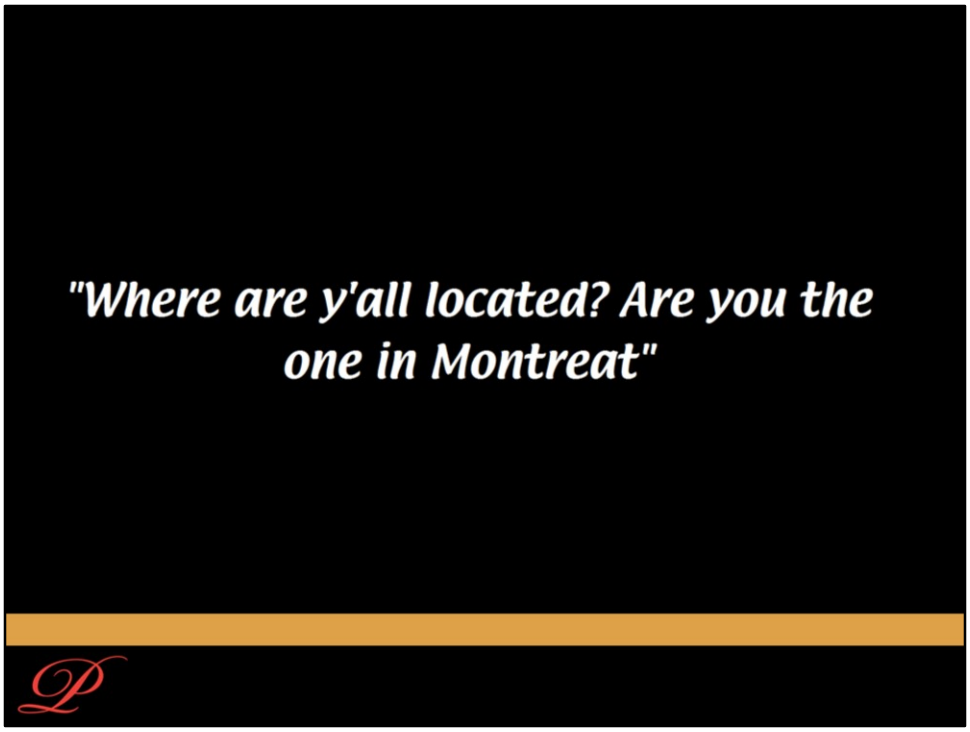
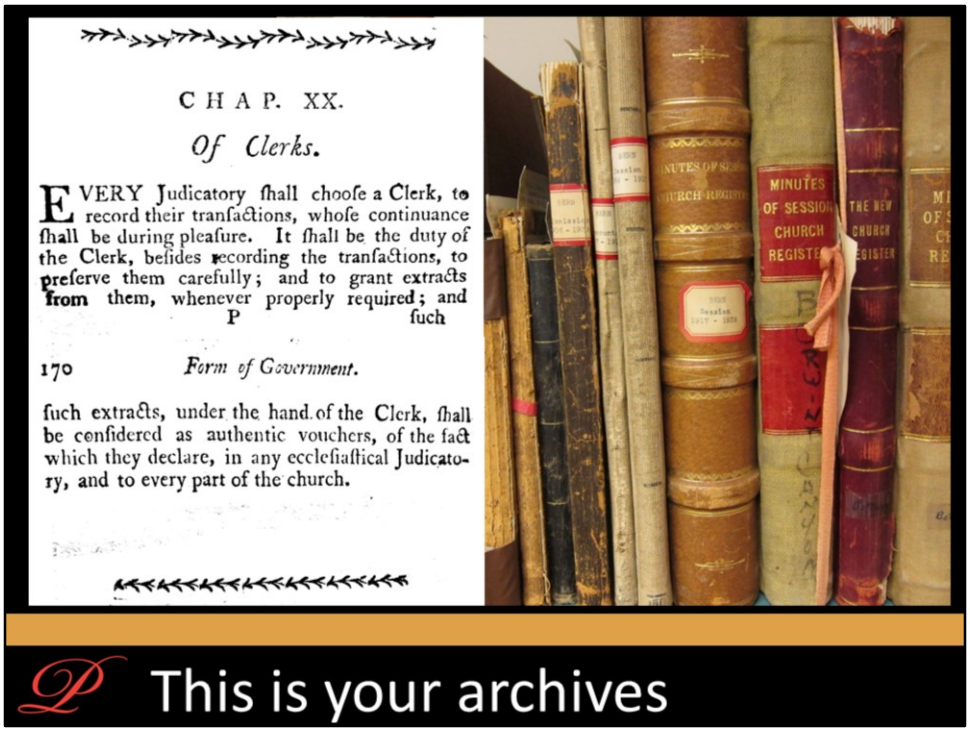
This is a soup-to-nuts guide to preserving records of congregations, structured as an FAQ, with intertitles using the “It’s Always Sunny in Philadelphia” font.

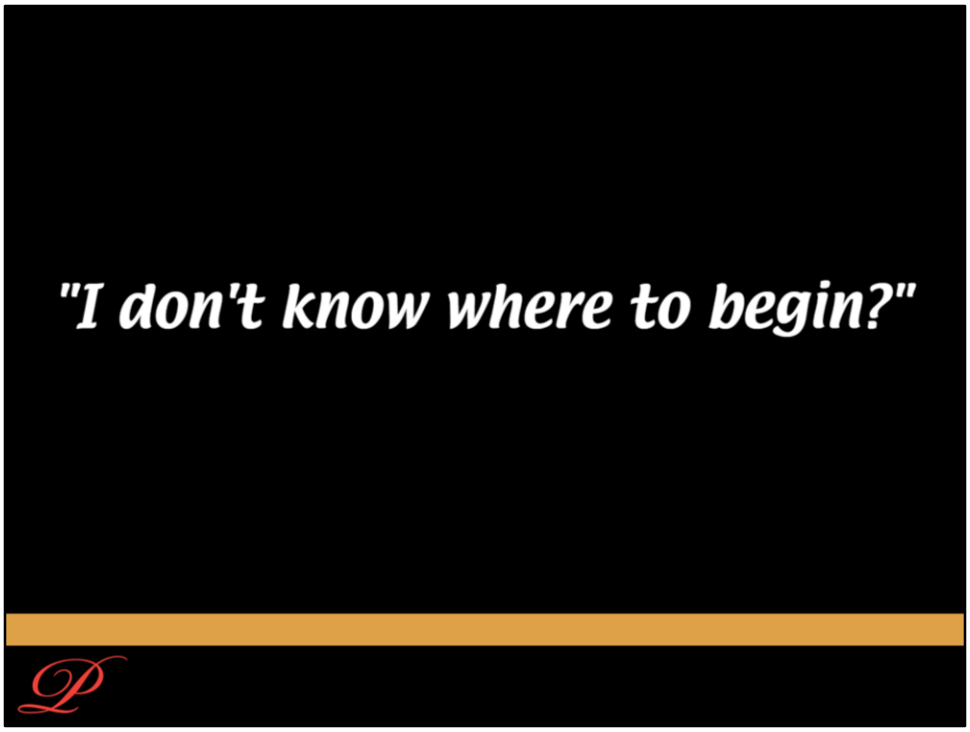

"Where are y'all located? Are you the one in Montreat"

Nope, we are in Philadelphia. We are the Presbyterian Historical Society, organized in 1852 as the national archives of the PCUSA (Old School); we have more than 34,000 cubic feet of archival records, dating to the first Presbytery in 1706.

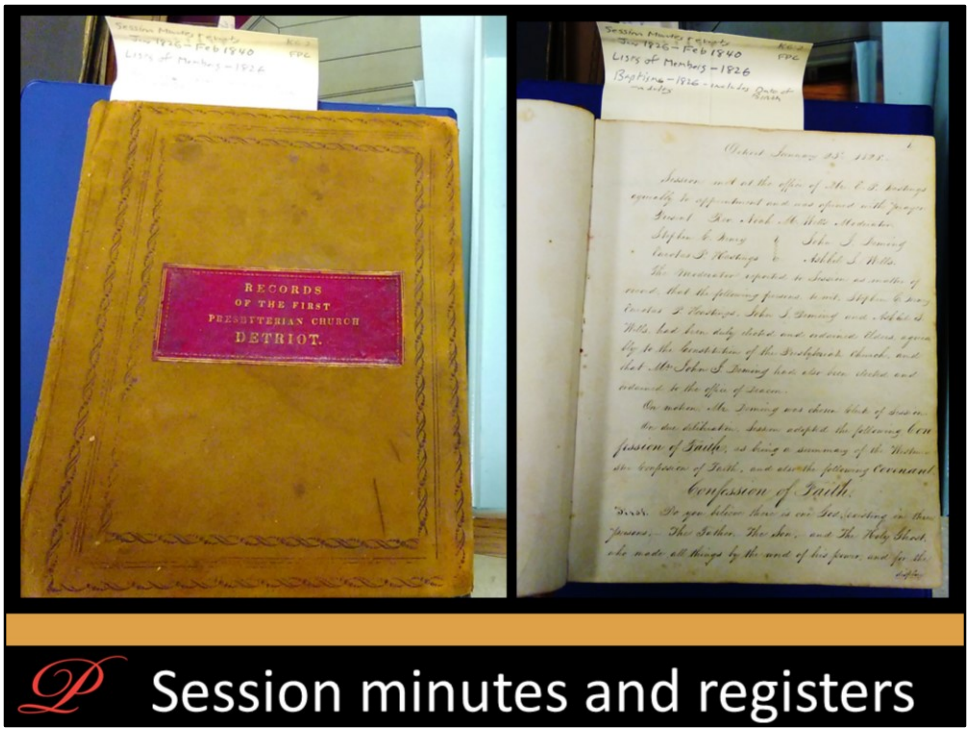
We bring in 700 cubic feet of records every year, from congregations, mid councils, and the national agencies. We field 3600 research requests every year and researchers put in 1000 hours of study into primary sources in our reading room every year. Very often we serve legal and administrative needs in real time: proving people were baptized, proving employment or education, helping to settle property disputes. We bring in 30 feet of records from 20 congregations every month. There have been about 40,000 Presbyterian congregations ever in existence; we hold records of about 12,000 of them. The rest are held by active congregations, by other institutions, or are unaccounted for. If by some accident we could account for all of them, we would be bringing in to the collections another 45,000 feet of records.



You have the largest and oldest denominational archives in the United States for a reason: recordkeeping is an essential part of your belief system. Other friends of ours joke that Presbyterians are people of the books: the Bible, the Book of Order, the Book of Confessions, and Robert's Rules. You maintain (relatively) orderly meetings and record (relatively) orderly minutes in order to give a full account of yourselves to each other. Recordkeeping, reinforced by annual minutes review, just like your gatherings, is an expression of the connectional nature of the church. The 1789 General Assembly codified this; PHS's stacks execute this.


"I don't know where to begin?"

Very often, when dealing with records of a congregation, clerks of session and church administrators may be confronted with a large body of heterogeneous material. In this kind of scenario, I've found that the best thing to do is to know the body you want to pull out of the wreckage: in our context, first look for session minutes, and registers.



Session minutes and registers

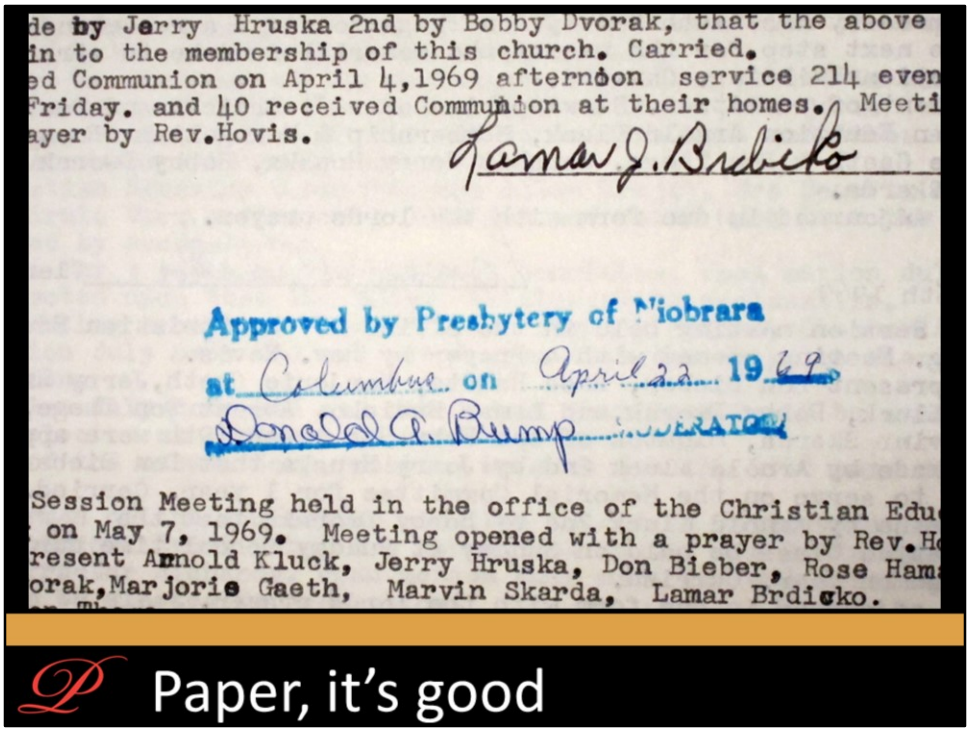
Session minutes and registers (pictured here, the first volume of minutes of the First Presbyterian Church (Detroit, Mich.) from 1826) are the vital records of the congregation. They make your organization real and accountable to its constituents; maintaining them is in the Book of Order.

There are also materials that we consider permanent, and will bring on deposit, but which may alternately carry more power locally. Minutes and registers may formally constitute an organization, but it's the creation of newsletters, histories, photographs, scrapbooks, and other material that documents the social presence of the congregation. Pictured here is a proud girl scout from Bay Ridge United Church (Brooklyn, N.Y.)

A constituent at an event I was at, upon hearing I was from the national agency offices of the PC(USA), asked me "What is it that you're doing for the impoverished children, the children who don't have anything to eat." I was put a little bit off my game, but I came up with something like: Not a thing. Your church does that work in the world. I work for the archives, and it's our job to remind you of that work, to remind you of who you are supposed to be. This is the discovery of the Baby Moses, part of a Children's Day pageant at the Bohemian Brethren Church of Omaha, Neb. Records here speak to contemporary concerns – how has the church spoken to young people; what is the purpose of a language-based presbytery. We recently fielded requests from synods regarding the history of the so-called "language" presbyteries – groups of Welsh, Bohemian, Magyar, and other European Presbyterians which came into the PCUSA in the late 19th and early 20th centuries, and kept minutes in a non-English language. We bear testimony to the historic capacity of American Presbyterians to welcome immigrant populations to the table.

Much of the rest of the material that you find will have been created in the course of the church doing its business. Most financial records – paid bills, bank statements, expired insurance policies – can be securely shredded after the period of their usefulness has ended.


"Wait, you're saying I should use paper for permanent records?"



Paper is the cheapest reliable steady-state carrier of text over the long term. Best practice, as far as I've seen is what the Presbytery of New York City does. They publish their minutes to the web in PDF, store PDF/A copies on their local network and in the cloud, and print an official version to acid-free paper, storing the paper in the classic three-post Cokesbury binder. When the binder is full, they have the 300 or 400 pages hardbound, and they deposit that with us.

In the very near term, committing text to acid-free paper helps with some of the intellectual problems of born-digital recordkeeping: bulk and authenticity. Having a solid-state version of text means you're less likely to throw the whole kitchen sink into your minutes; you're reducing the bulk of your records to essential functions, and making them comprehensible to future users. Also, having one version of your minutes that receives a stamp from a reviewing body clarifies the relationship among copies; you're reinforcing a source-surrogate relationship.

(As an aside, once you expand the preservation horizon to 100 years, printing minutes to paper gains an interesting advantage over PDF/A: its carbon footprint. A round figure for cloud storage of 1.5 GB of data – about what the average congregation digitization project, some 2000 pages, takes up in PDF/A for us – is 1.5kg of CO_2 per year: meaning 150kg over 100 years. The same 2000 pages, or 18kg of acid-free paper, has a carbon footprint of 2kg per 1kg, or 36kg for the century.)

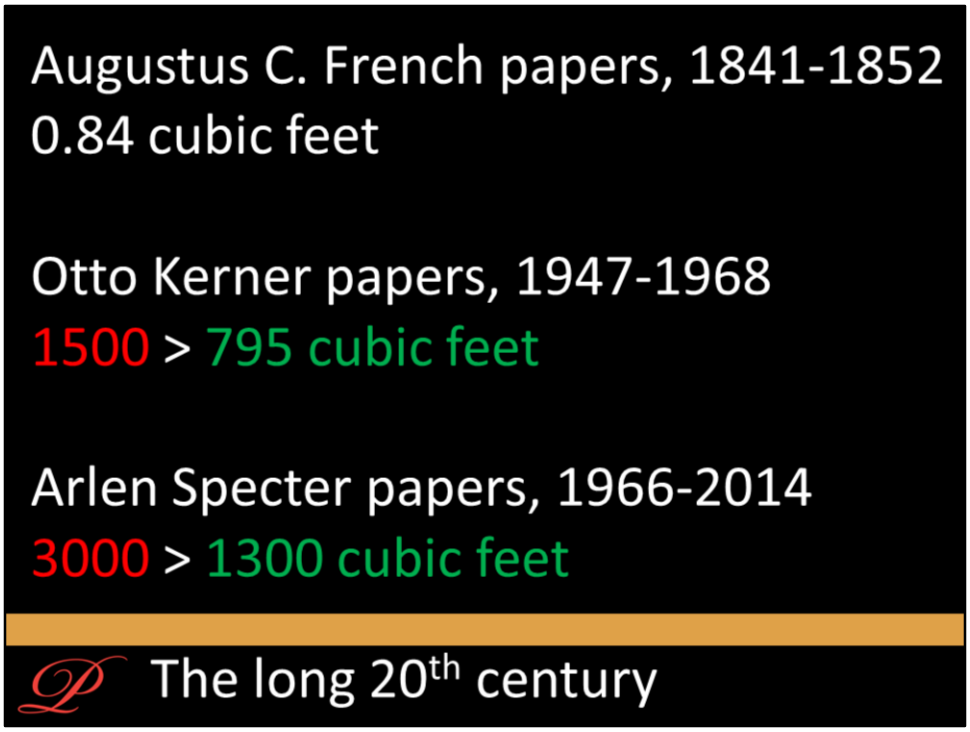
We continue to recommend acid-free paper as the preferred carrier of text for the very-long term. You will continue to live in a hybrid paper and born-digital environment because of the different virtues of paper and digital formats as carriers of text.

Well, no, in part because we're still waiting on the rest of the past to catch up with us.

Humans are always going to do some things manually, in analog terms, first of all, and the demise of paper as a recordkeeping medium was first written about in Business Week in 1974. That was 45 years ago. Let me talk a little about how we prepare original material to be digitized, and how we handle born digital content.

Before an archives becomes digital (or before a digital record enters the archives) it comes to us in an original state. There is no substitute for the record in its original condition; our work in preparing records to be comprehensible in a digital platform involves a kind of displaced office labor.

In the middle of the twentieth century, like any bureaucracy, the Presbyterian Church had dedicated administrative staff, who maintained paper filing systems. Over the course of the 1980s and 1990s, organizations began to eliminate secretarial-skill positions (overwhelmingly female) for computer-science positions (overwhelmingly male). Our presumption was that computers would **sort and file the things we create for us**. It's only in the 2000s, after enough CEOs had lost all the contents of their desktops, that corporations begin to conceive of "knowledge management," and began to hire information scientists to run enterprise-level KM platforms, all to replace what was once done by a bank of women in central filing. (Image, US Census Bureau, 1940)

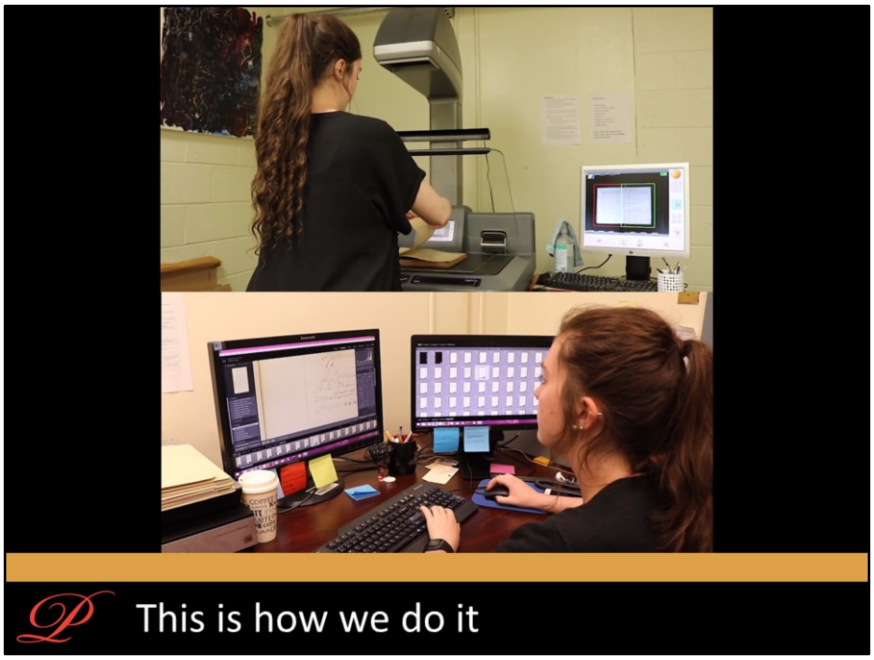


Thank God for the women in central filing, because we still haven't caught up with the efflorescence of 20th-century bureaucratic records. As the 1980s and 1990s progress – apart from disregarding the importance of secretarial labor – texts and images become easier and easier for individuals to create. **Far from eliminating the mediated environment, our technology *facilitates* the creation of media.** The course of the 20th century is a series of innovations in increasingly personal publishing.

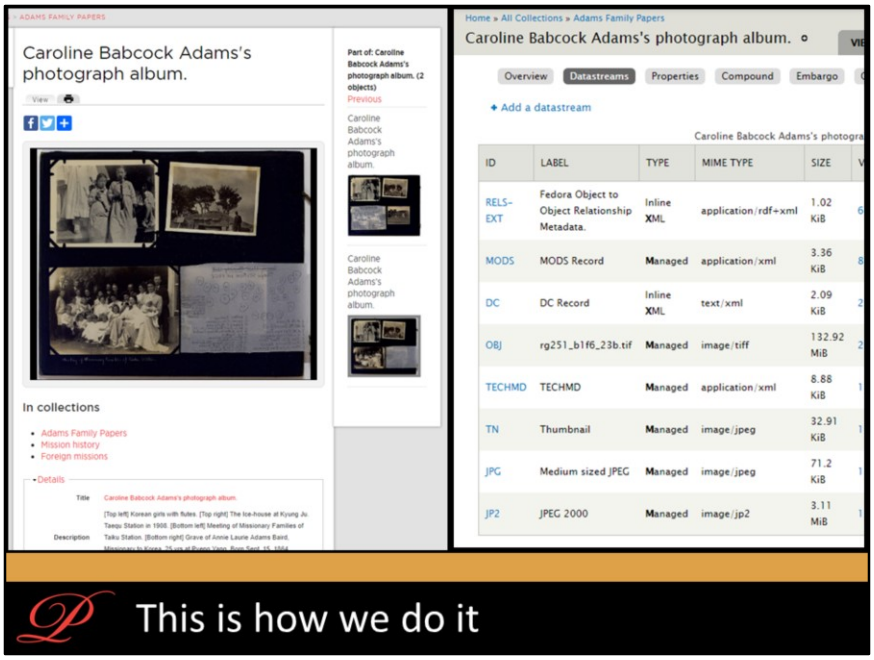
The renowned archivist Gerry Ham wrote in 1984 about the "bulk and redundancy" of 20th century paper records, arguing for archivists to dedicate themselves to fierce appraisal of collections, pointing out that the average 19th century governor of Illinois left behind 10 cubic feet of records, while at least one 20th century governor left behind 1500. (Gerry was responsible for cutting Illinois governor Otto Kerner's papers in half.) The remaining Arlen Specter papers, made visible last year at Pitt, are double the size of Kerner's, and began life as 3000 cubic feet. Appraisal is essential to our work.

Broadly speaking, the collections here dating to prior to 1972 come with the labor of administrative professionals embedded in them. The files are clean and comprehensible. Our main value-add is in grouping like with like. We use an advanced and secret archival technique called “stacking” to achieve this.

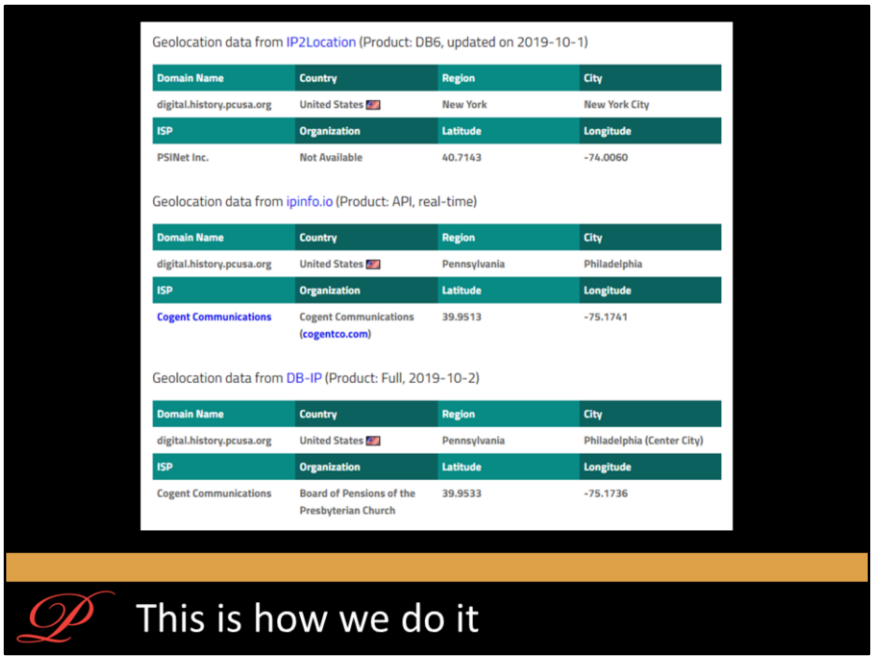
Processing, to my mind, is what we call all the steps necessary to make a heterogeneous collection digitizable. We add comprehensibility. Sometimes this means consolidating the original file structure. In order to spare the work of scanners, we also habitually remove large metal fasteners. Deduplication and deletion of temporary records also – where it hasn't been covered by our records management program – happens at processing.



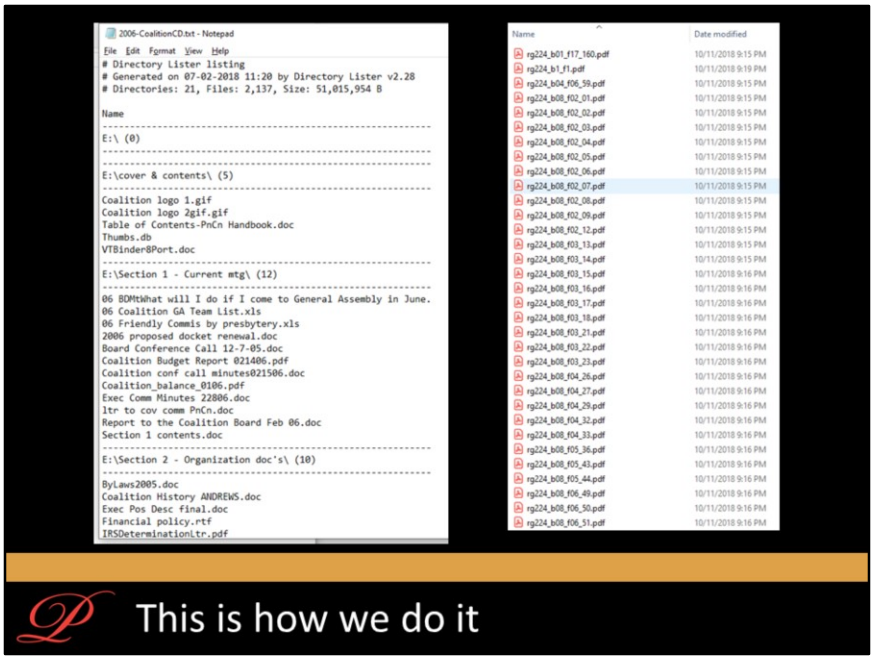
People do the work. Our digitization services started up in 2014; our archives technicians scan more than 100,000 pages of text and individual photographs every year. (We also digitize about 60 hours of audio and video every year.) Just about everybody tops out at 300 pages imaged per hour. Color correction, deskewing, error checking (skipped pages, hands in the frame, etc.), running optical character recognition on text, compiling documents to PDF, all take between 3 and 5 hours per hour of scanning.



Material that we can publish – which we own outright, is outside of copyright, and so on (feel free to take me on a sidebar about most archives' risk-averse copyright policies) – we put in our digital collections platform, called Pearl. Like all digital collections, this is a digital repository with a web front-end. The web content management system is Drupal, the repository is Fedora. The repository automates functions that we once did by hand – it creates derivatives of master files, generates technical metadata, manages object relationships, and runs checksum audits.



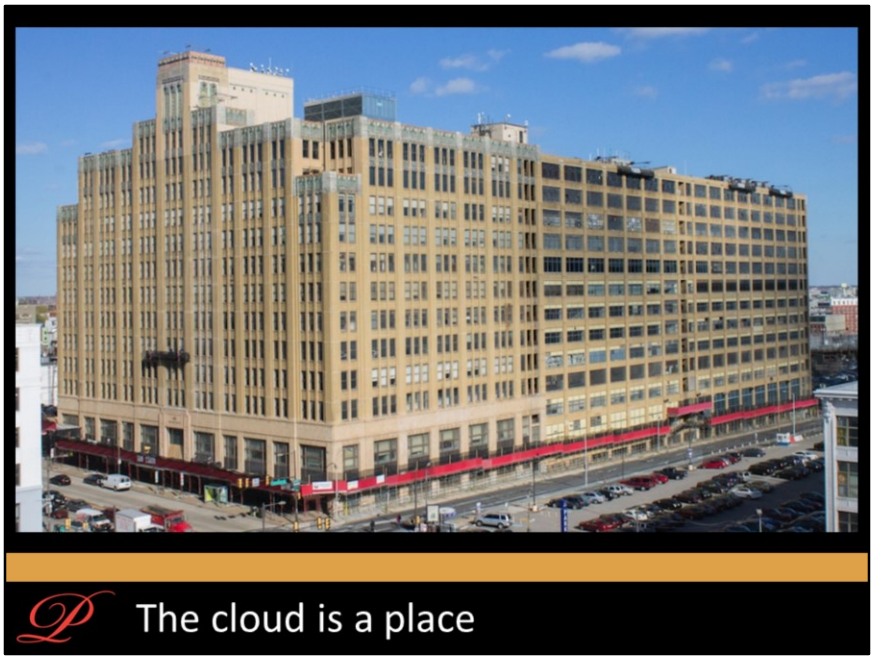
The servers that Pearl's application stacks sit on are run by Cogent Communications, a subcontractor of the Board of Pensions



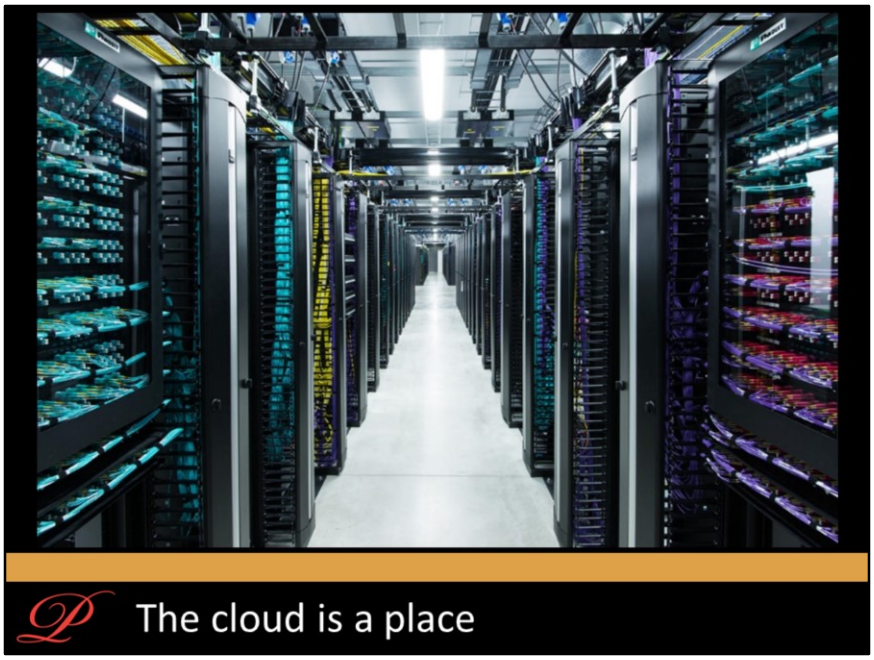
We also preserve select born-digital records created by national agency offices, and the born-digital portions of personal records collections. We remove files from media, “print” a directory list, and move the contents to our network attached storage array, and flag them for processing, putting them into their own share of the NAS. Just like with physical files, we appraise out content of temporary value, or things that are actively harmful (we weed individual user’s music files, we flag and delete executable files and thumbs.db files, etc.). We normalize files – mostly this means converting word processing files into PDF/A. And we use a Python script developed by the Library of Congress, BagIt.py, to create separate metadata files, and checksum files for the records. We move all of this to a new share of the NAS.

We have about 50TB of content in our dark storage; this is backed up locally to two other storage arrays, and it's synced to a corporate Dropbox account. Like other archives, there's a whole array of external services which we rely on to proffer content to our users: we use the Internet Archive to host our web captures; our website, databases, and digital collections all occupy space on Amazon Web Services servers; likewise, Dropbox is a major client of AWS; we use an audio transcription service called Trint; all of this is out in the cloud.

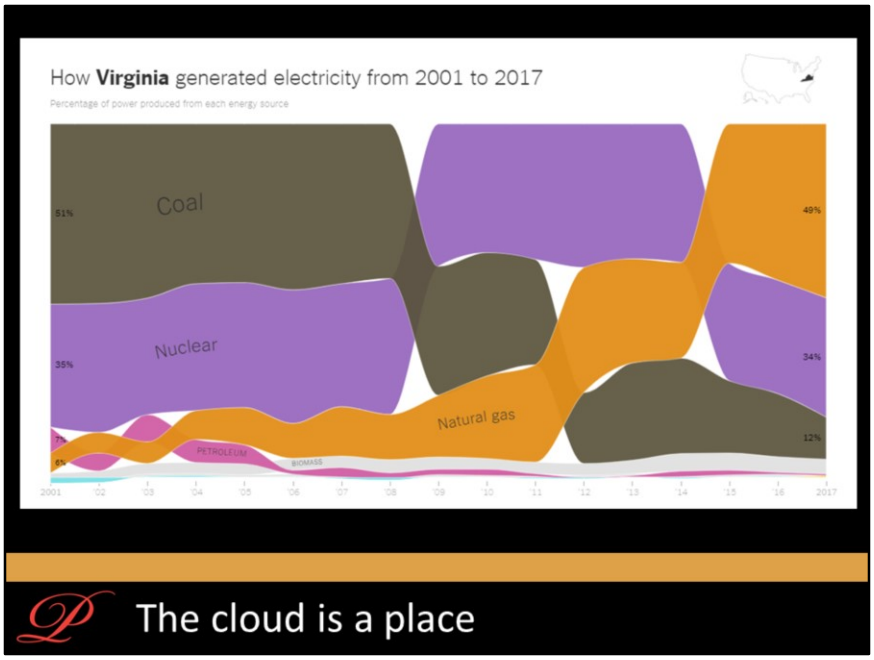
Well, there's space and there's space. Here are some of my thoughts.



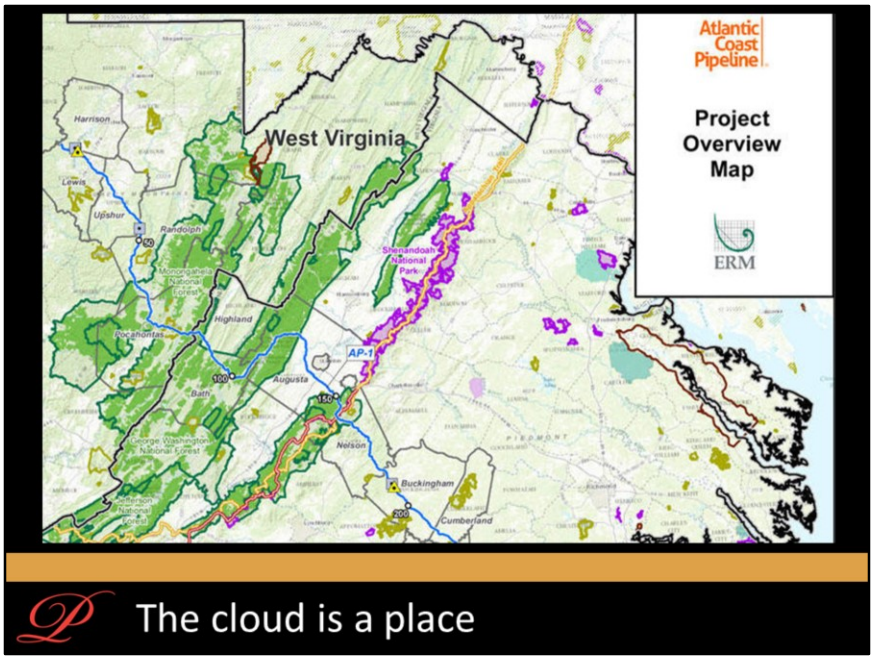
The cloud is a place. The physical plant of the internet is huge. Here's one little chunk in Philadelphia. Developers of the 1930 Terminal Commerce building have turned the "building with indestructible bones" into a "carrier hotel," a node for telecommunications companies https://hiddencityphila.org/2015/11/indestructible-at-broad-and-callowhill/



In a sense, this whole presentation is dedicated to one patron of mine, who maintained that we as a society have figured out how to handle archives, and it's to scan everything and dump it in the cloud. "Computers may change, but digital data in the cloud is actively maintained." So what is the cloud? The cloud is remote data centers, accessible by the internet, which are giant warehouses full of servers, giant versions of our network attached storage. Data is written redundantly across the array, and whole data centers are mirrored in geographically distinct places. Much of the east coast data storage is in Northern Virginia, where Amazon purchases electricity on futures contracts from Dominion Power. (https://www.greenpeace.org/usa/news/greenpeace-finds-amazon-breaking-commitment-to-power-cloud-with-100-renewable-energy/)



Virginia's power mix is about one-half natural gas, one-third nuclear, most of the rest coal. (https://www.nytimes.com/interactive/2018/12/24/climate/how-electricity-generation-changed-in-your-state.html)



Duke Energy and Dominion have partnered to push a natural gas pipeline through West Virginia and the Shenandoah Valley, in order to hook up fracked natural gas from the Midwest into the east coast. Just like the Dakota Access Pipeline, there has been opposition. The Fourth US Circuit Court of Appeals ruled recently that the pipeline cannot bisect the Appalachian Trail. CEOs of the companies are sanguine – they anticipate that the US Supreme Court will grant cert in October or November of 2019, and rule in 2020. Given the composition of the court, fossil fuel interests will win. (https://www.wfae.org/post/despite-court-rulings-atlantic-coast-pipeline-still-track-dominion-says#stream/0) There is a direct relationship between Amazon's thirst for electricity, and Dominion's need to bring fracked natural gas on line. [UPDATE: The Supreme Court ruled for Dominion in the spring of 2020, but the ongoing glut of supply in oil and gas markets made Dominon cancel the Central Virginia Pipeline. The Mountain Valley Pipeline, through the Blue Ridge, is still in progress.]

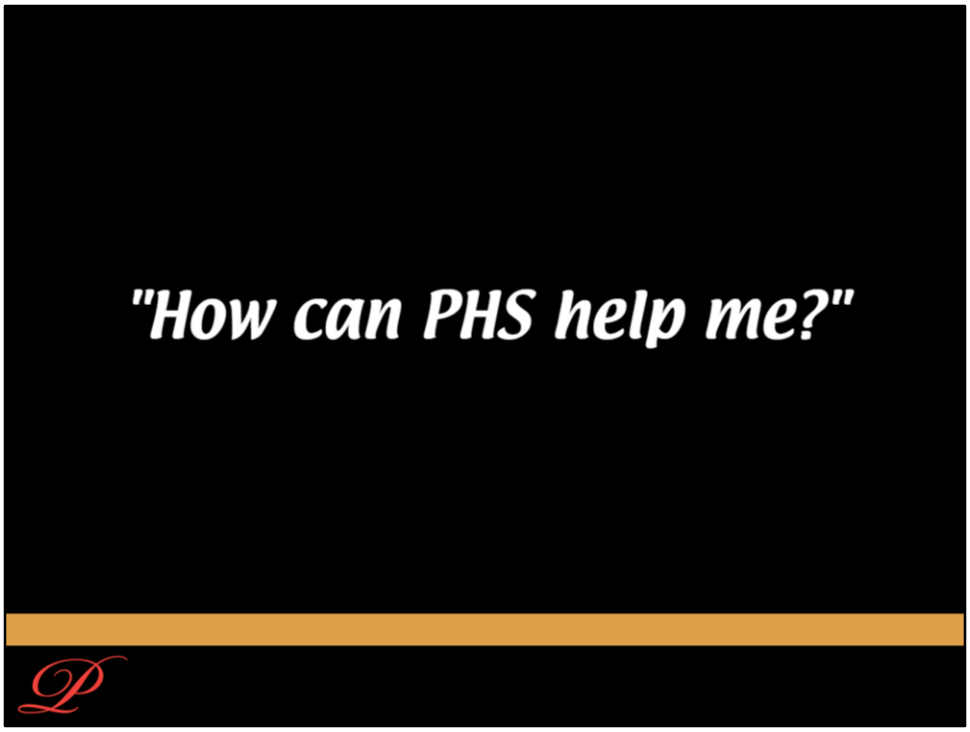
A 2014 study by the Natural Resources Defense Council estimated that data centers in the United States would consume 140 terawatt-hours of electricity per year, producing nearly 150 million metric tons of carbon emissions. The equivalent of 50 coal-fired power plants. (https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf) The study found that one of the obstacles to energy efficiency in data centers was server utilization – from 2006 to 2012, data centers operated at between 12 and 18 percent use. A 2017 EPA study (https://www.epa.gov/sites/production/files/2017-11/documents/2017-06_0.pdf) recommended that decommissioning "zombie servers" could shrink energy use by as much as 45% in data centers.

Common estimates find that the global data industry consumes between 3 and 5% of all electricity, and emits CO2 on par with the airline industry. (https://www.cnn.com/2019/09/10/asia/china-data-center-carbon-emissions-intl-hnk/index.html)

All of this is to suggest that we may have overgrown our capacity for remote digital storage. That irrational exuberance has had an effect on the climate.

This is not by way of saying that we should disconnect our services and stop being online (although, ask me in a sidebar about the television and movie industry's carbon footprint), just to emphasize that computers are not ephemeral, they are part of an energy-consuming infrastructure, and that anything that consumes energy can be measured alongside anything else.

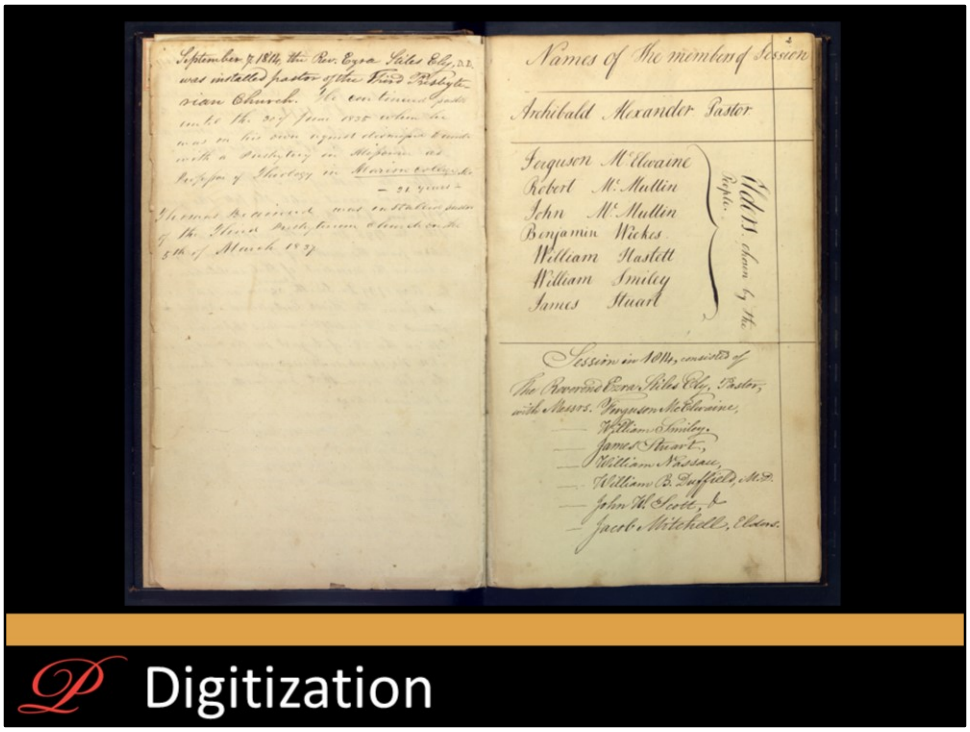
One of the interesting attributes of maintaining paper as a carrier of text is that (https://www.history.pcusa.org/blog/2019/09/our-carbon-footprint-archives) much of the carbon cost of maintaining paper is fixed: paper, once cut, distributed, and printed on, can sit on a shelf for 500 years. Our HVAC will give off an estimated 36,800kg of CO_2 per year, or about eight automobile drivers' worth. The total text in our holdings, if digitized and stored remotely, however (depending on a lot of variables) could amount to an annual emission of more than 700,000kg of CO_2.



Our chief services are deposit and digitization

Records of historic value document the decision-making processes of the institution, and the expression of those decisions in policy and action. These include minutes, registers, publications, newsletters, reports and photographs. We will happily take on deposit the original records of active and dissolved congregations and the original records of mid councils.

What do you do with paper records which are no longer frequently referred to – session minutes, registers, deacons' minutes, trustees' minutes, records of women's groups, historic photographs? Deposit them with PHS. Congregations can have them back for display or for writing a church history, with a letter from the clerk of session. Presbyteries can have records returned with a letters from the stated clerk. Simple requests, such as baptismal or marriage attestation, can be referred to us to answer. We have a secure, climate-controlled facility running the length of Lombard between Fourth and Fifth. Send original records here.



We image about 100,000 pages of original records every year, at cost for PC(USA) congregations. Our overhead planetary scanner shoots images at 600 dpi up to 16”x24” in full color. We compile those pages to PDF/A, and store those documents securely in our local storage array. We provide congregations with plain-vanilla PDF versions of their records.

Our chief services are deposit and digitization

Collection emphases

We are renewing our efforts to document Black lives, work, and witness in an increasingly multicultural church—from the organization of the First African Presbyterian Church in 1807 to the election of the first African American stated clerk of the PC(USA) in 2016.

We are bringing human and capital resources to bear on collecting records of the Black Presbyterian experience--both the personal records of servants of the church, and the original records of Black congregations. PHS seeks to represent in the archives the Black throughline: the integral presence of African Americans in what the authors of the historical volume *Periscope* called a “historically racist ecclesiastical body.”

(Above, the Westminster Neighborhood House in Watts, Los Angeles, 1980s. The only building on its block not burned down during the Watts rebellion in 1965.)

In the wake of the 2017 hurricanes which decimated much of Puerto Rico, we were called upon to dedicate resources to the preservation of the records of Puerto Rican Presbyterians. (Above, the rhythm club of the day school at Iglesia Presbiteriana La Marina (Mayaguez, P.R.) In 2018, we visited San Juan, Arecibo, and the Synod camp El Guacio, and we brought back PHS's first archival collection created by Puerto Rican Presbyterians

Last year, we visited Puerto Rico again, and digitized more than 6000 pages of the session minutes of Iglesia Presbyteriana en Rincon (P.R.)

Thanks for coming.