



ILLUSTRATION BY THE PROJECT TWINS

HOW TO DIGITIZE YOUR LAB NOTEBOOKS

Converting paper records to digital formats provides secure back-ups that researchers can access from anywhere. **By Julian Nowogrodzki**

When the coronavirus pandemic closed the University of Minnesota in St Paul, plant pathologist Linda Kinkel's laboratory team cast around for tasks they could do from home. They realized there was one job they'd wanted to do for some time: digitizing the team's 30-year-old collection of paper lab notebooks. "COVID really was what made us commit to the digital lab notebooks," says technician Andrew Mann. "All of us being alone and needing access to former students'

experiments so we can write grants and plan our next experiments."

Research groups digitize their old lab notebooks for a host of reasons. Digital records can be backed up so they are impervious to floods and fires, and encrypted to protect them from theft. They require no physical space, and can be used by multiple team members at the same time from different locations. The scanning process makes the text readable, accessible and suitable for archiving; if the software includes optical character recognition (OCR), scanned

typewritten text can also generally be searched – although OCR is not error-free, so the resulting text often needs manual correction.

Some researchers scan notebooks using smartphone apps or physical scanners; others outsource the work to specialized companies. "Digitization is on the increase, especially after COVID," says Jan Cahill, marketing director at Cleardata, near Newcastle upon Tyne, UK, which digitizes books and documents. Lab closures because of pandemic restrictions have highlighted the benefits of having

Work / Technology & tools

documents remotely accessible by every team member simultaneously, Mann says. For Glenn Lockwood, a computer scientist at Lawrence Berkeley National Laboratory in Berkeley, California, digitization was about providing peace of mind. “It just helps me sleep better at night,” he says.

Smart use of smartphones

Mann and his colleagues have taken a no-frills approach to digitizing their old notebooks: they use their smartphones. The collection includes dozens of bound, standard-sized lab notebooks with yellow paper and red covers, and each team member took a few home when the lab closed at the start of the pandemic restrictions. Scanning each page with a smartphone isn't fast, but because every lab member has one, it is efficient: there's never a queue to access a physical device. All it takes is time – a couple of hours per notebook, Mann says.

When selecting a scanning app, some of the most important considerations are the reliability and language specificity of its OCR software, if it has this feature. Even using software with an accuracy rate of 98%, a single typewritten page containing 2,000 characters can still generate around 40 errors that need to be corrected manually. Another consideration is how well the app automatically crops images, and whether you can manually adjust them easily and immediately after taking the picture. One popular choice is Adobe Scan, which offers OCR in 19 languages, including English, Spanish, Japanese and Korean, as well as traditional and simplified Chinese characters. The app is free and available for both Android and Apple iOS operating systems.

Mann uses Apple's free Notes app (iOS only), which does not provide OCR, although it does allow him to crop the resulting images on his computer. One setting takes and saves scans automatically, but by switching to the manual setting, you can crop each scan before saving it to your ongoing file, which is more efficient than doing so later. Other free apps include Microsoft Office Lens and Genius Scan, both of which are available as Android and iOS versions and have OCR. Or users can pay for apps, some of which feature OCR in more languages.

The team saves each lab notebook as a single PDF. Mann has found that the first and last pages are harder to scan legibly using a smartphone, because the spine is large enough that those pages don't lie flat. And the resulting files lack the easy navigation provided by physical tabs in a paper notebook, Mann says.

Heavy-duty digitization

Despite the simplicity of smartphone scanning, Lockwood bought a desktop scanner to drive his digitization project. As a computer scientist, his records have been digital for years, but he kept meticulous notebooks during his graduate work in materials science. “As a

student, I was trained to make sure everything was bulletproof in terms of provenance and intellectual property,” he says. That meant a physical lab notebook with carbon copies, with every page signed and dated. The originals remain with his graduate lab because of institutional policy, but Lockwood kept the unbound carbon copies, which he's carted between apartments for years. He decided to scan them as “an evenings and weekends project” so he could finally get rid of the physical notebooks – a task that is still ongoing.

“It's been kind of nice to go through all of these people's lab notebooks who I've never met. I feel like I know the research a lot better.”

Desktop scanners, or printers that include scanning features, can cost between US\$200 and \$600. Lockwood, who paid about \$200 for a Brother MFC L2750DW scanner with an automatic document feeder, recommends scanning in colour and at the machine's highest possible resolution – in his case, 600 dots per inch (dpi). “There's no point in cheaping out on that stuff,” he says. Some of his notes were in pencil on thin notebook paper and were not legible in lower-resolution scans. Because his notes are handwritten, OCR isn't much use, and the resulting files are large: a 116-page scan of one notebook came to nearly 190 megabytes.

When pages are consistent and even, scans take a couple of minutes, Lockwood says. But taped-in materials and uneven page sizes can complicate the process, making it more manual. “It turned out to be much more labour-intensive than I anticipated,” he says.

Developmental biologist Kelly Smith and her group used a Ricoh MP C4503, a combination photocopier–printer, to digitize protocols and key experiments when she moved her lab to the University of Melbourne, Australia, a year ago, because she had to leave the physical copies at her previous institution. Since moving, however, her lab has abandoned paper notebooks in favour of an electronic system from LabArchives in Carlsbad, California. “Being able to share data and access it immediately is awesome,” Smith says.

Scan and deliver

Scanning companies such as Cleardata, eRecordsUSA in Fremont, California, and Digiscribe near White Plains, New York, provide a third digitization option. Such firms typically provide OCR and ancillary services such as quality control, metadata attachment and confidential shredding of the original notebooks.

For example, eRecordsUSA scans various materials, from historical books and personal documents to magazine back catalogues, says co-owner Pankaj Sharma. The company handles about a dozen projects a year, which average 150 books per project, but it can scan up to 1,500 items per order. Sharma recommends scanning in colour and at a resolution of 300 dpi.

Because eRecordsUSA handles mainly historical documents, it has equipment that is specifically designed for delicate bindings, including a V-cradle scanner that stops the book from opening fully, and an overhead scanner. Every page with a folded, stapled or taped item is scanned twice: once in its original place, and again after the item has been unfolded or turned over. An employee does a page-by-page comparison of the original for quality control. “Most of the books are handwritten,” Sharma says. “We haven't found OCR to be very effective.”

The cost of using such companies varies with factors including the timeline, the notebook's dimensions, number of pages, binding type and whether a notebook has loose or irregular pages. Some firms offer discounts for high-volume projects; eRecordsUSA will digitize a sample book for approval before moving onto the rest, for instance. The company can also process confidential information such as health-care data, financial documents and litigation records, Sharma says. One standard-sized lab notebook with 100 pages would cost \$75–100, he estimates.

Cleardata scans around 5 million images each month in total, including lab notebooks and other items, says Cahill. The company can output scans into any required file format (PDF, JPEG or TIFF) or resolution (the default is 300 dpi), and every document is checked by two people for quality control. The firm also has a document collection and boxing service. When digitization is complete, notebooks “can either be returned, stored in our secure archive facility or destroyed using industrial shredding equipment”, Cahill says.

The cost ranges from £25 to £200 per book, depending on the item's characteristics and on whether the scans are in colour or black and white. Cleardata's minimum order is £500 (US\$645).

For Mann, scanning old lab notebooks has provided an unexpected benefit. He's new to the group, having started in February, just before the pandemic closed his lab. Reading through them has revealed insights that he might not have got from the team's papers. “It's been kind of nice to go through all of these people's lab notebooks who I've never met – actually go through every single page,” he says. “I feel like I know the research a lot better.”

Julian Nowogrodzki is a journalist in Boston, Massachusetts.