UIS Informatics Commission
Publications Exchange Working Group

Speleological Publications Exchange Guidelines

Prepared by Trevor Faulkner (UK)
with input from members of the Group

Issue 3: 17 October 2017

Contents

Plus supporting spreadsheets in Excel file: PEWG Spreadsheets Issue 3:
10. Sheet: PEWG Contacts
11. Sheet: PEWG Journal Titles
12. Sheet: PEWG Paper Exchanges received, excluding from Europe
13. Sheet: PEWG Paper Exchanges received, from Europe (Austria to Italy)
14. Sheet: PEWG Paper Exchanges received, from Europe (Netherlands to UK)

****

1. Introduction

The Publications Exchange Working Group (PEWG) of the UIS Informatics Commission was established to tackle the challenges and opportunities for the exchange, accessing, organisation, safe storage and archiving of world-wide published speleological and karstic information into the distant future; it is believed that this literature has enormous historical and research value. The information has been partly disseminated through journals and books published by international and national bodies and by local caving clubs. Since the 1950s, many of these organisations have made international bilateral arrangements to exchange new publications amongst themselves, usually by post, but in an *ad hoc* pattern that remained unknown internationally. The documents exchanged are commonly housed in speleological libraries of various sizes and physical accessibility. Many are thus stored in relative safety and are available for study, at least to local members. Three main issues confront this beneficial exchange of information:

- the rising costs of postal delivery;
- the opportunities presented by electronic publication, electronic delivery, electronic storage and digital searching;
- the intention of some Governments for all papers that report publicly-funded research to be freely available under an 'Open Access' policy.

2. Terms of Reference of the Publications Exchange Working Group

The aims of the PEWG are to:

1 Promote the exchanges of previous and new paper journals and books that will be housed in physical libraries internationally;
2 Publicise the extent of this international cooperation to help safeguard the preservation of important speleological information into the distant future, as a widely distributed paper archive;
3 Examine the opportunities and problems related to the production, dissemination and long-term storage of information electronically, to improve its international, searchable, 'access' and to produce a distributed digital archive.

These aims are achieved by the creation of various recommendations and the maintenance of several supporting spreadsheets within these Guidelines.

The vision is that the PEWG will assist in the networking of those existing speleological libraries internationally that each hold and catalogue a good selection of published cave and karst literature.

3. The record of the Publications Exchange Working Group, 2013–2017

The Working Group was formally established within the UIS Informatics Commission, with close liaison with the UIS Bibliography Commission, at the 16th International Congress of Speleology (ICS) held in Brno, Czech Republic, from 21–28 July 2013. This action followed a meeting in Brno on 25 July about Journal Exchanges that was convened by the British Cave Research Association (BCRA) and attended by 19 Editors, Librarians and other interested people from 12 Countries. That meeting discussed the BCRA experience in re-creating a British Caving Library (BCL) and in re-establishing contact with many overseas organisations with whom exchange agreements had lapsed. The second meeting of the PEWG was held on 16 August 2016 in Yorkshire, UK, as part of the 2016 European Speleological Congress, to discuss progress and the initial Terms of Reference document with its Appendices at Issue 2. A third meeting was held as part of the 17th International Congress of Speleology in Sydney, Australia, on 28 July 2017. Now that most of the initial work of the PEWG is complete, its documents have been rationalised into these *Speleological Publications Exchange Guidelines* at Issue 3. Following the meeting in Sydney, future work should mainly include regular reviews of these *Guidelines* and the maintenance of the supporting status spreadsheets. During the meeting in Sydney, Trevor Faulkner, the first Chairman of the PEWG, announced his retirement at the end of 2017. The meeting nominated and agreed that the next Chairman will be Michele Sivelli, the Librarian for the Societa Speleologica Italiana.

All attendees at previous meetings of the PEWG, Editors and Librarians of international speleological organisations, and other interested people, are regarded as part of this Working Group. The main method of communication has been by email, with emails from the Chairman sent as blind copies to most of the email addresses that are listed in the PEWG Contacts list. The latest versions of these Guidelines are distributed as email attachments and this Issue 3 will be placed on the UIS website at http://www.uisic.uis-speleo.org/publexch/. This will be publicly available, but without email addresses. The associated spreadsheets are designed to be printed on A4 paper in Portrait or Landscape format, although this does not apply to the spreadsheets held on the website.

4. Speleological libraries

Recommendations

4.1 Speleological libraries should keep permanently two paper copies of their own organisation's publications. Single paper copies can be lost, burned, flooded or attacked by vermin, and Libraries should try to mitigate these risks; insurance payouts would not recover original lost material. Having additional paper and digital copies would provide a partial level of security in the event of such a disaster. If one of the copies is disbound for scanning to make a digital version (Chapter 7), it is advisable to return that copy to the library and retain it in a disbound state. Such a disbound copy can then be used for scanning again in the future, if files are lost or scanning technology improves.

4.2 Online catalogues of library contents and especially the organisation's own journal back-catalogues should be created and maintained. The catalogues should also be registered with the appropriate national UIS Documentation Centre (see below), where possible, and the Karst Information Portal (KIP). The PEWG might be able to advise suitable, perhaps standardized, software to help create them.

4.3 It is better to locate libraries in permanent premises (e.g. in Universities or rented property) and not in members' houses, to avoid problems of ownership, physical access, and movement when the Librarian changes.

4.4 Try to ensure that more than one person or a team is knowledgeable about the Library and its operation, especially if it needs to be located in a member's house.

4.5 Organisations that cannot fund the renting of Library premises from their normal revenues could consider seeking bequests and donations from their dedicated older members, other benefactors, or suitable charities.

4.6 Organisations should consider the safe physical archiving of their own early paper records, those of distinguished early members, and (for large organisations) those of other local organisations that no longer exist. Such information should be coordinated at a national, rather than international, level.

4.7 Organisations should consider digitising (in searchable format) their own and their members' early records, journals and books and (for large organisations) those of other local organisations that no longer exist, to make them more accessible. See the Recommendations for Digital Scanning in Chapter 7.

Speleological Abstracts

Various publications have listed and summarised the speleological journals issued each year. The British Speleological Association published 7 issues of *Speleological Abstracts* for British speleological literature for years 1962 to 1968. *Speleological Abstracts* was then replaced by *Current Titles in Speleology* (*CTS*). Issues of CTS from 1 (1969) to 18 (1985) were published privately. It became 'International' from number 4 (1972). Numbers 19 (1986) to 25 (1992) were published by the British Cave Research Association (BCRA). *CTS* was then incorporated into the *Bulletin Bibliographique Spéléologique – Speleological Abstracts* (*BBS-SA*) from its number 32 (1993). This continued to be published by the Swiss Speleological Society, with the support of BCRA, the French Speleological Federation and the Italian Speleological Society, up to 51/52 (2012/2013). Issues 21 to 51/52 are available on paper. Issues from 34 were also provided on CD. Issues from 53 (2014) will be published online only. All issues of *BBS-SA* from 1 (1969) to 27 (1988) are now available for free download from www.ssslib.ch/bbs. A searchable database is available for 27 (1988) to 46 (2007).

UIS Documentation Centres

The primary Documentation Centre of the UIS is at present housed at the library of the Swiss Speleological Society in La Chaux-de-Fonds (CH: below). According to the UIS Internal Regulations, it must maintain complete collections of all official UIS publications, including *ICS Proceedings*, *UIS Bulletins*, all issues of the *International Journal of Speleology* and *Speleological Abstracts*, and all scientific reports and papers produced

by the UIS Commissions, Committees and Working Groups (Labegalini, 2015, pp. 109-113). A UIS Bureau meeting in 1978 published a list of 'National Documentation Centers of Speleology', to overcome the problems of personal access to one physical library. By 2015, 18 such libraries were assumed to participate in exchanging publications and providing access to their collections (Labegalini, 2015). The *BBS-SA* website ssslib.ch/bbs/public/anglais/index.htm now lists 15 registered UIS Documentation Centres that should enable people to study international caving literature and to order copies of articles that have been cited in *Speleological Abstracts, CTS and BBS-SA*, perhaps electronically. Marcelo Rasteiro and Lucas Malafaia of the Brazilian Speleological Society (SBE) have since proposed that other major national speleological Libraries should be encouraged to register as a UIS Documentation Centre. Although this was not strictly a function of the PEWG, a revised Questionnaire to assist this registration process was sent to participants in Word file 'UIS questionnaire about libraries', which all Librarians are encouraged to complete, *if they are not already registered.* Completed forms should be returned to Lucas Malafaia, the SBE Secretary. However, it is important to recognise that other, unregistered, speleological libraries also have a valuable part to play in preserving information of national and local interest.

The 34th Brazilian Congress of Speleology approved a motion in June 2017 "*To support the proposal for recognition of National Documentation Centers for UIS through the Publications Exchange Working Group of the Informatics Commission*". This proposal was sent to the President, Vice-President and Secretary General of the UIS on 17 July 2017, but probably arrived too late to be actioned officially at the UIS General Assemblies during the 17th ICS held in Sydney. Nevertheless, the Proposal was discussed at the PEWG Workshop on 28 July 2017 and it was agreed in principle.

The 15 registered UIS Documentation Centres, as shown on the BBS-SA website ssslib.ch/bbs/public/anglais/index.htm, are listed below, although it is possible that some of these Libraries might no longer exist.

A	Speläologisches Dokumentationszentrum des Institutes für Höhlenforschung, c/o Naturhist. Museum, Burgring 7 ; A-1014 Wien, **Austria.**
B	Centre de Documentation, Union Belge de Spéléologie UBS/SSW, Avenue Arthur Procès, 5, B - 5000 Namur, **Belgium**. Tél : + 32 (0) 81 230009, Fax : + 32 (0) 81 225798
CH	Bibliothèque centrale de la Société Suisse de Spéléologie/Centre de Documentation, UIS, c/o Bibliothèque de la Ville ; CH-2300 La Chaux-de-Fonds, **Switzerland**.
D	Bibliothek des Verbands der deutschen Höhlen- und Karstforscher e.V., Dechenhöhle Iserlohn, Dechenhöhle 5, D-58644 Iserlohn, **Germany**.
E	Espeleo Club de Gràcia, C/ Asturies, 83, 1-1, 08.024 - Barcelona, **Spain**.
F	Fédération française de spéléologie - Centre national de documentation spéléologique, 28, rue Delandine, F-69002 Lyon, **France**. T. : +33 (0)4 72 56 09 63 - F. +33 (0)4 78 42 15 98 - http://codoc.ffspeleo.fr
I	Centro Italiano di Documentazione Speleologica «F. Anelli»: Via Zamboni 67 I-40126 Bologna, **Italy.** tel-fax : ++39 051 25 00 49. www.cds.speleo.it ;
J	Natural Science Museum, c/o Dr.Uéno, Hyakunin-cho 3-23-1, Shinjuku, J-Tokio 160, **Japan**.
P	Biblioteca Sociedade Portuguesa de Espeleologia, rua Saraiva de Carvalho 233,P-1350 Lisboa, **Portugal**.
PL	Library of «Kras i speleologia». Laboratory of Research and Documentation of Karst Environment. University of Silesia. ul. Bedzinska 60, 41-200 Sosnowiec, **Poland**.
R	Institutul de Speologie «Emil Racovitza», Calea 13 Septembrie nr.13, R-050711 Bucharest 5, **Romania**.
RA	Biblioteca del Instituto Argentino de Investigaciones Espeleológicas (IN.A.E) Pasaje El Payén 1035, (5613) Malargüe, Mendoza, **Argentina**.
RL	Bibliothèque Sami Karkabi - Centre de documentation UIS. Speleo-Club du Liban. B.P. 70-923, Antelias, **Lebanon**. FAX +961-1-334571. BP 16-5792 Achrafieh, Beyrouth 1100-2070.
UK	British Caving Library, The Old Methodist Chapel, Great Hucklow, BCRA Library, The Old Methodist Chapel, Great Hucklow, Buxton, SK17 8RG, **United Kingdom**.
YV	Biblioteca Sociedad Venezolana de Espeleologia, Apartado 47.334, Caracas 1041-1, **Venezuela**.

The other 3 libraries listed by Labegalini (2015) are:

Library “Dr. Emilio Maury”, Grupo Espeleologico Argentino , Heredia 426, C1427CNF, Buenos Aires, **Argentina**.

Sociedade Brasileira de Espeleologia, Biblioteca Guy-Christian Collet, Caixa Postal 7031, 13076-970, Parque Taquaral, Campinas, SP, **Brazil**.

National Speleological Society Library, 1, Cave Avenue, Huntsville, Alabama 35810, **USA**.

Reference

Labegalini, JA. 2015. Fifty Years of the UIS: 1965-2015, Ljubljana, Slovenia. 522 pp.

5. Paper Publications and Digital Warnings

Observations

5.1 Publishing on paper is probably still the most reliable method of achieving a long-term (perhaps >100 year) archive, especially for text in black and white. It avoids most of the complications of electronic storage. Please refer to comments by the VP of Google and by British Library archivists, as reported in the next two pages.

5.2 For some organisations, exchanging journals electronically could impact their sales of paper journals, which they might use, quite legitimately, to cross-subsidise other speleological activities.

5.3 In some countries, making cave information widely available electronically could result in more cave vandalism and raise safety issues for the public, who do not commonly read paper caving journals. However, in other countries, it has not been the practice to give cave location information, even in paper journals. Thus, each country needs to agree its own site information standards. These might vary from putting most cave grid references in all paper and electronic journals and on websites (if there is little extra risk, as in the UK), to limiting this to paper journals only, or to completely refusing to publish any site location information (as in Australia).

5.4 Despite using digital electronic production methods, it is expected that many high-quality journals will continue to be published on paper for those who request it, although many will also be published electronically. Receiving speleological journals on paper remains popular with many club members and organisations. It is clearly more efficient for the publisher to do the printing of covers and contents and the binding, rather than for each individual reader to try to do that for himself, to the same quality standard. Thus, some members may be prepared to pay a premium for this 'service', whilst others may prefer electronic distribution, to avoid paper altogether. Some people will also welcome both delivery methods, to take advantage of digital searching.

5.5 Commercial publishers increasingly charge authors for publication, especially for colour pages, as a way to help cover their costs. Additionally, commercial charges to authors or their institutions for making papers 'open access' are several thousand per article. (See also Chapter 8). These practices provide niche opportunities for speleological journals that are published locally and not via an international publishing house. Such journals could make similar but much more modest charges or simply request some small financial support from each author's academic institution, if applicable. However, any such charges or requests need to ensure that they do not limit the volume of submissions.

5.6 It is recognised that to attract good academic papers, some publications need to be peer-reviewed. They also may need to obtain an ISI number and a high citation index, so that authors who submit to them continue to receive research funding.

5.7 It is recognised that not all published speleological literature has been listed in the UIS *BBS-SA*.

Recommendations

5.8 For organisations that choose to publish only electronically, it is recommended that, in order to create a permanent archive, some copies (perhaps at least 10) should be printed to paper. Two of these should be stored in the organisation's own library, with the others possibly distributed more widely as exchange copies.

5.9 Each journal should be specified in the supporting Titles spreadsheet file, to show which issues were published on paper, which of those have been scanned, and which were produced digitally. Note that some digital journals might contain material that is not suitable for printing.

5.10 For articles that are not written in English, consider having an English summary or abstract. The contents of these Guidelines are biased towards Europe and English language journals at present. It might be beneficial if versions could be provided for Latin American use, written in Spanish and Portuguese.

The Guardian | Saturday 14 February 2015

The Guardian, Kings Place, 90 York Way, London N19GU 020 3353 2000

Digital is decaying. The future may have to be print

Ian Sample
Science editor
San Jose

Piles of digitised material - from blogs, tweets, pictures and videos to official documents such as court rulings and emails - may be lost for ever because the programs needed to view them will become defunct, Google's vice-president has warned.

Humanity's first steps into the digital world could be lost to future historians, Vint Cerf told the American Association for the Advancement of Science's annual meeting in San Jose, California, warning that we faced a "forgotten generation, or even a forgotten century" through what he called "bit rot".

Cerf called for the development of "digital vellum" to preserve old software and hardware so that out-of-date files could be recovered no matter how old they are.

"When you think about the quantity of documentation from our daily lives that is captured in digital form, like our interactions by email, people's tweets, and all of the world wide web, it's clear we stand to lose an awful lot of our history," he said.

"We don't want our digital lives to fade away. If we want to preserve them, we need to make sure that the digital objects we create today can still be rendered far into the future."

The warning highlights an irony at the heart of modern technology, where music,

Continued on page 2 →

Google boss warns about digital files

← continued from page 1

photos, letters and other documents are digitised in the hope of ensuring their long-term survival. But while researchers are making progress in storing digital files for centuries, the programs and hardware needed to make sense of the files are continually falling out of use.

"We are nonchalantly throwing all of our data into what could become an information black hole without realising it. We digitise things because we think we will preserve them, but what we don't understand is that, unless we take other steps, those digital versions may not be any better, and may even be worse, than the artefacts that we digitised," Cerf told the Guardian. "If there are photos you really care about, print them out."

Ancient civilisations suffered no such problems, because histories written in cuneiform on baked clay tablets, or rolled papyrus scrolls, needed only eyes to read them. To study today's culture, future scholars would be faced with pdfs, Word documents and many other file types that can be interpreted only with dedicated software, and sometimes hardware too.

The problem is already here. In the 1980s, it was routine to save documents on floppy disks, upload Jet Set Willy from cassette to the ZX Spectrum, slaughter aliens with a Quickfire II joystick and have Atari games cartridges in the attic. Even if the disks and cassettes are in good condition, the equipment needed to run them is mostly found only in museums. The rise of gaming has its own place in the story of digital culture, but Cerf warns that important political and historical documents will also be lost to bit rot.

Cerf concedes that historians will take steps to preserve material considered important by today's standards, but argues that the significance of documents and correspondence is often not fully appreciated until hundreds of years later.

Researchers at Carnegie Mellon University in Pittsburgh have made headway towards a solution to bit rot, or at least a partial one. They take digital snapshots of computer hard drives while they run different software programs.

These can then be uploaded to a computer that mimics the one the software ran on. The result is a computer that can read

Google vice-president Vint Cerf says we are nonchalantly throwing all of our data into what could become a black hole without realising it

otherwise defunct files.

Inventing new technology is only half the battle, though. More difficult still could be navigating the legal permissions to copy and store software before it dies. When IT companies go out of business, or stop supporting their products, they may sell on the rights, making it a nightmarish task to get approval.

"To do this properly, the rights of preservation might need to be incorporated into our thinking about things like copyright, patents and licensing," said Cerf. "We're talking about preserving them for hundreds to thousands of years."

Leader comment, page 32 →

Guardian G2, 17-02-15.

"Going, going, gone ", Lewis Dartnell

A huge amount of the information we consume and transmit in our everyday lives is perilously ephemeral. Every second, thousands of new photographs are uploaded to social media. Most of the images we take today are uploaded straight from a digital camera or a phone, with the picture never actually existing as a physical artefact.

So how will future historians and biographers piece together our lives and times without bundles of diaries, paper letters and professional correspondence? Family photos and emails are important to us personally, but what about more significant losses of our collective heritage? How do we preserve our interaction on Facebook, Twitter, comment threads and citizen journalism across the web? And does the "grey literature" of official reports, briefings and policy statements that are only published online also risk being lost to the future? In a speech last week, Google's vice-president Vint Cerf warned that a whole century of digital material could be lost.

There are some attempts to preserve this digital data. In 2010, the US Library of Congress signed an agreement with Twitter to archive public tweets sent since the platform's birth in 2006, and to continue preserving tweets to make this data available for analysis and research. In the UK, the British Library is taking bold steps to rectify what it refers to as the "digital black hole", where information is lost once it is taken down from a webpage or an entire site shuts down. Since 2004, it has been working to archive websites for future generations, just like paper-based literature. This effort received a huge boost in 2013 when the non-print legal deposit regulations came into force and allowed the British Library, as well as the five other UK deposit libraries, including those at Oxford and Cambridge universities and Trinity College Dublin, to archive all digitally published material. Nearly 5m UK-based websites will be preserved for the historical record, with regular snapshots taken so future historians can track how webpages evolve over time. Online retailers are also getting in on the act - services such as Blurb.co.uk or MySocialBook.com will print a physical photo album from Facebook posts.

But it is not just words and images that we risk losing for ever. Huddie William Ledbetter was an influential American folk and blues musician at the turn of the 20th century, admired as the king of the 12-string guitar. As Lead Belly he is included in the Rock and Roll Hall of Fame in Cleveland, and is considered the godfather of modern music; Elvis Presley, Johnny Cash, Led Zeppelin, the White Stripes, Red Hot Chili Peppers and Nirvana have all covered his tracks. Yet, sadly, many of his original recordings have already been lost to time. Tapes of his sessions have degraded beyond salvaging - the recording on a tape is stored as a magnetic imprint in a thin film of metal oxide, and if this delicate coating flakes off, the music is irretrievably lost.

The sound archive at the British Library is one of the largest such repositories in the world, and the archivists here estimate that around two million of their recordings are fragile and at risk of being lost for ever. These historical recordings exist on large reel-to-reel tapes, cassettes, lacquer discs and even wax cylinders, and are vulnerable not just to physical degradation, but obsolescence and the disappearance of the technology needed to play them. If archivists don't get to the deteriorating media soon, the very act of trying to copy a recording could destroy it in the process.

Similarly, deciding on the best format to preserve them for the next hundred years »

British Library archivists estimate about two million of their recordings are at risk of being lost for ever

6. Postal Exchanges

Observations

6.1 Exchange of publications between speleological organisations, both within countries and internationally, is accepted as an excellent way of spreading information and ideas, publicising discoveries, meetings and publications, and promoting the forwarding organisation. Traditionally, exchanges involved posting hard copies to the cooperating organisations, who posted their publications in return.

6.2 Some organisations do not publish a regular journal, but are pleased to exchange their published books and monographs.

6.3 Each organisation may also wish to supply issues of its previous books and publications from unsold stock to its Exchange Partners, even if these were not previously part of an Exchange Agreement.

6.4 For libraries with limited funds to pay for journal subscriptions and books, Publication Exchanges may provide the most economic way to obtain new material.

Recommendations

6.5 Good international email communication between people responsible for bilateral Publication Exchanges is essential. Each organisation should therefore identify one person who is responsible for Publication Exchanges. This may be the Librarian, or the Editor, or another person. Try to achieve good communication between Librarians and Editors regarding the receiving and sending of Exchange Publications: the left hand needs to know what the right hand is doing!

6.6 After establishing or reinstating exchanges of paper journals, each organisation should identify gaps in its holdings of foreign journal issues from its Library catalogue. The partner organisation can then be asked to supply copies to fill the gaps, from its stock of unsold publications.

6.7 Some organisations already send out many more journals for exchange than they get back, because they believe it is important to spread information as widely as possible. However, rising postal costs could limit these exchanges in the future. To mitigate the costs of postal delivery, organisations can consider:

6.7.1 Putting journals online immediately and batching paper deliveries annually.

6.7.2 Arranging exchange transfers at suitable international meetings or visits.

6.7.3 Focussing their exchanges within the same Continent, or with organisations that use the same language.

6.8 Most publication exchanges within each Country do not need to be known internationally, but local exchange processes that are similar to these international processes could be utilised. To assist this, a suitable spreadsheet format to list local caving publication titles and record local journal exchanges can be provided by Michele Sivelli.

7. Scanning documents for digital archiving

This discussion covers making electronic copies of paper publications. Digital archiving of photographs, movie films, video tapes, and sound recordings is beyond its scope. Some of it will apply to survey notebooks and other personal documents. The reasons for scanning publications that exist only on paper can range from just having a backup version that can be stored in several places away from the original as protection against loss, to preparing a version that can be viewed on-screen or printed that is as good as or perhaps better than the original. In the first case, just acquiring files of raw scans that preserve all the information on the original pages and storing them, properly cataloged and dispersed, can suffice. Additional effort is needed to turn the raw scans into a "reprint" suitable for printing a facsimile of the original or convenient reading on a computer screen. Additional benefits can be gained if the resulting file is put through the optical-character-recognition process (OCR: see below), so that the text can be searched and, if appropriate, indexed by services such as Google.

The goal when making the raw scans is to acquire grayscale, or colour if necessary, scans with a resolution of 300 or 400 DPI. If the original page contains very small print or a map or other illustration with very fine detail, a resolution of 600 DPI may be needed; resolutions higher than that are unlikely to be useful because of the texture of the paper or the inherent resolution of the original printing technology. There is no need for more accuracy in levels than 8-bit grayscale or 24-bit color. If the original is extremely clear and contains only black text or line drawings, a black-and-white scan (1 bit per pixel) might suffice, but that risks losing information that might be useful during further processing of the scan. (While black-and-white, as opposed to colour, sometimes encompasses grayscale illustrations, here it is distinct from them.)

It is easiest to obtain raw scans that are not distorted if the publication can be disbound for scanning, either by just removing staples or taking apart the book. Staples can of course by easily replaced, and rebinding, though probably not of the original quality, can be done if necessary. Of course whether this is done depends on the balance between the intended use of the scan and the value of the original, which may be old or rare.

The most common devices for scanning are flatbed scanners that handle pages up to A4 or US letter in size. This size will accommodate most pages of most publications. If paper is printed on both sides, some enhancement in quality of the scans may be gained by backing up the sheet being scanned with a sheet of black paper. This will keep the material on the back from showing through on the scan of the front if the paper is thin. Some multi-function printer/copiers include scanners that sheets are fed through, sometimes automatically from a stack. Needless to say, the latter can be used only for loose pages in good condition.

Pages that must be scanned from bound volumes can still be scanned on a flatbed scanner, but distortion near the binding edge is likely unless the inside margin was generous; pressing down to reduce this can damage a valuable volume. Some libraries have overhead book scanners, which scan from above a book laying open on the table. If available, this is a good way to scan a bound volume, although distortion where the page curves into the binding will still be visible. It might be possible to construct something similar by using a digital camera, or perhaps even a smart-phone, to acquire the image. Hand-held scanners exist that can be moved across, rather than down, the page. These might make it possible to reduce the distortion at the binding.

The Karst Information Portal (www.karstportal.org) offers the service of scanning originals for its archive, and provides a copy of the result. The KIP runs on a Cloud-hosted Drupal platform. It seemed better for Portal managers to control the data rather than to use the dSpace contributor support model. The KIP collection has steadily increased the number of contained digital objects. In June of 2006, it hosted just over 3,000 metadata records without digital content. Today it hosts nearly 7,000 metadata records, 43 percent of which link to digital content included in the KIP collection. For questions or comments concerning the KIP project, please contact Todd Chavez.

Foldouts or large maps that were originally distributed with the publication in loose, folded form will have to be dealt with separately. Foldouts can probably just be scanned in two or three pieces using the same equipment as the rest of the book, or a hand-held scanner could scan them all at once. Large sheets can also be scanned in pieces, but it is much easier to have them scanned at a library, copy shop, or similar service that has a large scanner. Whether the pieces need to be reassembled by computer depends on the intended use. For just archiving, the considerable effort involved may not be needed.

It is critically important to prevent inappropriate compression of the raw scans. Some compression schemes, in particular, the popular JPEG, are designed only for photographs and obtain their impressive compression ratios by eliminating detail that the eye will not perceive. When applied to text or line drawings, however, the algorithm degrades the image in ways that can easily be seen, and JPEG compression is not reversible; the original cannot be recreated from a JPEG (.jpg) file. The best way of storing image files is as TIFF (.tif) files. They can be compressed reversibly without loss of information using ZIP or LZW compression schemes. Essentially all programs that can read or write image files can deal with TIFF files. Most can deal with those types of compression, which still leave the file as type .tif. Make sure the program you used to scan is not set to save JPEG files.

Whatever the choice of digital medium, which is beyond the scope of this section, it will be most convenient to package raw scans into a single file for the publication, or maybe even a string of issues that make up a volume, rather than store perhaps hundreds of files. A good way to do this is to convert them all to PDF files and concatenate all the PDFs into a single PDF file. Some programs can do the conversion and concatenation in one step. (Some scanning programs will create the PDFs themselves, but might not give control over compression, and they will not be easy to manipulate further.) PDF is just as standard a format as TIFF and not likely to become obsolete any sooner, and if the PDF is properly made, the scan images can be easily copied back out of it if necessary. Again, care is needed about JPEG compression; some programs that create PDFs use it for grayscale or color images by default. Be sure that the creation of the PDF does not resample the file to lower DPI. Experiment with the programs in use, to make sure the images can be extracted, and that the extracted images do not show JPEG artifacts and that they have the same DPI as the scans. Be sure to include loose maps, either the pieces or as oversize pages, at the end of the PDF. An alternative is to package all the scans in their folder into a single .zip file by compressing the folder. Even if the intended use of the scans is to produce a good "reprint," the raw scans should be kept, although in that case redundancy of storage may not be important once the final product has been distributed. If the scanned publication is to be made widely available in libraries or on a web site, it is best to make the "reprint" attractive and as close to the original as possible. In this case a number of additional steps will be needed. Just what is done will depend on the software and time available.

The raw scans are likely to be crooked or contain blemishes beyond the edges of the printed page. Straighten up the image and crop to include just the printed area. The alignment should be done as accurately as possible. The cropping needn't be tight. Type and line art such as maps that are strictly in black-and-white will occupy less file space and print more sharply if converted to black-and-white from grayscale or color. It may be necessary to adjust the density levels in the raw scans to get an optimum conversion. If the quality of the original is very poor, it may be impossible to find an adjustment that leaves the text legible after conversion. Leave it grayscale.

If a page has grayscale or color illustrations, leave the whole page grayscale or color, as scanned. However, to get the best results for all the parts of a page, it is possible to remove the grayscale or color images into separate files and optimize them for contrast and color separately; the remainder of the page can then be converted to black-and-white. If the illustrations were originally printed poorly, it may be possible at this stage to improve them, not just reproduce the original. When printed, such illustrations will have been half-toned, i.e. converted to black-and-white or colored dots. Half-toned photographs can be improved by "descreening," blurring out the dots, without affecting the perceived sharpness. Image-editing software usually has some filter that will accomplish that. Avoid doing it to maps or charts, as there the resulting blurring may be worse than leaving the dots. Some scanning software offers to de-screen, but do not do that, because it may end up compromising the raw scan in undesired ways.

Any program that is capable of accepting large image files on a sufficient number of pages can be used to assemble the new version of the publication. Place the cropped black-and-white image of the black-and-white material on a page, duplicating the original margins and location, and then add whatever illustrations have been removed to be treated separately. It may be necessary to stretch slightly illustrations that were printed to the very edge of the page, if scanning or straightening left them too small. At this stage, make sure that the original pagination of the book or magazine is preserved, with any blank pages, including back sides of covers. This will result in a file that has black-and-white, for sharpness and economy of space, except where grayscale or color is needed.

After scanning, save, export, or whatever it is called by the program, into a PDF file. Again, be careful that grayscale and color illustrations do not get JPEG-compressed, which is sometimes the default in the conversion process, and that the DPI of the images is not changed. Fortunately, there is no such thing as JPEG-compressed black-and-white, so only non-photo grayscale or color illustrations are a potential problem here. (If all the grayscale or color illustrations are photographs, high-quality JPEG compression will do no real harm, but the result will not be absolutely faithful to the original.) Permitting ZIP compression if given a choice will reduce the file size without compromising quality, and it is especially effective in reducing the size of the black-and-white material that probably makes up most of the publication. If in doubt, view the resulting PDF at high magnification to check for JPEG artifacts, which will manifest themselves as fuzz around lines or letters in grayscale or color graphics. If the original publication was large with many illustrations, the result can be a file of several hundred megabytes. Nevertheless, this is what should be kept or distributed to other libraries. For viewing on the web, a much smaller version might be created by allowing aggressive JPEG compression; this can be done from the good PDF, without going back to the layout. The tradeoff between the quality of some illustrations and data transmission time and storage space might be acceptable in that context.

If talent, ambition, and software permit, a nice touch is fixing the page numbers known to the PDF reader to match the page numbers in the document, so that, for instance, page ix is shown in the header on the screen as ix. Bookmarks can also be added for easy access to chapters or articles. Also, as mentioned earlier, it can be very helpful to produce an OCR file of a PDF, so that readers can search for text on-screen and, if it is publicly accessible on the Web, search engines can index it. Please refer to the section below.

The above guideline was drafted by Bill Mixon for the International Union of Speleology's Publications Exchange Working Group in April 2016. Comments, suggestions for improvement or clarification, and critical input are invited from interested parties. Please address all correspondence to Bill Mixon or the Chairman.

Optical Character Recognition (OCR)

On the advice of Bob McIntosh, this section is a summary of the description provided by Wikipedia. Reference to this source should be made for the latest information and for comparisons of available OCR software. OCR is the conversion of images of typed, handwritten or printed text into machine-encoded text from a scanned document or a photograph of a document. Documents can then be electronically edited, searched, stored more compactly, and displayed on-line. Advanced systems capable of producing a high degree of recognition accuracy for most fonts are now common, and with support for a variety of digital image file format inputs. Some systems are capable of reproducing formatted output that closely approximates the original page including images, columns, and other non-textual components.

In the 2000s, OCR was made available online as a service (WebOCR), in a cloud computing environment. Various commercial and open source OCR systems are available for most common writing systems, including Latin, Cyrillic, Arabic, Hebrew, Indic, Bengali (Bangla), Devanagari, Tamil, Chinese, Japanese, and Korean characters. With the advent of smart-phones, OCR can also be used in internet-connected mobile device applications that extract text captured using the device's camera. Those devices that do not have OCR functionality built-in to the operating system will typically use an OCR API to extract the text from the image file captured by the device. The OCR API returns the extracted text, along with information about the location of the detected text in the original image back to the device app for further processing.

OCR engines can be used to:

- Make textual versions of printed documents more quickly, e.g. book scanning
- Make electronic images of printed documents searchable
- Convert handwriting to text

Types of OCR

- Optical character recognition – targets typewritten text, one glyph or character at a time.
- Optical word recognition – targets typewritten text, one word at a time (for languages that use a space as a word divider). (Usually just called "OCR".)
- Intelligent character recognition (ICR) – also targets handwritten printscript or cursive text one glyph or character at a time, but this usually involves machine learning.

- Intelligent word recognition (IWR) – also targets handwritten printscript or cursive text, one word at a time. This is especially useful for languages where glyphs are not separated in cursive script.

Techniques

OCR is generally an "offline" process, which analyses a static document. Although probably of little use with speleological handwritten documents, handwriting movement analysis can be used as input to handwriting recognition. Instead of merely using the shapes of glyphs and words, this technique is able to capture motions, such as the order in which segments are drawn, the direction, and the pattern of putting the pen down and lifting it. This additional information can make the end-to-end process more accurate. This technology is also known as "on-line character recognition", "dynamic character recognition", "real-time character recognition", and "intelligent character recognition".

Pre-processing

OCR software often "pre-processes" images to improve the chances of successful recognition. Techniques include:

- De-skew – If the document was not aligned properly when scanned, it may need to be tilted a few degrees clockwise or counter clockwise in order to make lines of text perfectly horizontal or vertical.
- Despeckle – remove positive and negative spots and smooth the edges
- Binarisation – Convert an image from colour or greyscale to black-and-white binary image. The task of binarisation is performed as a simple way of separating the text from the background. Most commercial recognition algorithms work only on binary images. The effectiveness of the binarisation step influences to a significant extent the quality of the character recognition stage. Careful decisions are needed in the choice of binarisation employed for a given input image type, since the resulting quality depends on the type of the input image (e.g. scanned document or historical degraded document).
- Line removal – Cleans up non-glyph boxes and lines
- Layout analysis or "zoning" – Identifies columns, paragraphs, captions, etc. as distinct blocks. Especially important in multi-column layouts and tables.
- Line and word detection – Establishes baseline for word and character shapes and separates words if necessary.
- Script recognition – In multilingual documents, the script may change at the level of the words. Hence, identification of the script is necessary, before the right OCR can be invoked to handle the specific script.
- Character isolation or "segmentation" – For per-character OCR, multiple characters that are connected due to image artefacts must be separated and single characters that are broken into multiple pieces due to artefacts must be connected.
- Normalise aspect ratio and scale

Segmentation of fixed-pitch fonts is accomplished relatively simply by aligning the image to a uniform grid based on where vertical grid lines will least often intersect black areas. For proportional fonts, more sophisticated techniques are needed because white space between letters can sometimes be greater than that between words, and vertical lines can intersect more than one character.

Character recognition

There are two basic types of core OCR algorithm, which may produce a ranked list of candidate characters.

The matrix matching type compares an image to a stored glyph on a pixel-by-pixel basis. It is also known as "pattern matching", "pattern recognition", or "image correlation". This relies on the input glyph being correctly isolated from the rest of the image, and on the stored glyph being in a similar font and at the same scale. This technique works best with typewritten text and does not work well when new fonts are encountered.

The feature extraction type decomposes glyphs into "features" like lines, closed loops, line direction, and line intersections. These are compared with an abstract vector-like representation of a character, which might reduce to one or more glyph prototypes. General techniques of feature detection in computer vision are applicable to this type of OCR, which is commonly seen in "intelligent" handwriting recognition and indeed most modern OCR software. Nearest neighbour classifiers such as the k-nearest neighbours algorithm are used to compare image features with stored glyph features and choose the nearest match.

Software such as Cuneiform and Tesseract use a two-pass approach to character recognition. The second pass is known as "adaptive recognition" and uses the letter shapes recognised with high confidence on the first pass to recognise better the remaining letters on the second pass. This is advantageous for unusual fonts or low-quality scans where the font is blurred or faded. The OCR result can be stored in the standardised ALTO format, a dedicated XML schema maintained by the United States Library of Congress.

Post-processing

OCR accuracy can be increased if the output is constrained by a lexicon – a list of words that are allowed to occur in a document. This might be all the words in the English language, or a more technical lexicon for a specific field. This technique can be problematic if the document contains words not in the lexicon, like proper nouns. Tesseract uses its dictionary to influence the character segmentation step, for improved accuracy. The output stream may be a plain text stream or file of characters, but more sophisticated OCR systems can preserve the original layout of the page and produce, for example, an annotated PDF that includes both the original image of the page and a searchable textual representation. "Near-neighbour analysis" can make use of co-occurrence frequencies to correct errors, by noting that certain words are often seen together. Knowledge of the grammar of the language being scanned can also help determine if a word is likely to be a verb or a noun. The Levenshtein Distance algorithm has also been used in OCR post-processing to further optimize results from an OCR API. In recent years, the major OCR technology providers began to tweak OCR systems to deal with specific types of input better. Beyond an application-specific lexicon, better performance can be had by taking into account business rules, standard expression, or rich information contained in colour images. This strategy is called "Application-Oriented OCR" or "Customised OCR".

Accuracy

Recognition of typewritten Latin-script text is still not 100% accurate even where clear imaging is available. One study based on recognition of old newspaper pages concluded that character-by-character OCR accuracy for commercial OCR software varied from 81% to 99%. Other areas, including recognition of hand printing, cursive handwriting, and printed text in other scripts are still the subject of active research.

Accuracy rates can be measured in several ways, and how they are measured can greatly affect the reported accuracy rate. For example, if word context (basically a lexicon of words) is not used to correct software finding non-existent words, a character error rate of 1% (99% accuracy) may result in an error rate of 5% (95% accuracy) or worse if the measurement is based on whether each whole word was recognised with no incorrect letters.

Web-based OCR systems for recognising hand-printed text on the fly have become well known as commercial products in recent years. Accuracy rates of 80% to 90% on neat, clean hand-printed characters can be achieved, but that accuracy rate still translates to dozens of errors per page, making the technology useful only in very limited applications.

Recognition of cursive text is an active area of research, with recognition rates even lower than that of hand-printed text. Higher rates of recognition of general cursive script will likely not be possible without the use of contextual or grammatical information. For example, recognising entire words from a dictionary is easier than trying to parse individual characters from script. Using a smaller dictionary can increase recognition rates greatly. The shapes of individual cursive characters themselves simply do not contain enough information to recognise all handwritten cursive script to an accuracy greater than 98%.

8. The Creation, Exchange, and Storage of Digital Publications

The purposes of this chapter are:

1. to alert editors, publishers and distributors to issues they should consider in producing speleological papers, journals and books in the 21st century;
2. to set out the standards and procedures endorsed by the International Union of Speleology in relation to publications;
3. to facilitate the exchange of digital publications between members of the Union and with other libraries and institutions; and
4. to attempt to maximise the long-term survival and accessibility of those publications.

It is not the purpose of these guidelines to discuss the respective merits or shortcomings of physical printing of documents on paper versus digital production, nor any copyright issues. It accepts that both methods of "publication" are available to speleological organisations and have their strong points and their advocates.

Before the advent of open access, readers or their institutions paid heavily for most peer-reviewed academic paper journals or to download individual papers. Whilst this method of distribution continues, the introduction of the Gold Open Access and the Green Open Access processes for the publication of papers have been introduced, as summarised by Donovan (2017). Gold Open Access requires the payment of a large fee by the author or his institution, so that readers can freely download the paper. Although Green Open Access is free for authors, if offered by the publisher, authors can only 'self archive' their own unformatted typescript for online access after an embargo period. Hence, all these formal processes have undesirable limitations. These should provide niche, and hopefully expanding, opportunities for economic publication for what has previously been referred to as the speleological 'grey literature', especially if this is peer-reviewed to academic standards.

1. Creation of publications for digital distribution

When creating new publications, editors now need to be conscious of how they may be distributed, viewed and stored. While there is an increasing tendency for digital documents never to be reproduced on paper, in the early 21st century it is desirable that magazines (including scientific journals) and books should be able to be either viewed on screen or printed out as hard copy. This requires the imposition of limitations on the use of any digital techniques which cannot 'translate' onto paper, such as the use of HyperText Mark-up Language (HTML – useful for designing web pages and viewing on screens because it works in windows of different sizes, but that flexibility means loss of control of the spacing and arrangement of material on the pages, as well as the actual fonts used), the inclusion of audio or video clips, scrolling windows or pop-ups and the use of hyperlinks to connect to external sources.

Scientific publications (and we include in this caving club newsletters as well as more esoteric speleological journals) are intended for the dissemination of information and the progressive accumulation of knowledge. To fulfil those objectives, they must be widely distributed and accessible. They must also continue to be available for reference in perpetuity and facts reported in them must be able to be cited. The system of referencing developed within the scientific literature is an essential element in the development of knowledge, the cross-checking of information and the prevention of effort-wasting by re-inventing wheels. If the citation system (particularly that referred to as the 'Harvard system') is to survive, the reliability of pagination in publications must be maintained. It is essential that every page of a journal (and it would be helpful if each page of a book or monograph) has, within a header or footer, its title and volume and part or issue numbers, as well as page number. (This ensures that if only a page, or a few pages, of a publication is copied or printed out, details of its origin are not lost.)

From mid-2015, the standard format for a digital publication file is Portable Document Format (PDF). Whichever word-processing or layout program is used to put the text and graphics of the publication together, it should be saved as a PDF file. This will preserve its content, pagination and appearance, no matter what computer or other device it is subsequently viewed on. Properly prepared PDF files contain the fonts they need and preserve the exact appearance of the document. Hard copies of the publication can also be produced from the PDF at any time, either in large numbers by a printing house or as single copies at a library or privately. The

PDF is now so widely used that one can be certain that when it is replaced by an improved format in the future, provision will be made for PDFs to continue to be read.

It will facilitate all aspects of document exchange, whether as hard copy or digital file, if the following points are also fully considered in publication design:

Pages should be of a standard size that can be economically printed and easily viewed, which means no larger than A4 or US letter size. Smaller sizes can be used but paper will be wasted when the publication is printed on standard paper.

Margins on the binding edge of each page should be adequate, so that important information can still be read in a printed and bound, or at least stapled, copy. Maps and photos should not be run across a centre-fold. If photographs are "bled" (run right to the edge of the paper), nothing vital should be placed right at the edge, because laser printers always leave a small blank margin.

Page numbers and headers or footers that identify the publication should be at least 5 mm or a quarter inch from the edge of the paper. If these rules are not followed, a reduced version of the publication would have to be printed, making type and everything else more difficult to see. (Anyone with the ability to print to only one of the standard paper sizes may have to reduce material designed for the other in any case, as A4 is taller than US letter, and letter is wider than A4.)

Graphics (maps and diagrams) in colour should be designed so that they can still be understood if printed in black and white (or at least greyscale). It would be helpful if points or lines in different colours could also be distinguished by different symbols or dots and dashes. Maps should always have a graphic or bar scale or a caption stating the spacing of grid lines or marks, not a statement like 1:500.

Resolution in graphics should always be set as high as practicable. 600 DPI is ideal for black and white; 300 DPI is adequate for greyscale and colour. Higher resolutions than those will increase the final file size, with little benefit. There is no need to resample graphics to exactly those numbers. Any printing or display software will do its own re-sampling to whatever the actual resolution of the hardware is.

There are many ways to convert a document into a PDF file. Most of them use, directly or indirectly, Adobe Acrobat, but there are other options. Generally there are various choices as to just how the conversion is done, and the default choices are not always the best ones. Make sure that fonts are embedded, that the conversion to PDF does not change the resolution (DPI) of the graphics or at least does not reduce it below 600 DPI for strictly black-and-white material or 300 DPI for greyscale or colour, and that the conversion does not use JPEG compression if there are any greyscale or colour illustrations other than photographs. ZIP compression is acceptable.

A PDF file format that is standard in the printing industry is PDF/A. Files that meet that standard include the fonts, which are important to an editor. They should also contain information to make possible very accurate colour rendition by offset presses, which an editor may not regard as particularly important. Even if a file is not PDF/A, the fonts should at least be embedded. (Very rarely a font cannot be embedded for copyright reasons. In such a case, another font will need to be used and the layout revised. None of the fonts included with any home-computer operating system should have this problem.) Some old printer drivers might have trouble with a PDF file that contains advanced features such as transparent overlays or shadowed text, which should be avoided.

2. Exchange of digital publications

2.1 Posting hard copies

The rising cost of printing and postage (especially international postage) has led many groups to look for more economical ways of carrying out exchanges. A PDF file or collection of files could be mailed to exchange libraries on disks, but CDs and DVDs are rapidly becoming obsolete technology; some home computers already are not equipped to read them. Packaging disks and posting them still costs money, though not as much as printing and mailing paper copies. There is also a question as to the longevity of optical disks.

2.2 Utilising e-mail

Small publications or significantly compressed ones can be sent as digital exchanges by simply attaching them to e-mails. This is fairly convenient to all parties, especially as all the receiving librarian has to do is save the attachment. The quality of the publication may be severely degraded, however, and this technique would not work for large magazines or books.

2.3 Utilising the Internet

A high-quality PDF file of a publication may well be hundreds of megabytes in size. It will almost certainly be too large to e-mail as an attachment. A much smaller PDF might also be made with reduced DPI graphics, more compression, etc., for distribution only for on-screen reading like a web site, but the best possible file should be the one archived and distributed to exchanges. (A receiving librarian may choose to compress the file for storing in an electronic archive, if required.)

There are also various ways to put a PDF file on the Internet where exchange librarians can access it. Many are free unless you exceed specified limits (often ~10GB). Microsoft Cloud, Google Drive, iCloud, and Dropbox are examples. If submitting publications to the Karst Information Portal (see below), it could serve as a cloud source for other librarians, although there might be a delay while it is processed by KIP. Another option is to put the PDF file on your own or another organisation's web site. Note that it can be put there even if public access is denied by giving it a file name that no link points to. This will prevent anyone (including Google) who is not given the exact URL of the file from accessing it.

Exchange librarians then need to be sent e-mail messages that tell them the issue is available and contains the link they need to click to download it. Exchange librarians, who may be dealing with dozens of exchanges in their spare time, *should not be expected to check a number of sites periodically for new material*. Most on-line storage schemes have a way to send a link that does not require the recipient to wrestle with a web site, but just to click instead. The PDF file should be kept available at the stated link for a period of months, because the librarian is probably a volunteer with limited time and no vacation backup.

3. Storage of digital publications

Recipient librarians need to make some decisions regarding storage of exchanged (and, indeed, their own organisation's) digital publications.

3.1 Print and store hard copy

Some librarians may prefer to continue to maintain collections of hard-copy newsletters and journals. They would have the option of printing publications out from the downloaded PDF files (preferably on a colour laser printer using ISO 9706 acid-free, archive-quality paper, as the inks used in ink-jet printers are not regarded as sufficiently stable for archival storage). With increasing use of digital publication, this process could require considerable amounts of time, printing resources, and physical storage space. Additionally, there remains the problem of printing the covers and then binding the document.

3.2 Save digital copies

It is considerably easier, quicker, and cheaper to save copies of the digital files than to print them out.

Removable optical disks (CD-R and DVD±R). These are cheap, inexpensive, and reliable. If properly cared for, the data written on a CD or DVD will likely "outlast the technology." But this may be already happening, as some new computers are not equipped to read them. Another disadvantage of such disks is that they cannot be updated once written to, by, for instance, adding a new issue of a periodical. (Never use rewritable CD-RW or DVD±RW disks; they are not as reliable for long-term storage).

Hard disk drives. These are currently the fastest and easiest way to save digital files. They are, perhaps surprisingly, cheaper per gigabyte than DVDs, with prices continuing to fall. Because they contain moving parts, they are subject to mechanical failure and cannot be counted on to last indefinitely, even if they are not powered most of the time. They are also subject to technological obsolescence, as electronic interfaces or file-system software evolve. Hence, re-copying the information needs to be planned for the future.

USB sticks. These solid state devices can now store 64GB of data and should be more reliable, but are easily lost.

The "cloud". Many commercial services, some of which are mentioned above, are available for storing digital information and accessing it over the Internet. They are not free, but the cost is low and falling rapidly. Store material can be accessed by multiple people from multiple locations, as long as they know the password. These services also have the advantage that they will deal with matters of technological obsolescence in hardware and file systems themselves, and they are reliable because they do their own backups to distributed locations. However, this assumes that subscriptions continue to be received and that the company stays in business.

The Karst Information Portal (www.karstportal.org) could also serve as a cloud facility, but it is oriented toward making everything publicly available on the web, although this policy is evolving somewhat. It requires permissions and other agreements to accept new material. The International Union of Speleology is one of its partners. It is connected with a university library in Florida, USA. It is free, but it is not run by volunteers and long-term funding may not be assured.

3.3 Backups

In any case, a single storage medium is never sufficient, and there should always be at least one backup, with provisions to keep the copies synchronised by updating the backup as changes to the collection occur. Two or more hard disk drives in different locations or hard disk(s) with cloud backup are recommended. As with hard copies, some process of cataloguing accessions is needed, to keep track of publications received, even though searching through digital media is extremely fast and reliable. The catalogue itself, of course, should be backed up along with the data, and it should be made publicly available on the web, so that researchers can discover the receiving organisation as a source for the material.

This guideline Chapter has been drafted mostly by Greg Middleton and Bill Mixon for the International Union of Speleology's Publications Exchange Working Group in April 2016. Comments, suggestions for improvement or clarification, and critical input are invited from interested parties. Please address all correspondence to the authors or to the Chairman.

Reference

Donovan, SK. 2017. Open access and geology. Proceedings of the Geologists' Association 128, 163-164.

9. Review of previous actions

Actions from 13 November 2013

Action no.	Action	Action on	Target date	Action status
2013/1	Participants are asked to read and comment on the ideas in the draft Terms of Reference and to propose new ideas for consideration.	All	31/12/13	Complete. The ideas have been generally welcomed and the recommendations adopted by several organisations. Technical advice was provided by Greg Middleton and by Bill Mixon, who has written a preliminary paper (Appendix 4) on scanning documents to make good pdf files with Acrobat.
2013/2	Following /1, make any necessary amendments, and submit to the Chairman of the Informatics Commission for approval.	Chair-man	Q1 2014	Complete. ToR submitted on 5 February 2015.
2013/3	Propose a template by which each organisation can summarise its existing Publication Exchange arrangements	Chair-man	Q1 2014	Complete, by design of Journal Titles Published and Exchanges Summary spreadsheets.
2013/4	If participants agree with the proposed vision and approach, discuss these ideas locally, try to follow the recommendations made in Section 1, and provide feedback on progress and issues arising.	ALL	2014	Complete. Feedback and information has been provided from 2014 to 2017.
2013/5	Participants are asked to recommend volunteers with the requisite skills to form a subgroup to lead the Electronic Publications Exchange activity outlined in Appendix 5.	ALL	2014	Complete. Bob McIntosh, Greg Middleton and Bill Mixon have offered to work on this. Todd Chavez joined the team in April 2015.

Actions requested 12 February 2015

Action no.	Action	Action on	Target date	Action status
2015/1	Participants are asked to comment on and supplement the data in the Working Party Contacts spreadsheet by returning it to the Chairman with changes highlighted in green.	All	end June 2015	Updates received: complete
2015/2	Participants are asked to comment on and supplement the data in the Journal Titles Published spreadsheet by returning it to the Chairman with changes highlighted in green.	All	end June 2015	Complete.
2015/3	Participants are asked to comment on and populate the data in the Publications Exchanges Summary spreadsheet by returning it to the Chairman with changes highlighted in green.	All	end June 2015	Withdrawn. Only a few inputs were received. Perhaps the spreadsheet was too complicated. It is now replaced by a simplified Publications Exchanges Status spreadsheet in Appendices 8.1 and 8.2.
2015/4	Prepare a UIS Library Questionnaire, so that major speleological libraries can register as a 'UIS Documentation Centre'	Marcello Rasteiro	end June 2015	Complete.
2015/5	Prepare draft Recommendations for Digital Scanning	Bob McI. Greg M. Bill M.	end June 2015	Complete. Bob McIntosh suggested that Appendix 4 should start with a generic statement about scanning that is independent of the technology used. This could be based on the KIP guidelines. Other correspondence

				is added.
2015/6	Prepare draft Recommendations for Electronic Publications Exchange	Greg M. Bill M. Bob McI.	end June 2015	Complete. Input to Appendix 5 received from Bill Mixon, who raised two new issues: "Digital Preservation" and the advent of caving journals using multimedia.

Actions requested 10 February 2016

Action no.	Action	Action on	Target date	Action status
2016/1	Participants are asked for any further comments on the ToR and its Appendices by returning them to the Chairman with suggestions highlighted in green.	All	end June 2016	Complete. Inputs and comments received were included in the Appendices at Issue 2 draft.
2016/2	Participants are asked to populate the data in the Publications Exchanges Status spreadsheet in Appendices 8.1 and 8.2 by returning them to the Chairman with additions highlighted in green.	All	end June 2016	Mainly outstanding. Appendices 6 and 7 have been improved and simplified by introducing First and Second lists of exchanging organisations that are now organised by Continent. Appendix 8 follows the same organisation for the First list only, is now split into two sheets, and is gradually being populated. Thanks to Lucas Malafaia, Bob McIntosh and Michele Sivelli especially for their ideas and suggestions. See 2016/13.
2016/3	Arrange a Working Group workshop at the EuroSpeleo event in Yorkshire, UK, from 13-20 August 2016.	Chair-man	end June 2016	Complete. The workshop was held at 2pm on Tuesday 16 August 2016 in Yorkshire, UK.
2016/4	Participants are asked to inform the Chairman if they expect to attend the 2016 EuroSpeleo event.	All	end June 2016	Complete. The PEWG workshop was attended by 13 people.
2016/5	Complete the UIS Questionnaire about Libraries, to register as a 'UIS Documentation Centre'.	All librarians	As soon as possible	Suspended. About five libraries responded, but the Workshop on 16 August 2016 suggested that the Questionnaire should be simplified.
2016/6	Finalise Recommendations for Digital Scanning (Appendix 4)	BMcI, GM, BM, TC.	end June 2015	Partly complete (See 2016/8). A completely new version of Appendix 4 has been provided by Bill Mixon, with thanks.
2016/7	Provide next draft Recommendations for Electronic Publications Exchange (Appendix 5)	BMcI, GM, BM, TC.	end June 2015	Complete. A completely new version of Appendix 5 has been provided by Greg Middleton and Bill Mixon, with thanks.

Actions requested 16 August 2016, from the PEWG Workshop

Action no.	Action	Action on	Target date	Action status
2016/8	Provide recommendations for OCR during the scanning process, for Appendix 4.	BMcI	End Dec. 2016	Complete. OCR is included at Issue 3.
2016/9	Advise if any organisations should be transferred between the First and Second Lists	ALL	End Dec. 2016	Outstanding
2016/ 10	Investigate if Dropbox could be used for holding PEWG information.	PD	End Dec. 2016	Outstanding
2016/ 11	Recommend standard catalogue software	ML	End Dec. 2016	Outstanding
2016/ 12	Discuss the Questionnaire and Certificate for UIS Documentation Centres with Lucas Malafaia	Chair-man	End Nov. 2016	A revised Questionnaire is being distributed and the list of existing Documentation Centres is included in Chapter 4.
2016/ 13	Check and populate the information in Appendices 6, 7, 8.1 and 8.2.	ALL	End Dec. 2016	Several updates are included at Issue 3.
2016/ 14	Reissue the ToR and Appendices at Issue 2	Chair-man	End Nov. 2016	Complete. Documents were circulated on 9 November 2016.

Actions requested 28 July 2017, from the PEWG Workshop

Action no.	Action	Action on	Target date	Action status
2017/1	Provide guidelines for Chapter 7 about scanning to a JPG-file	Peter Matthews	Q1 2018	
2017/2	Continue to check and populate the 5 spreadsheets	ALL	2018	
2017/3	Decide name and status of the PEWG within the new UIS statutes	PM Michele Sivelli	2018	
2017/4	Agree process for creating new UIS Documentation Centres	PM, MS, Patrick Deriaz, Lucas Malafaia	2018	
2017/5	Place Issue 3 documents on the website, after removing email addresses	PM	End 2017	
2017/6	As offered separately, provide a website to enable people to swap spare journals	PM	2018	