

Руководящие принципы по наилучшим практикам спасения климатических данных

Издание 2016 г.

ПОГОДА КЛИМАТ ВОДА



ВСЕМИРНАЯ
МЕТЕОРОЛОГИЧЕСКАЯ
ОРГАНИЗАЦИЯ

ВМО-№ 1182

Руководящие принципы по наилучшим практикам спасения климатических данных

Издание 2016 г.



ВСЕМИРНАЯ
МЕТЕОРОЛОГИЧЕСКАЯ
ОРГАНИЗАЦИЯ

ВМО-№ 1182

РЕДАКТОРСКОЕ ПРИМЕЧАНИЕ

Терминологическая база данных ВМО «МЕТЕОТЕРМ» доступна по адресу: <https://public.wmo.int/ru/meteoterm>.

Читателям, копирующим гиперссылки, выделяя их в тексте, следует учесть, что могут появиться дополнительные пробелы, непосредственно следующие за <http://>, <https://>, <ftp://>, <mailto:>, а также за наклонными чертами (/), дефисами (-), точками (.) и неразрывными последовательностями символов (букв и цифр). Эти пробелы должны быть удалены из вставленного URL. Правильный URL отображается на экране, если навести курсор на ссылку или нажать на нее, а затем скопировать ее из браузера.

Дополнительная информация о наилучших практиках спасения данных представлена на Международном портале по спасению данных М-СД (www.idare-portal.org), в том числе иллюстрации и фотографии, более подробные спецификации практики и оборудования, истории успеха спасения данных, веб-ссылки, контакты и прочее. За любой дополнительной информацией следует обращаться на портал М-СД.

ВЫРАЖЕНИЕ ПРИЗНАТЕЛЬНОСТИ

Мы хотели бы поблагодарить г-жу Нэнси Уэсткотт, бывшего сотрудника Регионального климатического центра Среднего Запада, Служба водных ресурсов штата Иллинойс Университета Иллинойса, Соединенные Штаты Америки, за ее выдающийся вклад в эту публикацию.

ВМО-№ 1182

© Всемирная метеорологическая организация, 2016

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
Факс: +41 (0) 22 730 81 17
Э-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

	<i>Стр.</i>
ОБЛАСТЬ ПРИМЕНЕНИЯ	v
1. ВВЕДЕНИЕ — НАУЧНОЕ ЗНАЧЕНИЕ ДЕЯТЕЛЬНОСТИ ПО СПАСЕНИЮ КЛИМАТИЧЕСКИХ ДАННЫХ	1
2. ОБЗОР ДЕЯТЕЛЬНОСТИ ПО СПАСЕНИЮ КЛИМАТИЧЕСКИХ ДАННЫХ: АРХИВАЦИЯ, ОБРАБОТКА ИЗОБРАЖЕНИЙ И ОЦИФРОВКА	2
3. АРХИВАЦИЯ БУМАЖНЫХ НОСИТЕЛЕЙ	3
3.1 Поиск и определение местонахождения	3
3.2 Обеспечение сохранности и хранение	4
3.3 Создание электронных реестров массивов данных на бумажных носителях/микрофильмах и изображений	5
4. ОБРАБОТКА ИЗОБРАЖЕНИЙ НА ИСХОДНЫХ НОСИТЕЛЯХ	6
4.1 Основные этапы процесса обработки изображений	7
4.2 Проверка и хранение графических файлов	8
5. ПРЕОБРАЗОВАНИЕ В ЦИФРОВУЮ ФОРМУ ЗНАЧЕНИЙ ДАННЫХ	9
5.1 Минимизация ошибок при вводе данных	10
5.2 Сравнение с международными базами данных	11
6. АРХИВАЦИЯ ЦИФРОВЫХ НОСИТЕЛЕЙ (ИЗОБРАЖЕНИЙ И ЗНАЧЕНИЙ ДАННЫХ)	11
6.1 Перекрестная проверка печатных носителей, изображений и цифровых наборов данных	11
6.2 Обновление носителей	11
6.3 Рекомендации по архивации цифровых носителей	12
ПРИЛОЖЕНИЕ 1. ГЛОБАЛЬНАЯ ДЕЯТЕЛЬНОСТЬ	13
ПРИЛОЖЕНИЕ 2. ОЦЕНКА СПАСЕНИЯ ДАННЫХ	15
ПРИЛОЖЕНИЕ 3. ИНФРАСТРУКТУРА, ОБОРУДОВАНИЕ, ПРИНАДЛЕЖНОСТИ И ПЕРСОНАЛ	16
ПРИЛОЖЕНИЕ 4. РУКОВОДЯЩИЕ УКАЗАНИЯ ПО СОДЕРЖАНИЮ БУМАЖНЫХ АРХИВОВ	18
ПРИЛОЖЕНИЕ 5. ОПРЕДЕЛЕНИЕ ПРИОРИТЕТОВ ПРИ ОБРАБОТКЕ ИЗОБРАЖЕНИЙ И ОЦИФРОВКЕ ДАННЫХ	19
ПРИЛОЖЕНИЕ 6. МЕТОДЫ ЭЛЕКТРОННОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ	21
ПРИЛОЖЕНИЕ 7. ДРУГИЕ МЕТОДЫ ОЦИФРОВКИ	24
ПРИЛОЖЕНИЕ 8. ОШИБКИ ПРИ ОЦИФРОВКЕ И ПРОЦЕСС ОБЕСПЕЧЕНИЯ КАЧЕСТВА	26
ПРИЛОЖЕНИЕ 9. НАКОПЛЕННЫЙ ОПЫТ	28
ПРИЛОЖЕНИЕ 10. ПРИМЕНЕНИЯ ДЛЯ ДЕМОНСТРАЦИИ ЦЕННОСТИ ДАННЫХ	30
ПРИЛОЖЕНИЕ 11. КОНТРОЛЬНЫЙ ПЕРЕЧЕНЬ ДЛЯ СПАСЕНИЯ КЛИМАТИЧЕСКИХ ДАННЫХ	32
ПРИЛОЖЕНИЕ 12. ГЛОССАРИЙ	34

ПРИЛОЖЕНИЕ 13. ССЫЛКИ..... 36

ОБЛАСТЬ ПРИМЕНЕНИЯ

Данный технический документ представляет собой обновленный вариант публикации WMO/TD-1210, WCDMP-55, *Guidelines on Climate Data Rescue* (Руководящие принципы по спасению климатических данных, WCDMP-55, WMO/TD-1210) (2004). Он опирается на первоначальные руководящие принципы и учитывает как технологические изменения, произошедшие в последующие 12 лет, так и мировой опыт, накопленный в ходе более поздней деятельности по спасению климатических данных. Обзор спасения данных включает главы, посвященные его значению, архивации исходных носителей, обработке изображений, оцифровке и архивации цифровых изображений и цифровых данных. В двенадцати приложениях представлена вспомогательная информация.

Руководящие принципы по спасению климатических данных призваны обеспечить руководство в форме рекомендуемых наилучших практик. Ввиду разнообразия национальных метеорологических и гидрологических служб (НМГС) с точки зрения их размера и уровня технического развития, а также изменчивости типов погоды и климата 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 [Международном портале по спасению данных \(М-СД\)](#), который поддерживается ВМО при содействии Королевского Нидерландского метеорологического института и Экспертной группы по спасению данных Комиссии ВМО по климатологии.

Несмотря на то, что эти наилучшие практики касаются данных о погоде и климате, они также могут применяться для спасения данных в других научных областях как в рамках компетенции ВМО, так и за ее пределами. В частности, спасение гидрологических, морских и других данных об окружающей среде осуществляется в соответствии с аналогичными общими принципами и практиками и, по сути, относится к сфере применения настоящих Руководящих принципов. Однако специфику таких данных необходимо выявлять и учитывать в тесном сотрудничестве с соответствующими сообществами, включая, например, Комиссию ВМО по гидрологии и Совместную техническую комиссию ВМО и Межправительственной океанографической комиссии (ЮНЕСКО) по океанографии и морской метеорологии.

РУКОВОДЯЩИЕ ПРИНЦИПЫ ПО НАИЛУЧШИМ ПРАКТИКАМ СПАСЕНИЯ КЛИМАТИЧЕСКИХ ДАННЫХ

1. ВВЕДЕНИЕ — НАУЧНОЕ ЗНАЧЕНИЕ ДЕЯТЕЛЬНОСТИ ПО СПАСЕНИЮ КЛИМАТИЧЕСКИХ ДАННЫХ

Спасение климатических данных подразумевает организацию деятельности по сохранению зарегистрированных данных наблюдений приборов и климатических данных, 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 климата, а также для предоставления широкого спектра климатического обслуживания. Эти данные помогают ликвидировать пробел между палеоклиматическими данными и текущими наблюдениями. Многие из мировых наборов климатических данных содержат цифровые данные начиная с 1940-х и 1950-х годов, но многие наблюдения XIX и начала XX века существуют только в печатном виде и нуждаются в спасении и оцифровке. Некоторые регионы мира хорошо представлены вплоть до 1800-х годов, но данные по многим другим регионам отсутствуют. Достоверность климатических моделей и палеоклиматических данных наблюдений (например, в отношении годовых колец, ледовых кернов, пыльцы) значительно повысится при наличии длинных временных рядов инструментальных наблюдений во всех регионах мира.

Регионы с недостаточным покрытием данными сохраняются и в XXI веке. Поэтому получение и проверка более поздних данных имеют важнейшее значение. На надежность метеорологических наблюдений могут сильно повлиять местоположение и размещение станции, а также контроль качества (КК) наблюдений. Сравнение данных станции с данными с близлежащих объектов и возможность просмотра исходной документации наблюдений (и исходной информации о методах наблюдений) позволят климатологам обеспечить более качественную проверку прошлых и настоящих режимов метеорологических и климатических условий.

Расширение рядов климатических данных 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 исторической перспективе;
- c) обеспечивают основу для оценки исторической чувствительности природных и антропогенных систем к изменчивости и изменению климата, тем самым позволяя более точно оценить воздействие будущих изменчивости и изменения климата.

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 климатического обслуживания и исследователям данные (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 КЛИМАТИЧЕСКИХ ДАННЫХ: АРХИВАЦИЯ, ОБРАБОТКА ИЗОБРАЖЕНИЙ И ОЦИФРОВКА**

Спасение климатических данных подразумевает организацию деятельности по сохранению климатических данных, которые могут быть потеряны в результате повреждения, разрушения, запущенности, технического износа или простого рассредоточения массивов климатических данных с течением времени. Неоцифрованные данные подвергаются опасности в силу уязвимости исходных бумажных носителей данных. Спасение данных предусматривает организацию и сканирование данных на бумажной основе, микрофильмах и микрофишах; работу на клавиатуре с численными и текстовыми данными и оцифровку ленточных диаграмм в удобном формате, а также архивацию данных, метаданных и результатов и процедур проверок качества. Обзор компонентов деятельности по спасению климатических данных приводится в таблице 1.

Таблица 1. Компоненты деятельности по спасению климатических данных

<i>Компоненты спасения данных</i>	<i>Деятельность</i>	<i>Ключевые слова</i>
Архивация средств распространения информации на бумажных носителях и в виде микрофильмов/микрофиш (раздел 3)	Поиск и определение местонахождения	НМГС, пункты наблюдений, университеты, авиационные и морские учреждения, 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 электронного реестра массива данных на бумажных носителях/микрофильмах	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 CD/DVD или в каждом компьютерном каталоге	Обновление основного реестра изображений после их проверки на читаемость, включая метаданные Перекрестная проверка реестров файлов на CD/DVD с основным реестром изображений
Оцифровка значений данных (раздел 5)	Создание реестра цифровых данных	
	Ввод с клавиатуры, перенос данных диаграмм	Ввод данных в СУКД
	Контроль качества данных	Обновление реестра цифровых данных по мере оцифровки данных и прохождения различных проверок контроля качества

<i>Компоненты спасения данных</i>	<i>Деятельность</i>	<i>Ключевые слова</i>
Архивация цифровых носителей информации (раздел 6)	Перекрестная проверка печатных носителей информации, изображений и оцифрованных данных	Сравнение реестров изображений и цифровых данных с исходным электронным реестром массива данных на бумажных носителях и в виде микрофильмов
	Резервное копирование электронных носителей информации	Ежедневно
	Распределение дубликатов архивов изображений и цифровых данных	По разным местам хранения
	Обновление носителей и миграция технологий	Каждые 5—10 лет

Большинство стран вели добросовестные наблюдения за погодой, регистрировали их вручную или автоматически, заполняли бумажные формы или создавали копии микрофильмов и в конечном итоге сохраняли их на некоем типе компьютерных носителей (в идеале в Системе управления климатическими данными (СУКД)) 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. Многие бумажные формы и копии микрофильмов/микрофиш 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 климатических данных.

3. **АРХИВАЦИЯ БУМАЖНЫХ НОСИТЕЛЕЙ**

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 климатологических данных. Они могут храниться в национальных архивах и службах, на станциях наблюдений или в частных коллекциях и т. д., что не всегда бывает известно метеорологическому сообществу. Некоторые данные могут храниться даже в зарубежных архивах, особенно в архивах стран с колониальным прошлым. Без проведения такого поиска многие потенциально ценные наборы данных могут остаться невыявленными, и доступ к ним без ведома третьих лиц может быть невозможен. Поиск в следующих местах и источниках может выявить ценные климатические данные, которые могли быть забыты, в дополнение к тем, которые уже были упорядочены и оцифрованы и управление которыми реализовано в рамках существующей в организации или НМГС системы управления данными:

- a) НМГС и другие департаменты или агентства часто располагают собственными массивами данных;
- b) университеты, колледжи, средние школы и частные лица (например, давно работающие сотрудники метеорологических и связанных с ними организаций);
- c) международные центры данных;

- d) библиотеки или местные историки также могут располагать копиями данных на бумажных носителях, микрофильмах/микрофишах или даже цифровыми копиями данных;
- e) такие сельскохозяйственные организации, как крупные продовольственные корпорации, содержали плантации по всему миру и располагают метеорологическими данными наблюдений за сотни лет;
- f) военные ведомства также могут располагать данными, к которым может быть предоставлен доступ, как и авиационные и морские метеорологические данные в министерствах транспорта;
- g) религиозные организации, особенно заинтересованные в осуществлении наблюдений или имеющие научные интересы, например, монашеские ордены, занимающиеся обучением и воспитанием, и отдельные представители духовенства;
- h) научные общества, особенно ранее существовавшие организации (общества естествознания);
- i) музеи;
- j) судовые журналы, исторические газеты и личные дневники также являются 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.

Кроме того, ценными наборами данных располагают национальные библиотеки (ссылки на эти библиотеки можно найти на [портале М-СД](#)).

3.2 Обеспечение сохранности и хранение

Массивы данных на бумажных носителях и микрофильмах/микрофишах должны быть организованы логически и храниться в бескислотных архивных коробках на прочных полках или в картотечных шкафах:

- 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 в перчатках;

- f) в зависимости от количества пыли могут потребоваться защитная одежда и маски.

Под хранение архивных данных на бумажных носителях и микрофильмах часто выделяются специальные помещения:

- 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:

- a) от количества лет, за которые имеются данные, и от количества станций может зависеть наиболее логичный способ организации бумажных копий — 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, 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 (дополнительные сведения можно найти в приложении 4);
- 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 3. Предложения по содержанию бумажных архивов изложены в приложении 4.

3.3 **Создание электронных реестров массивов данных на бумажных носителях/микрофильмах 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. Когда изображения будут созданы, они должны быть доступны для оцифровки, после которой 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 микрофильмов/микрофишей), а также для последующей обработки, учета изображений и размещения графических файлов.

В целом, по мере обнаружения климатических данных следует включать их в электронный реестр массивов данных на бумажных носителях/микрофильмах, чтобы определить, какие имеются данные, какие усилия по их спасению уже были предприняты и что еще предстоит сделать. Данные электронного реестра включают 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 создания электронного реестра исторических климатических данных на национальном уровне 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 главе 4.

4. ОБРАБОТКА ИЗОБРАЖЕНИЙ НА ИСХОДНЫХ НОСИТЕЛЯХ

По завершении сбора и каталогизации бумажных копий и микрофильмов можно приступить к планированию, чтобы определить порядок обработки изображений и оцифровки данных. С чего начать и как выполнить эти задачи будет зависеть от потребностей конкретной НМГС. Приоритеты при обработке изображений и оцифровке данных, вероятно, будут отличаться. В приложении 5 рассматриваются вопросы, которые необходимо учитывать при определении приоритетов в порядке обработки и оцифровки.

Бумажные и/или микрофильмированные страницы фотографируются или сканируются для создания изображения или цифрового снимка, которые затем можно уберечь от дальнейшего разрушения и сделать широко доступными. Все страницы исходных метеорологических документов 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 6 рассматриваются относительные преимущества цифровых фотокамер и сканеров, а также желательные характеристики фотокамер.

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 носителях/микрофильмах	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 CD/DVD 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 CD/DVD или компьютера
Выполнить перекрестную проверку основного реестра изображений и реестров графических файлов на предмет количества изображений на CD или других носителях информации	

4.1 Основные этапы процесса обработки изображений

Практика и методы обработки изображений могут быть разными, но все они требуют ведения документации и отслеживания файлов. К числу настоятельно рекомендуемых основных процедур относятся следующие:

- а) создать основной реестр изображений на основе электронного реестра массива данных на бумажных носителях/микрофильмах, 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 CD/DVD (предоставить шаблон);

- b) по мере обработки изображений проверять обработанные файлы на читаемость, чтобы убедиться, что изображение считывается, а его информационное наполнение соответствует имени файла;
- c) большое количество изображений будет создано с помощью цифровой фотокамеры или сканера. Каждый раз при создании изображения и его загрузке на компьютер программное обеспечение фотокамеры/сканера генерирует имя файла для каждого изображения. Имена файлов могут не иметь отношения к их информационному наполнению. Если изображения многократно загружаются на компьютер в один и тот же день, некоторые имена файлов могут повторяться. 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 М-СД](#) и приложение 6;
- 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 каталоге, имеющем структуру «станция/год/месяц». Для 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 CD/DVD;
- f) обновлять основной реестр изображений с учетом количества обработанных страниц и сопоставлять исходное количество страниц с количеством изображений, фактически созданных и сохраненных на CD/DVD или жестком диске. Количество страниц, подсчитанных и внесенных в реестры, должно быть равно количеству созданных изображений. Указать местонахождение изображений (CD/DVD или наименование каталога);
- g) документирование процесса обработки изображений, включая соглашение о порядке именования файлов, 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 CD/DVD, или жесткий диск, или в СУКД:

- 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 освещения 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 целесообразно проверить несколько страниц изображений

и сравнить их 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 CD/DVD, 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 CD/DVD;
- b) копия имен файлов, связанных с каждым CD/DVD, должна быть указана на таком диске и внесена в основной каталог;
- c) маркировка CD должна быть включена в основной реестр изображений.

Чрезвычайно важно продублировать архив изображений на CD/DVD и внешних жестких дисках и/или в СУКД. Если данные с CD копируются на жесткие диски или в СУКД, каждый CD/DVD должен иметь каталог, и все реестры файлов также подлежат копированию. Рекомендуется использовать DVD, а не CD ввиду их большей вместимости.

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 не только храниться в региональном центре, но и распределяться по странам происхождения данных. При планировании обработки данных в региональном центре необходимо предварительно утвердить соглашение между 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 ленточных диаграмм самописца, так и преобразование текстовых данных в цифровой формат, пригодный для записи и чтения на электронных носителях. Первый случай рассматривается в приложении 7; оцифровка буквенно-цифровых данных посредством ввода данных рассматривается в этом разделе. Оцифровка часто производится в НМГС, но иногда и в региональных центрах силами компаний или в рамках гражданских научных инициатив, обычно под руководством экспертов в соответствующей области, например, климатологов. Если печатные данные находятся в национальном архиве, то, скорее всего, вводить данные потребуется на месте. При планировании обработки данных вне НМГС или даже за пределами страны необходимо предварительно утвердить соглашение между сторонами. Опять же, это может быть сделано при содействии Секретариата ВМО.

Оцифровка посредством ввода данных обычно занимает гораздо больше времени, чем обработка изображений: на 20 часов оцифровки приходится примерно час обработки изображений. 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 5.

В некоторых случаях в отсутствие средств для ввода больших объемов данных используется краудсорсинг. Добровольцы вводят данные в электронную форму или переносят временные ряды с ленточных 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, ленточных диаграмм и ОРС рассматриваются в приложении 7 и на [портале М-СД](#).

Первым шагом в оцифровке 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.

5.1 Минимизация ошибок при вводе данных

Существует ряд приемов, позволяющих минимизировать ошибки при оцифровке. 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 сопоставить каждое значение данных. В качестве альтернативы один или несколько операторов вводят суточные данные, а другой(ие) оператор(ы) — данные за месяц, после чего оценивается согласованность месячных общих и усредненных значений. Двойной ввод данных может осуществляться вне СУКД, а данные импортироваться после их сопоставления. Таким образом, данные могут вводиться непосредственно в СУКД.

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. 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 точную запись формул, использованных для перекодировки данных. Если необходимо использовать кодирование, следует позаботиться о том, чтобы оно было хорошо задокументировано и обратимо (т. е. чтобы можно было 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 исправлены.

Двойной/тройной ввод данных, создание шаблона и воспроизведение один в один значительно сократят количество ошибок. После ввода или импорта данных в СУКД в качестве предварительных данных необходимо произвести контроль качества (см. приложение 8). Этот этап невозможно переоценить.

Наконец, все данные на странице рекомендуется вводить сразу, главным образом потому, что вряд ли появится другая возможность вернуться к этой странице и ввести

пропущенные данные (например, широту, долготу, имя наблюдателя, время наблюдения, другие метеорологические переменные и все численные и текстовые 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 цифровые данные не помещались на хранение внутри организации в момент их появления, и доступ к ним может быть затруднен из-за проблем обеспечения пропускной способности, необходимой для загрузки очень больших файлов данных. Получить эти данные представляется желательным, поскольку они уже оцифрованы и могут содержать больше временных и переменных параметров, чем хранится в собственной климатической базе данных. Однако во многих случаях они могут содержать ошибки, допущенные при вводе, кодировании и передаче. При этом они, скорее всего, прошли процедуры обеспечения качества. Однако если значение на исходных бумажных формах было неясно, оно может быть восстановлено 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 наличия данных из международных баз данных, вновь оцифрованных данных и изображений во всех возможных случаях.

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. Это достигается 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 на основе исходных данных на бумажных носителях/микрофильмах был создан электронный реестр массивов данных на бумажных носителях/микрофильмах. К концу работы количество созданных цифровых данных (согласно реестрам изображений и оцифрованных данных) сопоставляется 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 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. Если цифровые данные все еще хранятся на этих носителях, следует обратиться в национальные архивы или университеты, чтобы перенести данные на такие современные носители, как DVD или жесткий диск. В противном случае 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 следующие новые, но уже зарекомендовавшие себя носители (например, CD, жесткие диски). В настоящее время широко используются CD/DVD, но и они могут выйти из употребления в ближайшем будущем. Кроме того, поскольку компьютеры (а также CD и жесткие диски) периодически выходят из строя без возможности восстановления, рекомендуется наладить процесс резервного копирования.

Следует подчеркнуть необходимость обеспечить миграцию программного обеспечения для записи, хранения, чтения и интерпретации климатических данных со старых компьютеров и компьютерных технологий и его обновление. Если делать это достаточно регулярно, то спасение данных обеспечивается до появления проблемы и утраты данных. 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, формат ASCII является предпочтительным для текстовых файлов данных с точки зрения удобства чтения цифровых данных и переноса 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) Следует защитить цифровые графические файлы и цифровые файлы данных. Графические файлы могут рассматривать как официальные документы и поэтому должны быть защищены;
 - b) упорядочить данные для удобства доступа к ним в будущем (так, чтобы доступ к данным можно было получить на CD или по структуре каталога: станция, год, месяц). Следует сохранять реестр файлов на каждом CD и в основном каталоге;
 - c) обеспечить ежедневное резервное копирование СУКД;
 - d) хранить цифровые изображения и данные на соответствующих носителях, чтобы данные оставались доступными в течение неопределенно долгого времени, с миграцией наборов данных по мере необходимости для снижения рисков морального износа и разрушения;
 - e) использовать 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 всех этих данных в нескольких местах, в том числе в многонациональном Региональном климатическом центре за пределами собственного региона.
-

ПРИЛОЖЕНИЕ 1. ГЛОБАЛЬНАЯ ДЕЯТЕЛЬНОСТЬ

1. КТО ДОЛЖЕН ЗАНИМАТЬСЯ СПАСЕНИЕМ ДАННЫХ?

Любая группа или отдельное лицо, в распоряжении которых находятся данные (на бумаге, микрофильмах или на основе цифровых технологий), должны пытаться содействовать спасению климатических данных. Лица, которые несут ответственность за управление регистрацией национальных климатических данных, 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 климатических подразделениях НМГС, а спасение данных является одним из элементов их обязанностей в качестве администраторов климатических данных. При этом сторонников спасения данных можно обнаружить во многих учреждениях, как в государственных (сельскохозяйственные департаменты), так и в частных (университеты, сельские муниципалитеты, агробизнес). Кроме того, добровольческие организации, такие как Международная организация по спасению данных об окружающей среде (МОСДО) и международный проект «Модели циркуляции атмосферы Земли» (АКРЕ) были созданы для содействия спасению данных. К другим сторонам, заинтересованным в оказании помощи деятельности по спасению данных, относятся работающие и вышедшие на пенсию климатологи, библиотекари, историки, студенты и супруги. Такие лица будут также весьма полезны в реализации привлечения 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 в этих отделениях может быть накоплен опыт и собрано оборудование в объемах, превышающих возможности отдельных стран. 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 зависит 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.

3. МЕЖДУНАРОДНЫЙ ПОРТАЛ ПО СПАСЕНИЮ ДАННЫХ (М-СД)

[Портал М-СД](#) является сетевым ресурсом для всех лиц, 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 может использоваться применительно в другим научным дисциплинам.

Поскольку его целями является повышение уровня общественного восприятия существующих видов деятельности по спасению данных, стимулирование развития ее новых направлений и улучшение координации международных усилий по спасению данных, [портал М-СД](#) представляет собой полезное средство коммуникации. Каждый включенный в него проект позволяет лицам, имеющим схожие интересы, узнать об отличной работе, проделанной каждой страной. [Портал М-СД](#) 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 привлечь финансовые ресурсы для реализации проектов.

ПРИЛОЖЕНИЕ 2. ОЦЕНКА СПАСЕНИЯ ДАННЫХ

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

Таблица 3. Контрольный перечень для оценки спасения данных

1. Изучить портал М-СД на предмет технических ресурсов и наилучших практик, связанных со спасением данных
2. Загрузить любые известные реестры данных из международных баз данных, библиотек, литературы для оценки известных массивов метеорологических данных страны
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 НМГС 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, 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, связанные с управлением данными (период, резервное копирование, содержание)
10. Изучить текущие процессы оцифровки и ОК, чтобы понять, как новые оцифрованные данные могут вписаться в этот 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 (WMO/TD-1186)
14. Провести консультации по вопросам ввода данных
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 М-СД

ПРИЛОЖЕНИЕ 3. 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 климатических 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. 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. Как обработка изображений, так и оцифровка проходят более эффективно, если персонал знаком с компьютерными технологиями и если работа выполняется в паре: один 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 6 и на [портале М-СД](#).

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
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 6) Репроустановка ПК Подставка для книг Стол Технические руководства	СУКД

	<i>Организация</i>	<i>Обработка изображений</i>	<i>Оцифровка</i>
Принадлежности	Перчатки, спецодежда, защитные маски, тряпки для пыли	CD, DVD и/или два самых больших внешних жестких диска USB-кабель или беспроводная сеть	
Персонал, ответственный за спасение данных	Два человека для очистки документов, организации и разработки архива и создания реестров	Два человека для производства фотосъемки и 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 для контроля за поступлением новых данных и метаданных	Климатолог для контроля качества и разработки продукции

ПРИЛОЖЕНИЕ 4. 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 и бескислотные архивные коробки. Помещение должно быть хорошо освещено, а коробки промаркированы и легко доступны. 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 изображений. Важно сделать несколько электронных копий реестров и всего архива изображений и обеспечить их удаленное хранение в нескольких разных местах.

Предпочтительно всегда сохранять бумажные копии:

- 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 со стороны этих национальных учреждений следует обратиться в международные архивы.

ПРИЛОЖЕНИЕ 5. 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 климатические явления (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, т. е. 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 метеорологических параметрах могут определять приоритеты в отношении оцифровки данных;
- h) 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 важнейших климатических переменных выберите метеорологические переменные, которые можно перевести в месячные значения, такие как температура и осадки (Vojinski 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. Однако 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 сразу вводить все имеющиеся на странице данные, включая метаданные (поскольку другая возможность вернуться к этой странице и ввести 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.

ПРИЛОЖЕНИЕ 6. МЕТОДЫ ЭЛЕКТРОННОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

Создание электронного изображения весьма желательно, поскольку сокращает объем работ с бумажными носителями или микрофильмами. Это также позволяет изменить размер и улучшить качество изображения, чтобы сделать надписи более разборчивыми при вводе данных. Как правило, подходят изображения в формате JPEG. 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. При использовании фотокамеры этого можно было бы избежать;
- b) микрофильмы и микрофиши требуют применения специальных сканеров;
- c) сканеры имеют много подвижных деталей и обычно требуют обслуживания и наличия лицензии;
- d) 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;
- e) за исключением очень дорогих систем, сканеры, как правило, работают медленно и занимают много места.

2. ФОТОКАМЕРЫ

Для создания цифровых изображений чаще всего используются 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 (АС) 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 их на компьютер. Процесс архивации изображений также происходит очень быстро.

3. ХАРАКТЕРИСТИКИ ФОТОКАМЕРЫ

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 глянцевая бумага, которая иногда используется в качестве бумажного носителя, может 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
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 ПК требуется программное обеспечение и USB-кабель или доступ к беспроводной сети связи. Если такое программное обеспечение позволяет задействовать затвор фотокамеры при помощи мыши, кабель дистанционного управления может не понадобиться. Если USB-порт или беспроводная сеть 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 и архивариусу. 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 с помощью USB-кабеля или беспроводной сети связи
5.	В качестве альтернативы можно использовать самую большую из доступных карт памяти
6.	Возможность и программное обеспечение для переименования графических файлов
7.	Увеличительная линза, при наличии таковой, и макролинза для съемки крупным планом
8.	Возможность изменения настроек 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. Следует помнить, 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

ПРИЛОЖЕНИЕ 7. ДРУГИЕ МЕТОДЫ ОЦИФРОВКИ

Ниже приводится общее описание методов оцифровки, отличных от ручного ввода данных. 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, но не способная обрабатывать рукописные данные наблюдений. Интеллектуальное распознавание символов (ИРС) предназначено для считывания отдельных печатных рукописных символов и позволяет использовать различные шрифты и шаблоны печати. ИРС лучше всего подходит для считывания форм, имеющих четкую структуру, и может достигать точности до 97 процентов. Интеллектуальное распознавание слов позволяет извлекать целые рукописные слова, будь то написанные печатными буквами или прописью. Однако исторические формы, как правило, не соответствуют требованиям ИРС/ОРС и не будут пригодны для обработки с помощью этой современной технологии. Последние два метода используют технологию машинного обучения для автоматического обновления баз данных распознавания новых образцов почерка.

На момент подготовки настоящей публикации эти технологии были недопустимо дороги и, как и все процессы оцифровки, требовали контроля качества для обеспечения точности преобразований. Эти технологии все еще находятся 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

В условиях краудсорсинга информация обрабатывается онлайн-добровольцами, которых часто называют «учеными-любителями». Краудсорсинг используется в ряде уже существующих и успешных программ, связанных с климатом, в частности в проекте по старой погоде (OldWeather) и в программе общественной объединенной сети добровольных наблюдений за дождевыми осадками, градом и снегом (the Community Collaborative Rain, Hail & Snow) (CoCoRaHS). Добровольцы проекта OldWeather вводят в онлайн-форму текстовые данные из судовых журналов (в основном начала XX века). Данные из каждого судового журнала вводят три человека, а результаты сравниваются на этапе конечной продукции (ссылку на OldWeather см. на [портале М-СД](#)). Добровольцы CoCoRAHS измеряют количество осадков с помощью 10-сантиметрового (четырёхдюймового) дождемера, который обычно расположен недалеко от дома, и вносят сведения о суточном количестве осадков в онлайн-форму. Эти данные помещаются в базу данных и наносятся на карту.

Хотя краудсорсинг использует для ввода данных добровольцев, это 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. Как и все другие технологии, краудсорсинг требует контроля качества.

3. **ОЦИФРОВКА ДИАГРАММ**

При оцифровке диаграмм наблюдатели обычно переносят значения диаграммы в табличную форму, что очень утомительно, особенно если речь идет о резких колебаниях, как в анемометрах, или множестве линий на одной диаграмме. Хотя на диаграмме могут отображаться данные с высоким временным разрешением, во многих случаях сообщаются только пиковые или суточные значения. Однако при использовании программного обеспечения для оцифровки диаграмм человек проводит линию, отмечая точки на диаграмме, и эти точки преобразуются в цифровые значения. Это обеспечивает временное разрешение значений данных в диапазоне от нескольких минут до нескольких часов. В 1990-х годах отдельными группами был разработан ряд программ для оцифровки данных об осадках (дождемеров-самописцев), но они могут использоваться только в сочетании с определенными компьютерами и операционными системами. Исходный код этого программного обеспечения доступен, но не поддерживается (например, не обновляется по мере появления новых языков программирования и компьютерных технологий), и специалисту по программному обеспечению придется его адаптировать. Кроме того, необходимо обучить персонал, обслуживающий и использующий программное обеспечение.

За последнее время было разработано и используется несколько новых программ для оцифровки. Функционал этих программ включает возможность вычисления количества дождевых осадков в разных временных масштабах (от нескольких часов до нескольких минут); возможность чтения различных типов диаграмм и обработки множества линий на одной диаграмме. Документацию по этому программному обеспечению можно найти на [портале М-СД](#). Некоторые из этих программ доступны для бесплатного использования при условии, что данные будут помещены в бесплатную и открытую базу данных общего доступа.

ПРИЛОЖЕНИЕ 8. ОШИБКИ ПРИ ОЦИФРОВКЕ И ПРОЦЕСС ОБЕСПЕЧЕНИЯ КАЧЕСТВА

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 СУКД страны, являющейся Членом ВМО. В этом случае ошибки будут отмечаться по мере ввода данных в СУКД и впоследствии исправляться. При отсутствии этих функций в рамках СУКД 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, содержащихся в метеорологических формах, 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. Другие распространенные ошибки включают: а) ошибки наблюдателя (например, использование нестандартных единиц измерения); б) неисправность прибора; в) проблемы с метаданными (несоответствие в части идентификатора станции или соглашения о порядке именовании, неверный год или месяц); г) перестановка или дублирование конкретного элемента применительно к данному месяцу и станции и е) 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: а) верные, соответствующие форме, разумные и не подлежащие изменению; б) помеченные как ошибка, но без изменения данных; в) ошибка, но с указанием заменяющего значения; г) не соответствующие исходной форме, исправленные для приведения в соответствие с исходной формой; д) ошибка с указанием на отсутствие данных, или е) как нуждающиеся в вводе с последующим вводом данных. В общем случае, если помеченное значение неоднозначно и может оказаться верным, его не изменяют.

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 СУКД, 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 и ежемесячным итоговым и среднемесячным значениям), проверки внутренней согласованности данных (сопоставление измерений между типами данных) и проверки экстремальных значений (по отдельным значениям). Дополнительную информацию можно найти в *Руководстве по климатологической практике* (ВМО-№ 100 (2016), раздел 3.4), а также в Westcott et al. (2011) и других источниках на [портале М-СД](#).

ПРИЛОЖЕНИЕ 9. НАКОПЛЕННЫЙ ОПЫТ

1. СТИМУЛИРОВАНИЕ БОЛЕЕ ТЕСНОГО СОТРУДНИЧЕСТВА 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. 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, из которой черпается вся продукция. 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 сравнение данных с печатными отчетами для проверки экстремальных значений.

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

Отдел климата НМГС часто разрабатывает собственный метод ввода данных 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 и контроля качества текущих данных. См. раздел «Best practices» (Передовая практика) на [портале М-СД](#).

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, 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. **ДЕМОНСТРАЦИЯ ПРЕИМУЩЕСТВ 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 НМГС, которым предстоит наглядно обосновать необходимость спасения данных. Это может включать идентификацию практических вопросов, 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 выбрасывать ценные исторические записи 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 10 может помочь в предоставлении дополнительного обоснования для спасения данных. Отдельные лица и группы, которые уже провели спасение данных, могут выступать в качестве активистов, способных 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 будут представлены на [портале М-СД](#). Чтобы попросить помощи у других Членов, НМГС могут обратиться в Секретариат ВМО.

ПРИЛОЖЕНИЕ 10. 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 (годовые кольца, ледовый керн). Данные, которые удалось спасти, имеют огромную ценность для проверки обоснованности этих методов.

2. 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, бури с градом 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. Последствия этих прошлых явлений и реакция на них со стороны людей могут оказаться полезны в планировании действий на случай будущих экстремальных погодных явлений. Использование исторических данных для исследования прошлых климатических явлений может вызвать большой энтузиазм у потенциальных спонсоров, особенно если эти же явления имеют место в настоящем или ожидаются в будущем. Возможность продемонстрировать (изменяющуюся) климатологию этих явлений может также помочь спрогнозировать вероятность их наступления в будущем. Полные временные ряды данных также помогают настраивать и проверять климатические модели.

3. ПРЕДОСТАВЛЕНИЕ ДАННЫХ ПОЛЬЗОВАТЕЛЯМ

Важно обеспечить наличие и легкодоступность климатических данных для пользователей. Предоставление достоверных исторических данных позволяет научным исследователям участвовать в научном обмене и может открыть доступ к студенческой помощи. Руководители сельскохозяйственных и водохозяйственных организаций могут стать хорошими союзниками в обосновании необходимости спасения климатических 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 климатическим явлениям. 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 11. КОНТРОЛЬНЫЙ ПЕРЕЧЕНЬ ДЛЯ СПАСЕНИЯ КЛИМАТИЧЕСКИХ ДАННЫХ

1. Осуществить поиск и установить местонахождение исторических наборов данных: на бумажных носителях, в виде микрофильмов/микрофиш и цифровых данных в местных, региональных, национальных, международных библиотеках и климатических центрах
2. Упорядочить, очистить и хранить бумажные документы и микрофильмы/микрофиши 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 с указанием станции, года, месяца, типа носителя, типа формы и местонахождения 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 компьютерах (СУКД), 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 на CD
14. Сделать дубликат архива изображений и обеспечить его удаленное хранение
15. Сохранить оригиналы на бумажных носителях/микрофильмах
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 улучшения процесса

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 с компьютера/СУКД
 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 климатическую продукцию на основе данных, прошедших контроль качества
 24. Обеспечить предоставление данных пользователям в удобной, легкодоступной форме
-

ПРИЛОЖЕНИЕ 12. 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 метеорологических измерений, как температура, направление и скорость ветра и атмосферное давление (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 метеорологических измерений, как температура, направление и скорость ветра и атмосферное давление (то же, что и диаграммы самописца).

Метаданные. Набор признаков или элементов, необходимых для описания ресурса. При спасении данных они 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 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 ленточной 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. В контексте этого документа имеются в виду программы для работы с таблицами Microsoft EXCEL или Lotus 1-2-3.

СД. Спасение данных. 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

СУКД. Система управления климатическими данными

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

Электронный реестр. Сводная таблица с информацией об архивах данных

ПРИЛОЖЕНИЕ 13. ССЫЛКИ

- Всемирная метеорологическая организация (ВМО), 2010: *Руководство по метеорологическим приборам и методам наблюдений*. (Руководство КПМН), ВМО-№ 8 (обновление издания 2010 года).
- , 2016: *Руководство по климатологической практике*. ВМО-№ 100 (обновление издания 2011 года), ВМО, Женева.
- Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons and M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy («Концепция важнейших климатических переменных в поддержку научных исследований, применений и политики в области климата»). [*Bull. Amer. Meteor. Soc.*, 95:1431–1443].
- Brunet, M. and P. Jones, 2011: Data rescue initiatives: bringing historical climate data into the 21st century («Инициативы по спасению данных: использование исторических климатических данных в XXI веке»), *Climate Research*, 47:20–40.
- Westcott, N.E., K. Andsager, L. Stoecker, M.L. Spinar, R. Smith, R. Obrecht M.C. Kruk, R.T. Truesdell and D. O’Connell, 2011: Quality control of 19th century weather data («Контроль качества метеорологических данных XIX века»). Midwestern Regional Climate Center Contract Report 11–04, Illinois State Water Survey, 68 pp.
- World Meteorological Organization (WMO), 2003: *Guidelines on Metadata and Homogenization* (Руководящие принципы по вопросам метаданных и обеспечения однородности). WMO/TD-1186, WMO, Geneva.
- , 2004: *Guidelines on Climate Data Rescue* (Руководящие принципы по спасению климатических данных). WMO/TD-1210, WMO, Geneva.
- , 2009: *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation* (Руководящие принципы по анализу экстремальных значений в условиях изменяющегося климата в поддержку принятия обоснованных решений в области адаптации). WMO-TD 1500, WCDMP-No. 72, WMO, Geneva.
-

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

Электронная почта: communications@wmo.int

public.wmo.int