляются как с переплетенными материалами, так и с отдельными бумажными листами, равно как и с такими негабаритными документами, как карты. Цифровые фотокамеры стали обычным явлением, и даже автоматические фотокамеры могут обеспечить получение надежных изображений, особенно при наличии постоянного источника хорошего освещения. При наличии стационарного штатива и системы освещения, пульта дистанционного управления для съемки и адаптера переменного тока (AC) все, что приходится делать, это менять страницы, что обеспечивает

быстроту и эффективность процесса. Кроме того, в настоящее время доступно недорогое программное обеспечение, позволяющее переименовывать цифровые изображения и переносить изображения на компьютер. Процесс архивации изображений также происходит очень быстро.

3. **ХАРАКТЕРИСТИКИ ФОТОКАМЕРЫ И ОБОРУДОВАНИЕ РАБОЧЕГО МЕСТА ДЛЯ ОБРАБОТКИ ИЗОБРАЖЕНИЙ**

Оборудование, необходимое для организации рабочей станции для получения и обработки фотоизображений, приведено в таблице 5. Во-первых, требуется стол с компьютером и фотокамерой (включая трипод или штатив). Важнейшее значение имеет достаточное количество заземленных электрических розеток. Хотя для работы с изображениями идеально подходят естественное освещение или солнечный свет (необходимо следить за тем, чтобы свет был рассеянным, и избегать тени), помещение с естественным освещением не всегда бывает доступно. Таким образом, установка фотокамеры часто предполагает использование двух или четырех осветительных приборов. Трипод или штатив фотокамеры, установленные в увязке с осветительными приборами, обеспечивают ее неподвижное положение, что позволяет быстро сделать несколько снимков. Важно неоднократно проверять освещение, поскольку некоторые бумажные носители имеют глянцевую поверхность и могут частично засвечивать изображение, отражая свет обратно в объектив фотокамеры. Функция дистанционного управления позволяет делать снимки без присутствия возле фотокамеры фотографа, а также снижает вероятность дрожания ее корпуса. Компьютер, в котором хранятся цифровые изображения, имеет решающее значение.

Таблица 5. Основные требования к инфраструктуре использования фотокамеры

1	Стол
2.	Электрические розетки
3.	Штатив фотокамеры/трипод
4.	Осветительные приборы
5.	Фотокамера и аксессуары (включая запасную фотокамеру)
6.	Компьютер

Батареи фотокамер обычно разряжаются в довольно короткие сроки. Адаптер переменного тока позволяет делать неограниченное количество снимков и значительно экономит время и средства на покупку и замену батарей. Если с имеющейся фотокамерой нельзя использовать адаптер переменного тока, рекомендуется воспользоваться запасными перезаряжаемыми батареями и зарядным устройством. Важно отметить, что в отсутствие возможности использовать адаптер переменного тока и при необходимости воспользоваться батареями конструкция фотокамеры должна предусматривать возможность их замены таким образом, чтобы не снимать фотокамеру со штатива. Большинство цифровых фотокамер снабжены пластиковой резьбой для крепления на штатив. В результате многократного снятия фотокамеры со штатива для замены батареи эта резьба изнашивается.

Для автоматической передачи изображения с фотокамеры на компьютер требуется программное обеспечение и USB-кабель или доступ к WIFI. Если такое программное обеспечение позволяет задействовать затвор фотокамеры при помощи мыши, кабель дистанционного управления может не понадобиться. Если USB-порт или WIFI недоступны, следует использовать самую большую из доступных карт памяти. Цифровые фотокамеры часто включают программное обеспечение для создания имен файлов для каждого изображения и передачи фотографий непосредственно на компьютер. Такое программное обеспечение значительно экономит время фотографу. В зависимости от имеющейся фотокамеры полезно иметь увеличительную линзу и возможность изменять настройки фотокамеры. В таблице 6 приведена краткая информация о желательных аксессуарах для фотокамеры, обеспечивающих эффективность процесса съемки.

Таблица 6. Аксессуары для фотокамер, способствующие эффективному процессу съемки

1. Репроустановка, на которой размещаются фотокамера и от двух до четырех осветительных приборов
2. Адаптер переменного тока для питания фотокамеры
3. Пульт дистанционного управления для съемки или программное обеспечение для фотокамеры, как в пункте 4
4. Программное обеспечение для фотокамеры, позволяющее снимать и передавать фотографии непосредственно на компьютер через USB-кабель или WIFI
5. В качестве альтернативы можно использовать самую большую из доступных карт памяти
6. Программное обеспечение, которое позволяет переименовывать графические файлы
7. Увеличительная линза, при наличии таковой, и макролинза для съемки крупным планом
8. Аппаратное или программное обеспечение, которое позволяет изменять настройки фотокамеры (через фотокамеру или компьютер)
9. Безбликовое стекло для бережного распрямления документов с переплетом

Необходимо учитывать размер изображения, особенно в случае использования карты памяти. В зависимости от качества изображения, самое большое изображение, скорее всего, не потребуется, и среднего размера может оказаться достаточно. Установлено, что даже 4-мегапиксельной камеры хватает для создания файлов JPEG с разрешением, достаточным для получения читаемых буквенно-цифровых значений параметров. В настоящее время даже камеры мобильных телефонов имеют разрешение около 12 мегапикселей. Важно отметить, что чрезмерно высокое разрешение хотя и не требуется для восстановления буквенно-цифровых данных, может понадобиться для использования с программами автоматической компьютерной оцифровки. Как правило, речь идет о цветной, а не о черно-белой фотосъемке. Цвет облегчает чтение исправлений и других пометок, встречающихся в исходных записях.

Другие распространенные функции фотокамеры, на которые следует обратить внимание, — это возможность отключения вспышки, режим макросъемки для крупных планов, возможность регулировки светочувствительности фотопленки и возможность регулировки баланса белого. В таблице 7 приведены общие характеристики фотокамеры для обеспечения качества изображения.

Таблица 7. Общие характеристики фотокамеры для обеспечения качества изображения

1. Возможность выбора размера изображения
2. Возможность отключения вспышки
3. Режим макросъемки крупным планом
4. Возможность регулировки светочувствительности фотопленки
5. Возможность регулировки баланса белого

4. ПРАВИЛА И ТОНКОСТИ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

- Никогда не используйте вспышку при создании изображений.
- Всегда используйте штатив или трипод.

- Всегда используйте дистанционный переключатель для управления камерой. Это гораздо быстрее, чем нажимать кнопку спуска затвора, и позволяет избежать лишних движений камеры. Некоторые виды программного обеспечения для фотокамеры позволяют управлять камерой с помощью компьютера.
 - Убедитесь, что звук фотокамеры отключен, чтобы не мешать окружающим.
 - Обеспечьте достаточное и стабильное освещение и избегайте тени на документе.
 - Не занимайте документом весь кадр; оставьте по краям небольшой отступ.
 - При наличии переплета фотографируйте разворот при условии, что том можно уложить достаточно ровно. Размер и формат могут диктовать необходимость применения другого подхода.
 - Всегда фотографируйте документ целиком, включая обложку (на ней может быть указан архивный номер), титульный лист, любые страницы, содержащие метаданные прибора, инструкции по ведению наблюдений и заметки о наблюдениях.
 - Используйте цветную фотографию: нередко цвета имеют определенный смысл (например, температура, обозначенная красным цветом, может относиться к области отрицательных температурных значений, даже если знак «минус» не написан).
 - Обработайте изображения всех имеющихся страниц (сюда могут входить пустые страницы и обложки). Некоторые обозначения могут распространяться на другие страницы (например, иногда знаки «плюс»/«минус» указываются только при смене знака).
-

ПРИЛОЖЕНИЕ 8. ДРУГИЕ МЕТОДЫ ОЦИФРОВКИ

Ниже приводится общее описание методов оцифровки, отличных от ручного ввода данных отдельными лицами. Все методы требуют контроля качества, поскольку при любом воспроизведении данных возможны ошибки.

1. ОПТИЧЕСКОЕ РАСПОЗНАВАНИЕ СИМВОЛОВ, ИНТЕЛЛЕКТУАЛЬНОЕ РАСПОЗНАВАНИЕ СИМВОЛОВ И ИНТЕЛЛЕКТУАЛЬНОЕ РАСПОЗНАВАНИЕ СЛОВ

Оптическое распознавание символов — это общедоступная программная технология, используемая для оцифровки машинописного текста или печатных таблиц; однако в проектах, связанных с рукописными наблюдениями, она показала неоднозначный результат (Wilkinson et al., 2019; Lorrey et al., 2022). Интеллектуальное распознавание символов (ИРС) предназначено для считывания отдельных печатных рукописных символов и позволяет использовать различные шрифты и шаблоны печати; оно лучше всего подходит для считывания форм, имеющих четкую структуру. В свою очередь, интеллектуальное распознавание слов (ИРСл) позволяет извлекать целые рукописные слова, будь то написанные печатными буквами или прописью. Исторические формы, как правило, не соответствуют требованиям ИРС/ОРС и не пригодны для обработки с помощью этой современной технологии. ИРС и ОРС используют технологию машинного обучения для автоматического обновления баз данных распознавания новых образцов почерка.

На момент подготовки настоящей публикации ОРС, ИРС и ИРСл находились в стадии развития и, вероятно, со временем станут более надежными. В настоящее время чаще всего используется оцифровка посредством ввода данных, поскольку считается, что человеческий ум наиболее приспособлен для распознавания рукописных материалов.

2. КРАУДСОРСИНГ

Введение

Оцифровка методом краудсорсинга предполагает использование для обработки информации онлайн-добровольцев, которых часто называют «учеными-любителями».

Несмотря на то, что в этом случае ввод данных осуществляется добровольцами, это по-прежнему сопряжено с определенными расходами. Для разработки и поддержания инфраструктуры для ввода данных через Интернет, развития базы данных, контроля качества, разработки инструкций по обучению операторов/наблюдателей и информационно-просветительской работы для привлечения потенциальных добровольцев необходимы время, талант и финансовые ресурсы.

Основные преимущества использования сетевой гражданской науки для осуществления деятельности по спасению данных заключаются в том, что людские ресурсы можно привлекать со всего мира, время добровольцев бесплатно, а прогресс достигается более или менее непрерывно. Кроме того, спасение данных с помощью гражданской науки обеспечивает возможность вовлечения широкой общественности и информирования ее о важности долгосрочных метеорологических наблюдений и изменения климата. Оно также повышает прозрачность массивов данных в конкретных странах и архивах.

Приведенные ниже основные рекомендации и накопленный опыт в области краудсорсинга взяты из проекта гражданской науки Southern Weather Discovery (подробнее см. Lorrey et al., 2022).

Предварительное условие

Данные, подлежащие спасению, были идентифицированы, и имеются электронные изображения хорошего качества.

Общий подход: определение проекта и плана работы

Создайте внутреннюю проектную группу, разработайте идентификатор проекта гражданской науки, описание проекта и веб-сайт:

- рассмотрите два основных компонента веб-сайта: фронтальный компонент, видимый и доступный широкой общественности, и внутренний компонент, который содержит элементы дизайна и содержания, необходимые для организации рабочих процессов и создания полей ввода данных;
- придайте проекту человеческое измерение, показав, что науку двигают люди, и укажите основных контактных лиц;
- обеспечьте возможность для связи ученых-любителей с внутренней проектной группой по любым вопросам и налаживания более тесных контактов с гражданскими лицами, участвующими в преобразовании данных. Выделите достаточно времени на то, чтобы проектная группа могла отвечать на вопросы добровольцев, особенно на ранних этапах после запуска проекта;
- определите архитектуру базового плана работы, чтобы установить порядок разделения труда;
- проектная документация: убедитесь в том, что все относящиеся к проекту исходные данные наблюдений/журналы, их цифровые копии (изображения), использованные формы, версии программного обеспечения и кода, описания процессов и т. д. остаются в сохранности после окончания проекта.

Аспекты коммуникации

Популяризируйте проект, взаимодействуйте со средствами массовой информации и обеспечьте активное участие общественности в общем успехе проекта. Рассмотрите следующие моменты, связанные с коммуникацией по проекту:

- определите и реализуйте сильную коммуникационную стратегию с «ловушками» для привлечения внимания задействованных лиц;
- убедитесь, что проектная группа готова взаимодействовать со СМИ и участниками проекта;
- популяризируйте проект на различных платформах социальных сетей;
- поддерживайте постоянный контакт с сообществом ученых-любителей, используя электронную почту и обновляя информацию по мере осуществления задач;
- подготовьте многоэтапный план информационной работы для представления и популяризации проекта, включая видеоматериалы и кампании в печатных средствах массовой информации для привлечения участников и поддержания интереса;
- подумайте о том, какое из метеорологических явлений (например, сильный шторм, снегопад, волна тепла и т. д.) могло бы вызвать интерес для начала проекта.

Технические аспекты

- Рассмотрите специализированные рабочие процессы, включая рабочие процессы для тиражируемого ввода данных, чтобы обеспечить добровольцам возможность участвовать в преобразовании данных;
- рабочие процессы должны быть максимально простыми и задействовать фрагменты журнала, как это рассмотрено ниже, а не отображать страницу целиком;
- для обеспечения простоты выполнения рабочих процессов (с учетом обратной связи, установленной с участниками) необходимо проводить их тестирование и повторное тестирование;
- прежде чем приступать к полноценной кампании по спасению данных, проведите оценку первичного ввода данных и их поиска на небольшом наборе данных;
- заранее подготовьте достаточно материала, чтобы обеспечить постоянное поступательное движение.

Тиражируемый ввод сегментов журнала предназначен для обеспечения консенсуса большинства (и меры достоверности благодаря тиражированию), определяющего, какое численное значение стоит в каждой ячейке таблицы. По результатам тестирования можно найти компромисс между временем, затрачиваемым добровольцами из числа ученых-любителей на ввод данных, и достижением полноты и точности преобразованного набора данных при создании, например, трех, восьми или любого другого количества дублей данных. Стоит также отметить, что эти результаты зависят от характера воспроизводимых данных. Целые числа без десятичных знаков являются наиболее простыми для ввода и требуют небольшого количества дублей для обеспечения верности консенсусного значения. И наоборот, буквенно-цифровые значения и значения с большим количеством значимых цифр создают больше сложностей или возможностей для вариативности ответов ученых-любителей (например, обозначение десятичной дроби точкой, запятой, пробелом или полное пренебрежение десятичной дробью). В проекте Southern Weather Discovery использовалось восемь дублей данных для достижения консенсуса с минимальными погрешностями.

Фрагменты журнала: отдельные страницы метеорологического реестра могут быть разделены на небольшие части, чтобы отдельный доброволец имел возможность преобразовать фрагмент, а не целую страницу. Такой подход помогает i) обеспечить ввод данных короткими сериями, а не в течение длительных промежутков времени, ii) свести к минимуму риск того, что добровольцы прекратят заполнять форму ввода данных прежде, чем данные будут воспроизведены полностью, iii) уменьшить количество ошибок, связанных с воспроизведением не того столбца или строки, и iv) снизить вероятность распространения наиболее частых ошибок по всему журналу, если, например, одному добровольцу трудно расшифровать почерк на конкретной странице.

Также можно интегрировать деятельность по спасению данных учеными-любителями в учебный процесс. Это дает учащимся многочисленные преимущества в обучении и одновременно увеличивает ресурсный потенциал проекта по спасению данных. Деятельность по спасению данных может быть включена в школьные мероприятия (например, в рамках партнерства между учреждением, занимающимся спасением данных, и средней школой) или в качестве аккредитованного проекта для учащихся высших учебных заведений (что также предполагает партнерство между преподавателями и специалистами-практиками по спасению данных). Как показывают опросы, проведенные в Ирландии в рамках недавних проектов в области гражданской науки, учащиеся узнали, как создаются данные, в чем их важность и как их можно использовать, и увидели пользу от своего вклада. Дополнительную информацию о том, как использовать спасение данных в учебном процессе, и руководство по порядку организации работы в классе можно найти в публикациях Ryan et al. (2018) и Mateus et al. (2021).

3. **ОЦИФРОВКА ДИАГРАММ**

Примерами данных диаграмм могут служить ленточные диаграммы и диаграммы самописца, на которых фиксируются непрерывные измерения таких метеорологических и гидрологических переменных, как температура, скорость и направление ветра, атмосферное давление, осадки и уровень воды. Оцифровка может включать в себя i) ручное преобразование значений диаграммы в табличную форму или ii) использование специализированного программного обеспечения для оцифровки.

Проводить оцифровку вручную очень утомительно, особенно если речь идет о резких колебаниях или одна диаграмма состоит из множества линий. Хотя на диаграмме могут отображаться данные с высоким временным разрешением, во многих случаях сообщаются только пиковые или суточные значения. В свою очередь, программное обеспечение для оцифровки диаграмм позволяет с большей легкостью получать более высокое временное разрешение значений данных в диапазоне от нескольких минут до нескольких часов.

Существует различное программное обеспечение для оцифровки, позволяющее квалифицированным специалистам вручную перенести линию диаграммы и преобразовать точки вдоль этой линии в цифровые значения, часто с использованием специальных планшетов для переноса данных (аппаратных средств). Одна из потенциальных проблем заключается в том, что это программное обеспечение может быть совместимо только с определенными компьютерами и операционными системами. Совместимость оцифрованных данных (типа файла) с соответствующими СУКД и СУГД имеет решающее значение и может ограничить выбор программного обеспечения для оцифровки. Хотя можно получить исходный код программного обеспечения для оцифровки, чтобы настроить его под конкретные форматы вывода данных, этот код может потребоваться поддерживать (обновлять по мере появления новых языков программирования и компьютерных технологий) и адаптировать специалисту по программному обеспечению.

Разработано и используется несколько программ для оцифровки. Функционал этих программ включает извлечение кривых на изображениях необработанных данных диаграмм, возможность чтения различных типов диаграмм и возможность обработки множества линий на одной диаграмме. Документацию по этому программному обеспечению можно найти на портале C3S по спасению данных. Некоторые из этих программ доступны для бесплатного использования при условии, что данные будут помещены в бесплатную и открытую базу данных общего доступа.

Преобразование линий диаграммы в цифровые значения — это только часть процесса оцифровки диаграмм. Другая часть известна как «редактирование» и представляет собой процесс контроля качества, необходимый для того, чтобы снизить возможность переноса ошибок со стороны человека и погрешностей оборудования в цифровой формат. Например, в ходе редактирования необходимо убедиться, что линия, отображаемая на бумажной диаграмме, расположена правильно относительно осей для времени начала и окончания, даты и величины измерения.

В процессе редактирования используется такая информация, как калибровочные проверки и другие аннотации для того, чтобы предоставить подробные сведения о необходимых исправлениях в тех случаях, когда предполагается, что линия диаграммы требует корректировки при ее преобразовании в цифровой формат. Например, если из-за ошибки со стороны человека линия диаграммы началась с неверной величины измерения, то при переносе линии потребуются внести корректировку в программное обеспечение для оцифровки, чтобы обеспечить точность цифровых значений. Поэтому установление процедур редактирования на начальном этапе проекта по спасению данных диаграмм имеет важнейшее значение для обеспечения согласованности между действиями специалистов-практиков и производством точных и достоверных цифровых данных (см., например, de Smeth et al. (2024)).

Последовательность редактирования и оцифровки определяется по усмотрению НМГС и может зависеть от временных ограничений и приоритетов в области спасения данных.

Редактирование обычно требует больше времени, чем оцифровка, и должно выполняться специалистами (например, гидрологами или гидрометристами). Из-за ограниченности времени и ресурсов НМГС может принять решение оцифровать необработанные данные диаграмм, а затем отредактировать только определенные гидрологические события в своей СУГД. Это может произойти, например, если целью спасения данных является помощь в понимании риска паводков. В качестве альтернативы НМГС может решить отредактировать набор данных (например, по календарным или гидрологическим годам), а затем оцифровывать их годовыми партиями.

Редактирование вводит дополнительные требования к ведению учета в рамках спасения данных диаграмм:

- 1) неотредактированные и отредактированные данные диаграмм должны сохраняться и быть доступными для пользователей СУКД/СУГД, с тем чтобы они имели возможность сверить исходные и произведенные записи данных наблюдений;
- 2) такая контекстуальная информация, как калибровочные проверки и аннотации, должна фиксироваться наряду с отредактированными данными;
- 3) решения по вопросам редактирования должны быть зафиксированы:
 - a) в СУКД/СУГД, где хранятся оцифрованные и отредактированные выходные данные, в виде аннотаций к произведенным записям данных наблюдений;
 - b) в виде подробного отчета о решениях, принимавшихся в отношении каждой диаграммы, для внутреннего пользования НМГС в справочно-информационных целях;
 - c) в виде агрегированной ежемесячной/ежегодной сводки для использования будущими пользователями данных для содействия пониманию исторических проблем с измерениями или оборудованием и истории обеспечения качества данных.

Эта информация является важным источником метаданных о качестве данных и должна храниться в том же месте, что и метаданные станции, и быть доступной для будущих пользователей данных. Эти метаданные в основном отражают проблемы измерений, выявленные в процессе спасения данных, и решения, применявшиеся для их устранения. Предоставление этой информации будущим пользователям данных позволяет им получить сведения о качестве измерений, которые могут иметь значение в контексте предполагаемого использования данных.

Решения по вопросам редактирования также должны отражаться в подходе к кодированию качества произведенных цифровых данных. С более подробной информацией можно ознакомиться в приложении 9.

ПРИЛОЖЕНИЕ 9. ОШИБКИ ПРИ ОЦИФРОВКЕ И ПРОЦЕСС ОБЕСПЕЧЕНИЯ КАЧЕСТВА

1. КОНТРОЛЬ КАЧЕСТВА НА ПРЕДМЕТ ОШИБОК ПРИ ВВОДЕ ДАННЫХ

Ниже приводится краткий обзор контроля качества применительно к ошибкам оцифровки, допускаемым, в частности, при вводе данных. Некоторые или все эти функции контроля качества могут быть доступны в системах управления данными (например, СУКД, СУГД и т. д.) учреждений Членов. В этом случае ошибки будут отмечаться по мере ввода данных в систему управления данными и впоследствии исправляться. При отсутствии этих функций контроля качества в рамках системы управления данными настоятельно рекомендуется рассмотреть возможность их включения. На начальном этапе контроль качества новых вводимых цифровых данных сосредоточен на ошибках ввода данных, приборов, наблюдателя и ошибках в метаданных, а не на более точном анализе, выполняемом в ходе обеспечения однородности, иногда называемом контролем качества в «режиме с задержкой». Вводимые данные следует рассматривать (и помечать) как предварительные до тех пор, пока они не пройдут проверку контроля качества.

Использование двойного ввода данных, шаблонов и ввода данных один в один сводит к минимуму вероятность ошибки при вводе отдельных значений, но не гарантирует безошибочного ввода данных. Проблемы, связанные с метаданными (касающиеся в основном названия и идентификационного номера станций), качеством отсканированных изображений, неисправностями приборов и нестандартной практикой наблюдений, по-прежнему сохраняются. Для обеспечения наиболее точного цифрового представления данных следует провести проверку контроля качества, чтобы пометить сомнительные значения как метаданных (например, в отношении названия или идентификационного номера станции, года или месяца, типа элемента), так и метеорологических переменных. Сомнительные значения помечаются для проведения проверки вручную, при этом добавляются флажки, указывающие на тип ошибки и результат проверки.

К распространенным ошибкам, допускаемым при вводе данных, относятся перестановка чисел, отсутствие или неправильное расположение десятичных знаков, пропущенные цифры (например, ноль), неправильные периоды аккумуляции и отсутствующие значения (соответствующие пункты рекомендуемых инструкций для специалистов по оцифровке перечислены ниже). Другие распространенные ошибки включают: ошибки наблюдателя (например, использование нестандартных единиц измерения); неисправность прибора; проблемы с метаданными (несоответствие в части идентификатора станции или соглашения о порядке именования, неверный год или месяц); перестановка или дублирование конкретного элемента применительно к данному месяцу и станции и неясности в формах. Эти ошибки могут указывать на проблемы с форматированием и системные проблемы с качеством данных.

Для обозначения причины каждой ошибки используются флажки. После завершения проверки помеченного флажком ошибочного значения специалист по оценке выбирает соответствующий тип проверки данных. Значения могут рассматриваться как: верные, соответствующие форме, разумные и не подлежащие изменению; помеченные как ошибка, но без изменения данных; ошибка, но с указанием заменяющего значения; не соответствующие исходной форме, исправленные для приведения в соответствие с исходной формой; ошибка с указанием на отсутствие данных, или нуждающиеся в вводе с последующим вводом данных. В общем случае, если помеченное значение неоднозначно и может оказаться верным, его не изменяют.

2. ПРОВЕРКА КОНТРОЛЯ КАЧЕСТВА

Проверки и процедуры контроля качества гарантируют пользователям, что оцифрованные данные точно представляют данные наблюдений, отраженные в исходных документах. Если проверки не проводятся автоматически с помощью системы управления данными, следует разработать сводную таблицу реестра контроля качества для отслеживания каждой проверки по каждой станции. Необходимо перепроверять вручную резко отклоняющиеся значения в рамках каждой проверки и вносить исправления по ее завершении. Данные сводной таблицы учета контроля качества обновляются после выполнения каждого этапа контроля качества в отношении станции.

Проверки контроля качества включают проверки диапазона (по суточным значениям, ежемесячным итоговым и среднемесячным значениям), проверки внутренней согласованности данных (сопоставление измерений между типами данных) и проверки экстремальных значений (по отдельным значениям). Однако иногда проблемы с качеством данных могут быть обнаружены только при добавлении выходных данных проекта по спасению данных к имеющимся данным по конкретной измерительной станции (если таковые имеются). В качестве примера можно привести данные об абсолютном уровне воды, которые могут быть правильно оцифрованы на основании необработанных данных на бумажном носителе или в формате диаграммы, но при этом отображаться слишком высоко или слишком низко из-за применения неправильного исходного уровня. В данном случае исходные уровни в метаданных могли быть неправильно истолкованы, но ошибка обнаруживается только при рассмотрении в контексте остальных данных наблюдений.

В таблице 8 ниже приведены примеры проверок контроля качества, которые могут быть выполнены в отношении спасенных данных в рамках процедур контроля качества, установленных на начальном этапе проекта. Дополнительную информацию можно найти в [Руководящих принципах по контролю и обеспечению качества данных станций приземных наблюдений для климатических применений](#) (ВМО-№ 1269), в [Руководстве по климатологической практике](#) (ВМО-№ 100), раздел 3.3, а также в публикации Westcott et al. (2011) и других источниках на портале C3S по спасению данных.

Таблица 8. Примеры проверок контроля качества в отношении спасенных данных

Проверка контроля качества	Описание
Последовательность дат	Проведите проверку на предмет пропущенных или дублированных дат. Обращайте внимание на високосные годы.
Средние значения	Рассчитайте средние значения за период времени и отобразите их на графике, чтобы выявить любые выпадающие значения, которые могут указывать на ошибки, допущенные при спасении данных.
Агрегированные значения	Суммируйте данные по наиболее значимому периоду времени (например, ежедневно в отношении периодов, составляющих менее суток, ежемесячно в отношении суточных данных) и постройте график, чтобы выявить любые выпадающие значения, которые могут указывать на ошибки, допущенные при спасении данных.
Диапазон значений	Определите минимальное, максимальное значения и диапазон значений за период времени (например, за сутки, месяц, год) и постройте график, чтобы выявить любые выпадающие значения, которые могут указывать на ошибки, допущенные при спасении данных.
Дублированные значения	Сравните длину записей значений с длиной записей дат, чтобы определить наличие дубликатов. Применительно к суточным значениям пометьте флажком последовательно повторяющиеся суточные значения данных (например, одинаковые значения на протяжении более трех дней) для сверки с исходными данными.

<i>Проверка контроля качества</i>	<i>Описание</i>
Выпадающие значения	Определите интервал (например, два стандартных отклонения по обе стороны от среднего значения) и выполните поиск отдельных выпадающих значений, чтобы сверить их с исходными данными.
Отсутствующие значения	Проведите проверку на предмет отсутствующих записей (значений) или значений, помеченных как «NA» или «-9999» или снабженных иными подобными обозначениями. Обеспечьте последовательное представление пропущенных значений.
Пробелы в данных	Путем подсчета количества отсутствующих записей за период времени либо путем визуальной проверки спасенных данных выявите пробелы в данных, чтобы сверить их с исходными данными. Обратитесь за консультацией эксперта, прежде чем применять методики интерполяции или заполнения пробелов, и только по мере необходимости.
Скачкообразные изменения	С помощью статистических тестов, визуальной оценки или определения порогового изменения (например, превышающего определенное количество стандартных отклонений от среднего значения) выявите скачкообразные изменения и выясните, объясняются ли они ошибками при спасении данных, связаны ли они с другими проблемами качества в наборе данных, или же это реальные изменения значений.
Присоединение наборов данных	Добавьте восстановленные данные к любым существующим данным по этой станции и проверьте, сколько времени занимает процесс присоединения. Он включает в себя поиск потенциальных скачкообразных изменений в данных, оценку формы/плавности представления непрерывных данных и анализ диапазона значений в обоих наборах данных.

3. ИНСТРУКЦИИ ДЛЯ СПЕЦИАЛИСТОВ ПО ОЦИФРОВКЕ

В инструкциях для специалистов по оцифровке следует обратить особое внимание на следующие пункты:

- обсуждение топографических ошибок в источнике;
- отсутствующие данные. Необходимо определить четкий порядок обозначения отсутствующих данных;
- специальные символы. Специалисты по оцифровке должны быть проинструктированы воспроизводить каждый символ в том виде, в котором он встречается в источнике, не интерпретируя его значение;
- отрицательные температуры. Одна из самых распространенных ошибок при оцифровке данных о температуре заключается в потере отрицательного знака;
- десятичные субъединицы. При оцифровке ранних данных инструментальных наблюдений часто приходится иметь дело с десятичными субъединицами;
- средства визуальной поддержки. В шаблонах ввода данных полезно добавить границы и/или разные цвета заливки для разделения столбцов и в целом для выделения структуры данных;
- отключение автозаполнения. По умолчанию большинство программ для работы с электронными таблицами предлагают текст для подстановки в ячейку, основываясь на том, что было написано в предыдущих ячейках. Эта функция может привести к ошибкам и должна быть отключена перед началом оцифровки;

- комментарии. Даже при тщательном изучении источников многие детали могут остаться незамеченными. У специалистов по оцифровке гораздо больше шансов их увидеть, и они должны быть проинструктированы вносить любые соответствующие метаданные (примечания, неизвестные символы, изменения единиц измерения и т. д.) в колонку для комментариев. Кроме того, в колонке комментариев можно сообщить о проблемах (например, опечатке в источнике или нечитаемом значении) и принятых мерах;
- рукописный текст. При работе с рукописным источником важно, чтобы специалисты по оцифровке потратили несколько минут на знакомство с почерком, прежде чем приступать к оцифровке;
- даты. Иногда страницы могут отсутствовать, дублироваться или располагаться в неправильном порядке. Специалисты по оцифровке должны всегда удостовериться в том, что дата (особенно месяц), указанная в документе, соответствует дате в шаблоне;
- суммарные значения за месяц. При наличии данных ежемесячные средние/суммарные значения или аналогичные величины могут быть полезны для обеспечения контроля качества. Они должны быть набраны таким образом, чтобы не мешать преобразованию данных (например, в специальном столбце или строке);
- самопроверка. Многие ошибки оцифровки можно легко обнаружить в электронной таблице, например, проверив максимальные и минимальные значения в столбце или нанеся данные по каждому столбцу на график. Специалистов по оцифровке следует поощрять к тому, чтобы они выполняли эти базовые проверки и исправляли собственные опечатки.

Несмотря на то, что большинство инструкций может использоваться применительно к любому шаблону, для каждого шаблона обычно требуется определенный поднабор инструкций. Также полезно привести в шаблонах образец напечатанных данных.

4. **ПОДХОДЫ К КОДИРОВАНИЮ КАЧЕСТВА ДАННЫХ ОБ ОКРУЖАЮЩЕЙ СРЕДЕ**

Для обеспечения качества измерений в интересах будущих пользователей может потребоваться кодирование спасенных данных. Например, общепринятой гидрометрической практикой является кодирование качества таких гидрологических данных, как уровни воды, кривые зависимости между расходом и уровнем и данные наблюдений за речным стоком (подробные указания см. в [Руководстве по гидрологической практике](#) (ВМО-№ 168). Однако эту практику можно распространить на любые данные диаграмм, которые подверглись редактированию в процессе спасения данных. Присвоение кодов качества спасенным данным может быть затруднено в случае отсутствия подробной информации или метаданных, обеспечивающих важный контекст для интерпретации данных наблюдений, особенно если на исторических диаграммах присутствуют признаки какого-либо сбоя или ненадлежащего сбора данных.

Коды качества, применяемые в отношении спасенных исторических данных, должны дополнять современную практику кодирования качества НМГС. Это особенно важно в тех случаях, когда спасение данных расширяет массив данных наблюдений на станции или заполняет некий пробел. Последовательный подход к определению и применению кодов качества в отношении существующих и спасенных данных позволяет оценивать качество измерений с течением времени. Может потребоваться пересмотреть цифровые

или буквенно-цифровые уровни современного подхода к кодированию таким образом, чтобы его можно было применить в отношении спасенных данных (см. пример Ирландии в публикации de Smeth et al. (2024)).

ПРИЛОЖЕНИЕ 10. НАКОПЛЕННЫЙ ОПЫТ

Примечание: это приложение опирается главным образом на опыт спасения климатических данных. Уроки, извлеченные из этого опыта, могут быть обобщены и использованы применительно к деятельности по спасению гидрологических, морских и других данных о состоянии окружающей среды.

1. СТИМУЛИРОВАТЬ БОЛЕЕ ТЕСНОЕ СОТРУДНИЧЕСТВО МЕЖДУ ОТДЕЛАМИ ПРОГНОЗИРОВАНИЯ И КЛИМАТА НМГС

Отдел климата НМГС часто осуществляет надзор за работой местных операторов, действующих на метеорологических объектах, контролирует качество национальных климатических данных (почасовые, суточные таблицы) и вводит часть данных в свою СУКД. Как правило, закодированные данные SYNOP отправляются метеорологами-наблюдателями с помощью Глобальной системы телесвязи ВМО, но во многих случаях эти цифровые данные доступны только отделу прогнозирования НМГС и часто только в режиме реального времени. Помимо этого, зачастую в отдел климата отправляются только рукописные таблицы почасовых и суточных данных, а не закодированные цифровые данные. Кроме того, если отдел прогнозов архивирует декодированные метеорологические данные со станций SYNOP или автоматических станций, они часто недоступны для отдела климата. В результате отдел климата должен получать данные по своей стране из международных баз данных. Было бы идеально, если бы передаваемые цифровые данные архивировались в доступной как прогнозистам, так и климатологам центральной базе данных, и если бы эта центральная база данных использовалась в отношении всей продукции. Более эффективная координация деятельности между этими группами обеспечит отделу климата доступ к более актуальному, а главное более полному набору данных для использования в рамках климатологического анализа. Она также сделает возможным сопоставление набора данных с печатными отчетами для проверки экстремальных значений.

2. ПРОВОДИТЬ ОБЗОР ТЕКУЩЕЙ ПРАКТИКИ ОЦИФРОВКИ, СБОРА МЕТАДААННЫХ И КОНТРОЛЯ КАЧЕСТВА; ПОТРЕБНОСТИ В ОБУЧЕНИИ

Отдел климата НМГС часто разрабатывает собственный метод ввода данных, переноса данных с диаграмм и контроля качества, который развивается в зависимости от имеющихся кадровых, коммуникационных и компьютерных ресурсов и количества/типа метеорологических станций в стране. Рекомендуется регулярно анализировать существующую практику и внедрять соответствующие руководящие принципы ВМО в структуру инвентаризации, ввода, переноса с диаграмм и контроля качества текущих данных. См. передовую практику, приведенную на портале СЗС по спасению данных.

Рекомендуется рассмотреть любые соображения, возникшие в ходе выполнения процессов спасения данных, которые могут улучшить процесс спасения данных или текущий процесс поступления данных или которые могут привести к выявлению других источников данных, нуждающихся в спасении. Климатические данные хороши лишь настолько, насколько хороши наблюдатели и синоптики.

Лицам, выполняющим работы по спасению данных, придется иметь дело с незнакомым новым оборудованием и программным обеспечением. Потребуется подготовка и опыт в области использования цифровых фотокамер, сканеров и программного обеспечения для оцифровки; доступа к базам данных и осуществления контроля качества. Поскольку специалисты переходят с работы на работу, необходимо при помощи практических пособий обеспечить обучение нескольких лиц в период, пока они знакомятся с процессом, чтобы исключить зависимость от единой точки сбоя. Эти пособия подлежат пересмотру по мере изменения программ и пользователей.

3. **ДЕМОНСТРИРОВАТЬ ПРЕИМУЩЕСТВА СПАСЕНИЯ ДАННЫХ**

Меры по упорядочиванию, обработке изображений, оцифровке и контролю качества спасаемых данных могут быть дорогостоящими и требуют значительных временных затрат. Возможно, это одна из причин, по которой спасение данных еще не реализовано во многих НМГС. Начальная стадия и выполнение программы по спасению данных начинается с самой НМГС, которой предстоит четко сформулировать, в чем заключается необходимость спасения данных. Для этого может быть нужно определить практические вопросы, на которые можно будет ответить улучшением климатических данных.

Например, исторические климатические данные позволят НМГС точно и быстро ответить министру на вопросы гидрометеорологии, касающиеся повторяемости явлений засух или ливневых дождей, или сообщить направление самого сильного ветра за последние 100 лет, чтобы правильно расположить новую взлетно-посадочную полосу аэропорта стоимостью 10 000 000 долл. США. Такая информация представляет собой продукцию с «добавленной стоимостью», за которую администраторы НМГС смогли бы взять плату, принеся дополнительный доход НМГС.

Кроме того, упорядочивание климатических данных могло бы привести в результате к высвобождению сотен квадратных метров общей площади помещений НМГС без того, чтобы отказываться от ценных исторических записей с данными. Увеличенную общую площадь можно переоборудовать под дополнительные офисы или учебные классы. Наконец, если НМГС будет готова поделиться спасенными и оцифрованными данными, она получит международное признание как лидер в этой области.

Деятельности по спасению данных часто требуется многолетняя внешняя поддержка. Приложение 11, которое касается вопросов применения, может помочь в предоставлении дополнительного обоснования для спасения данных. Отдельные лица и группы, которые уже провели спасение данных, могут выступать в качестве активистов и могут быть способны предоставить консультации и аргументировать финансирование усилий по спасению данных.

Консультации с другими лицами, заинтересованными в спасении данных, могут предоставить техническую информацию и возможные решения неожиданных проблем, возникающих на этом пути. Некоторая техническая информация и возможные контакты представлены на портале C3S по спасению данных. НМГС также могут обратиться в Секретариат ВМО за содействием в получении помощи у других Членов.

4. **ПРИМЕР ОЦЕНКИ РЕСУРСОВ И ЗАТРАТ (МЕТОДОЛОГИЯ ВВОДА ДАННЫХ)**

Оценки затрат на спасение данных в различных сценариях представлены в [Предлагаемом долгосрочном ресурсном плане в отношении управления климатическими данными и деятельности по спасению данных](#). В целом для достижения устойчивых результатов на осуществление трехлетнего национального проекта по спасению данных может потребоваться от 30 000 до 80 000 евро. В то время как с меньшими средствами будет трудно чего-то добиться (если только базовое спасение данных уже не является частью оперативной деятельности учреждения), есть возможность использовать больше средств для полной оцифровки национальных данных об окружающей среде.

Согласно неофициальным оценкам:

- для обработки изображений может потребоваться сделать от 800 до 1500 снимков, в среднем 1200 изображений на человека в день;
- документы о климате за период от тридцати до пятидесяти лет могут быть оцифрованы (методом ввода данных) за год силами одного человека, работающего

полный рабочий день, способного обрабатывать по десять метеорологических параметров и по три субсуточных значения в день (с учетом фотографирования, ввода данных и контроля качества).

ПРИЛОЖЕНИЕ 11. ПРИМЕНЕНИЯ ДЛЯ ДЕМОНСТРАЦИИ ЦЕННОСТИ ДАННЫХ

1. ПРОВЕРКА ЭКСТРЕМАЛЬНЫХ ЗНАЧЕНИЙ МЕТЕОРОЛОГИЧЕСКИХ ПАРАМЕТРОВ И КОСВЕННЫХ ДАННЫХ

Считается, что изменяющийся и более изменчивый климат приведет к более экстремальным значениям температуры, осадков, давления и ветра в будущем. Эти будущие значения не могут быть признаны «рекордными» без информации об экстремальных значениях, имевших место в прошлом. Сохранение исходной документации и оцифровка старых данных о погоде позволит легко проверять текущие и будущие экстремальные значения. При наличии поддающихся проверке значений метеорологических параметров будет больше известно об общем распределении экстремальных значений и о том, меняется ли распределение значений. Прошлые исследования климата в отношении периодов времени или регионов, по которым недостаточно метеорологических данных или они отсутствуют, основывались на косвенных данных (данные о годовых кольцах, данные о ледовом керне). Данные, которые удалось спасти, имеют огромную ценность для проверки обоснованности этих методов.

2. ГИДРОЛОГИЧЕСКИЕ ДАННЫЕ ДЛЯ МОНИТОРИНГА ИЗМЕНЕНИЙ

Деятельность в области гидрологического мониторинга всегда подчеркивала необходимость длительных данных гидрологических наблюдений для планирования использования водных ресурсов, оценки паводков и максимально полного понимания различных явлений. В последнее время возникла дополнительная потребность в долгосрочных наборах данных, с тем чтобы понять, как гидрологические режимы реагируют на климатические колебания и антропогенные воздействия.

В то время как глобальные климатические модели могут предоставить информацию об ожидаемых последствиях изменений окружающей среды, для проверки этих моделей требуются реальные данные. Что еще более важно, обществу необходимо понять последствия этих изменений на национальном уровне и на уровне водосбора и выявить соответствующие появляющиеся тренды или изменения в гидрологических режимах. Это можно сделать только путем оценки долгосрочных данных наблюдений, отражающих естественную изменчивость. Спасение данных обеспечивает возможность выявить ключевые данные наблюдений, которые ранее могли считаться несущественными (например, данные о небольших водосборах, используемых в сельскохозяйственных целях). Актуальность таких данных наблюдений для планирования использования национальных водных ресурсов и для глобального понимания последствий изменения климата должна стать одним из основных стимулов для деятельности по спасению данных.

Дополнительные причины, обуславливающие необходимость спасения гидрологических данных, включают следующее:

- более длительные данные наблюдений могут обеспечить более точную и достоверную оценку водных ресурсов и изменчивости паводков, что потенциально может привести к созданию более рентабельной и надежной инфраструктуры водоснабжения и защиты от паводков;
- более плотная сеть наблюдений или сеть, осуществляющая мониторинг более широкого спектра типов водосборов, может обеспечить лучшее понимание гидрологических процессов и улучшить результаты методов регионализации, которые лежат в основе многих гидрологических моделей;

- благодаря спасению данных можно расширить массив данных гидрологических наблюдений, затратив на это меньше средств, чем на постоянные полевые измерения;
- совершенствование практики управления данными и архивирования может принести организационные выгоды как за счет повышения эффективности, так и за счет улучшения доступа к информации;
- сохранение и использование качественных метаданных может привести к улучшению понимания и, следовательно, более эффективному использованию гидрологической информации;
- всеобъемлющий реестр массивов данных может быть полезен для демонстрации ценности архива данных и для обоснования текущих расходов на содержание такого архива.

3. **ИСПОЛЬЗОВАНИЕ В ИССЛЕДОВАНИЯХ, ПРИВЯЗКА К ИССЛЕДОВАТЕЛЬСКИМ ПРОЕКТАМ**

Лучший способ продемонстрировать ценность восстановления исторических климатических данных — это использовать их в научных исследованиях или обеспечить такое использование другими лицами. Длинные временные ряды позволяют определить и изучить интенсивность и повторяемость в прошлом таких явлений, как засухи и наводнения, волны тепла и холода, торнадо, тропические циклоны и выпадение града и пыльные бури. В то время как метеорологи и климатологи проявляют большой интерес к изучению причин возникновения погодных явлений, других интересует, какое влияние эти явления оказали на общество и даже на ход истории страны, поскольку последствия этих прошлых явлений и реакция на них со стороны людей могут оказаться полезны в планировании действий на случай будущих экстремальных явлений окружающей среды. Помимо этого, использование исторических данных для исследования прошлых явлений окружающей среды может вызвать большой энтузиазм у потенциальных спонсоров, особенно если эти же явления имеют место в настоящем или ожидаются в будущем. Возможность продемонстрировать историческую и современную статистику в отношении этих явлений может также помочь спрогнозировать вероятность их наступления в будущем. Полные временные ряды данных также помогают настраивать и проверять модели системы Земля.

4. **ПРЕДОСТАВЛЕНИЕ ДАННЫХ ПОЛЬЗОВАТЕЛЯМ**

Важно обеспечить наличие и легкодоступность данных об окружающей среде для пользователей. Предоставление достоверных исторических данных позволяет научным исследователям участвовать в научном обмене и может открыть образовательные возможности для учащихся. Руководители сельскохозяйственных и водохозяйственных организаций могут стать хорошими союзниками в обосновании необходимости спасения данных и располагать своими данными или ресурсами, которые они могут предоставить. Эти руководители, вероятно, сочтут спасенные данные важными для оценки чувствительности природных и антропогенных систем к экстремальным и прочим явлениям окружающей среды. Кроме того, разработка продукции, которая будет представлять общий интерес для населения (такой как климатические данные о погоде в дни национальных праздников), может стимулировать интерес к погоде и способствовать развитию групп ученых-любителей.

ПРИЛОЖЕНИЕ 12. КОНТРОЛЬНЫЙ ПЕРЕЧЕНЬ ДЛЯ СПАСЕНИЯ ДАННЫХ ОБ ОКРУЖАЮЩЕЙ СРЕДЕ

1. Осуществить поиск и установить местонахождение исторических наборов данных: на бумажных носителях, в виде микрофильмов/микрофиш или других старых носителей и цифровых данных в местных, региональных, национальных, международных библиотеках и климатических, гидрологических и морских центрах.
2. Упорядочить, очистить и хранить бумажные документы и микрофильмы/микрофиши/ другие носители в безопасных помещениях с регулируемой температурой и влажностью в промаркированных архивных коробках на полках или в картотечных шкафах
3. Создать электронный реестр массива данных на бумажных носителях/ микрофильмах/ хранящихся иным образом, включая данные по станции, год, месяц, тип носителя, тип формы и местонахождение коробки или ящика картотечного шкафа.
4. Создать электронную сводную таблицу реестра изображений со всеми обработанными и необработанными, проверенными и непроверенными изображениями.
5. Создать электронную сводную таблицу реестра цифровых данных со всеми данными, которые были и не были оцифрованы, прошли или не прошли контроль качества.
6. Определить приоритеты в отношении данных, подлежащих обработке в виде изображений, и данных, которые необходимо оцифровать.
7. Изучить текущие меры по оцифровке и контролю качества, чтобы понять, как новый поток оцифрованных данных может вписаться в рабочий процесс НМГС и нуждаются ли текущие меры по оцифровке и контролю качества в пересмотре.
8. Определить потребности в аппаратном обеспечении, программном обеспечении, персонале, помещении, обработке изображений и оцифровке.
9. Подготовить план работы по обработке изображений и оцифровке данных.
10. Приобрести необходимое оборудование и нанять персонал, если это необходимо.
11. Разработать процесс обработки изображений, включающий обработку и архивирование данных с использованием сводных таблиц реестра и контроль качества изображений и метаданных.
12. Обработать изображения всех копий данных и метаданных и надлежащим образом поместить их в архив.
13. Внести реестр изображений в каталоги и цифровые устройства хранения.
14. Сделать дубликат архива изображений и обеспечить его удаленное хранение.
15. Сохранить оригиналы на бумажных носителях/микрофильмах/других носителях.
16. Разработать процесс оцифровки, включающий разработку шаблонов и базы данных, обновление реестра и контроль качества данных и метаданных.
17. Анализировать процессы обработки изображений и оцифровки по мере продвижения работы, чтобы определить, существуют ли возможности для их улучшения.

18. Для обеспечения полноты процесса провести перекрестную проверку реестра изображений и цифровых данных относительно периода наблюдений по станциям.
 19. Ежедневно создавать цифровые резервные копии данных в компьютере.
 20. Регулярно обновлять носители, реестр и реестры цифровых данных с учетом текущих данных и метаданных.
 21. Создание дубликатов цифровых данных и размещение их в нескольких местах.
 22. Разработать план миграции технологий и обеспечить регулярную миграцию всех электронных данных (изображений и цифровых данных) на новые носители информации.
 23. Создавать продукцию на основе данных, прошедших контроль качества.
 24. Обеспечить предоставление данных пользователям в удобной, легкодоступной форме.
-

ПРИЛОЖЕНИЕ 13. ГЛОССАРИЙ

Ввод данных. Ввод или перепечатка данных в компьютер с помощью клавиатуры

Диаграммы самописца. Диаграммы, чаще всего вычерчиваемые на поворотном барабане в процессе автоматического производства таких постоянных измерений параметров окружающей среды, как температура, направление и скорость ветра, атмосферное давление и осадки, или таких гидрологических параметров, как уровень воды (то же, что и ленточные диаграммы)

Изображение. Документ, содержащий климатические данные, в форме изображения, полученного с помощью цифровой фотокамеры или сканера

Код ASCII. Американский стандартный код для обмена информацией, наиболее распространенный формат компьютерных текстовых файлов

Ленточные диаграммы. Диаграммы, чаще всего вычерчиваемые на поворотном барабане в процессе автоматического производства таких постоянных измерений параметров окружающей среды, как температура, направление и скорость ветра, атмосферное давление и осадки, или таких гидрологических параметров, как уровень воды (то же, что и диаграммы самописца).

Метаданные. Набор признаков или элементов, необходимых для описания ресурса. При спасении данных метаданные относятся как к информации о станциях, так и к информации о реестре климатических данных.

Миграция. Средство преодоления морального износа технологий аппаратного и программного обеспечения путем переноса данных с одного компьютерного носителя на другой для обеспечения сохранности интеллектуального контента цифрового объекта

Микрофильм. Метод воспроизведения изображений данных в значительно уменьшенном виде на фотопленке

Микрофиша. Форма микрофильма, изготовленная на прямоугольной пластине

НМГС. Национальная метеорологическая и гидрологическая служба

Носитель. Бумага, микрофильм, микрофиша, CD/DVD, магнитные ленты, жесткие диски и т. д. или компьютеры, на которых записаны данные

Обновление. Копирование информации с одного носителя на этот же или другой носитель без изменения формата

Оцифровка. Процесс преобразования аналоговых данных в цифровую форму для обработки компьютером, обычно выполняется путем ввода текстовых данных или переноса точек с диаграммы

Переформатирование. Копирование информации с одного носителя на другой или преобразование файлов из одного формата в другой

Репроустановка. Конструкция, удерживающая фотокамеры и осветительные приборы в положении для фотографирования документов

Сводная таблица. Формат, главным образом используемый при создании реестров

СД. Спасение данных, часто используется в отношении деятельности ВМО по спасению климатических данных

Сканер. Устройство, используемое для сканирования документов в формат цифрового файла

Сканировать. Воспроизводить с помощью сканера в виде графического файла данные с аналогового носителя или печатной копии цифровой записи

Сохранность. Обеспечение защиты данных от повреждения или утраты

Старый или устаревший носитель. Это выражение используется как синоним использовавшихся ранее носителей информации, включая микрофильмы, микрофиши, магнитные ленты, дискеты, WORM (write once, read many), CD-ROM и т. д.

СУГД. Система управления гидрометрическими данными

СУКД. Система управления климатическими данными

Цифровые данные. Данные, которые были введены в компьютер и сохранены в числовом формате. Эти числа могут быть либо в виде обычного текста, либо организованы в такую реляционную базу данных, как Microsoft ACCESS или ORACLE, либо предоставлены СУКД.

Цифровые устройства хранения. Цифровые системы хранения данных, включающие в себя различные устройства хранения, в том числе CD/DVD, жесткие диски, твердотельные накопители и т. д.

Электронный реестр. Сводная таблица с информацией об архивах данных

ПРИЛОЖЕНИЕ 14. ССЫЛКИ

Цитируемые ссылки

- Всемирная метеорологическая организация (ВМО). *Руководство по гидрологической практике* (ВМО-№ 168), том I. Женева, 2008, обновлено в 2020 г.
- . *Руководящие принципы по наилучшим практикам спасения климатических данных* (ВМО-№ 1182). Женева, 2016.
- . *Руководящие принципы по контролю и обеспечению качества данных станций приземных наблюдений для климатических применений* (ВМО-№ 1269). Женева, 2021.
- . *Руководство по климатологической практике* (ВМО-№ 100). Женева, 2023.
- Aguilar, E.; Auer, I.; Brunet, M. et al. *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186). WCDMP-No. 53; World Meteorological Organization (WMO): Geneva, 2003.
- Bojinski, S.; Verstraete, M.; Peterson, T. C. et al. The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bulletin of the American Meteorological Society* **2014**, 95 (9), 1431–1443. <https://doi.org/10.1175/BAMS-D-13-00047.1>.
- Brunet, M.; Brugnara, Y.; Noono, S. et al., 2020: *Best Practice Guidelines for Climate Data and Metadata Formatting, Quality Control and Submission of the Copernicus Climate Change Service Data Rescue Service*; Copernicus Climate Change Service (C3S), 2020. https://climate.copernicus.eu/sites/default/files/2021-05/C3S_DC3S311a_Lot1.3.4.2_2020_BestPracticeGuidelines_Part2.pdf.
- Climatic Data Rescue (CLIMDRE): www.climdre.com.
- de Smeth, K.; Comer, J.; Murphy, C. Hydrometric Data Rescue and Extension of River Flow Records: Method Development and Application to Catchments Modified by Arterial Drainage. *Geoscience Data Journal* **2024**, 11 (2), 176–196. <https://doi.org/10.1002/gdj3.206>.
- Lorrey, A. M.; Pearce, P. R.; Allan, R. et al. Meteorological Data Rescue: Citizen Science Lessons Learned from Southern Weather Discovery. *PATTERNS* **2022**, 3 (6). <https://doi.org/10.1016/j.patter.2022.100495>.
- Mateus, C.; Potito, A.; Curley, M. Engaging Secondary School Students in Climate Data Rescue through Service-learning Partnerships. *Weather* **2021**, 76 (4), 113–118. <https://doi.org/10.1002/wea.3841>.
- Ryan, C.; Duffy, C.; Broderick, C. et al. Integrating Data Rescue into the Classroom. *Bulletin of the American Meteorological Society* **2018**, 99 (9), 1757–1764. <https://doi.org/10.1175/BAMS-D-17-0147.1>.
- Tan, L. S.; Burton, S.; Crouthamel, R. et al. *Guidelines on Climate Data Rescue* (WMO/TD-No. 1210), WCDMP-No. 55; World Meteorological Organization (WMO): Geneva., 2004.
- Westcott, N.; Andsager, K.; Stoecker, L. et al. *Quality Control of 19th Century Weather Data*. Contract Report 2011-02; Midwestern Regional Climate Center, Illinois State Water Survey Institute of Natural Resource Sustainability, University of Illinois at Urbana-Champaign: Champaign, Illinois, USA, 2011. <https://core.ac.uk/download/pdf/4827106.pdf>.
- Wilkinson, C.; Brönnimann, S.; Jourdain, S. et al. *Best Practice Guidelines for Climate Data Rescue v1, of the Copernicus Climate Change Service Data Rescue Service*; Copernicus Climate Change Service (C3S), 2019. <https://doi.org/10.24381/x9rn-mp92>.
- Wilkinson, C.; Freeman, E. *Best Practice Guidelines for Keying Data from Historic Marine Documents*; Copernicus Climate Change Service (C3S), 2021. <https://doi.org/10.24381/68fy-2307>.
- World Meteorological Organization (WMO). *Climate Data Management System Specifications* (WMO-No. 1131). Geneva, 2014.
- . *Guidelines for Hydrological Data Rescue* (WMO-No. 1146). Geneva, 2014.
- . *A Proposed Long Term Resource Plan for Climate Data Management and Rescue Activities*; WMO: Geneva, 2017.

Дополнительная литература

- Всемирная метеорологическая организация (ВМО). *Руководство по приборам и методам наблюдений*. (ВМО-№ 8), том I. Женева, 2023.
- Brönnimann, S.; Brugnara, Y.; Allan, R. J. et al. A Roadmap to Climate Data Rescue Services. *Geoscience Data Journal* **2018**, 5 (1), 28–39. <https://doi.org/10.1002/gdj3.56>.
- Brunet, M.; Jones, P. Data Rescue Initiatives: Bringing Historical Climate Data into the 21st Century. *Climate Research* **2011**, 47 (1), 29–40. <https://doi.org/10.3354/cr00960>.

- Poli, P.; Dee, D. P.; Saunders, R. et al. Recent Advances in Satellite Data Rescue. *Bulletin of the American Meteorological Society* **2017**, 98 (7), 1471–1484. <https://doi.org/10.1175/BAMS-D-15-00194.1>.
- World Meteorological Organization (WMO). *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation* (WMO/TD-No. 1500). WCDMP-No. 72. Geneva, 2009.
-

За дополнительной информацией просьба обращаться:

World Meteorological Organization

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

Strategic Communications Office

Тел.: +41 (0) 22 730 83 14

Электронная почта: sra@wmo.int

wmo.int