

FRASER[®] Digitization Standards

It is the intent of the FRASER team of the Federal Reserve Bank of St. Louis to use imaging standards that produce the highest-quality image (for both optical character recognition and user legibility) with the smallest possible file size and investment of scanning time.

FRASER scanning standards vary by item type. In general, these standards apply only to items owned by the Federal Reserve Bank of St. Louis. Items scanned while on loan from other institutions or digitized in partnership with other institutions may use these standards or may have custom scanning parameters agreed upon by all parties.

Additional information about the standards that inform FRASER practice can be found in the Technical Guidelines for Digitizing Cultural Heritage Materials from the Federal Agencies Digitization Guideline Initiative:

http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf

Contents

Scanning standards.....	2
Archival items and collections.....	2
Rare books	2
Common publications	2
Textual grey literature.....	3
Data-intensive or tabular grey literature	3
Photographs.....	4
Born-digital documents.....	4
Image processing standards	5
Documents	5
Photographs.....	5
Metadata standards.....	6
Filenames	6
PDF metadata.....	6
Photograph metadata.....	7
FRASER website metadata.....	7
File storage standards.....	7

Scanning standards

All types of documents should be digitized at the highest possible resolution suggested for the document type and positioned to require the least amount of post-scanning processing.

However, situations may arise that require documents to be cleaned or otherwise manipulated to meet the FRASER standard.

Archival items and collections

Archival collections are those that consist of “Materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control; permanent records.”¹ FRASER’s archival collections include those owned by the Federal Reserve Bank of St. Louis, those loaned by other institutions for FRASER digitization, and those digitized by other institutions for the purpose of depositing the digital copies in FRASER. In general, the base standard for FRASER archival scanning is:

- 400-dpi 24-bit color, single-page TIFF, uncompressed or using lossless compression
- Full-page image captured with all page edges visible
- All notes and marks on a page captured
- All pages (including verso²) with visual information captured, even repeated pages and blank forms
- Blank reverse sides of pages and pages that are blank except for a watermark or maker’s mark of the printing or paper company may be omitted
- Prior to scanning, photographs should be taken of any archival boxes, making sure to capture the box label and the original organization of the folders within the box (if any). Folders should be scanned with the folder closed; then if necessary, they should be scanned with the folder open to capture all text on the folder label. If no original label exists for the box or folder, this should be noted in the project notes.

Rare books

For FRASER, rare books are defined as bound items published before 1900. Books with historically significant markings by the author or a previous owner should be treated as archival items. The base standard for FRASER rare book scanning is:

- 400-dpi 8-bit grayscale, multipage TIFF, uncompressed or using lossless compression
- Full-page image captured with all page edges visible; this may be cropped to page size in post-processing

Common publications

Common publications are those items that are mass-produced for general access and use. Common publications are produced by commercial publishing houses, are easily found in

¹ Society of American Archivists definition: <http://www2.archivists.org/glossary/terms/a/archives>

² Society of American Archivists definition: <http://www2.archivists.org/glossary/terms/v/verso>

libraries, and may be available for purchase. For FRASER purposes, books published before 1900 should be considered rare books and not common publications.

Common publications with paper in good condition are disbound (have the spines cut off) and are scanned via automatic document feeder (ADF) scanner. Common publications with paper in poor or delicate condition may be disbound and scanned on a planetary or flatbed scanner, or scanned as a bound book on a planetary scanner.

The base standard for FRASER publication scanning is:

- 600-dpi, bitonal, multipage TIFF, uncompressed or using lossless compression, for publications that are more than 85 percent black or dark text on white or light background; individual pages with photographs, multicolor illustrations, or other information that would be insufficiently represented with a bitonal image should be scanned in 600-dpi 8-bit grayscale or 24-bit color, whichever most accurately reflects the original
- 400-dpi 8-bit grayscale multipage TIFF for publications that contain significant amounts (more than 15 percent) of black-and-white photographs, single-color illustrations, charts, or graphs; or for publications that are printed in color or on colored paper that produces low-contrast images when scanned bitonal. Individual pages for these documents that contain color photographs, multicolor illustrations, or other information that would be insufficiently represented with a grayscale image should be scanned in 400-dpi color
- 400-dpi 24-bit color multipage TIFF for publications that contain significant amounts (more than 15 percent) of color photographs, multicolor illustrations, or other information that would be insufficiently represented with a grayscale; or for publications that are printed in color or on colored paper that produces low-contrast images when scanned in grayscale
- Page images should accurately reflect the size and orientation of the original page

A note on multipage TIFFs: The TIFF format becomes corrupt at file size of approximately 2GB. An uncorrupted TIFF generally has a maximum page count of 600 pages of bitonal (black-and-white) scans, 100 pages of uncompressed grayscale scans, and 45-50 pages of uncompressed color scans (depending on page size). Because this limitation generally causes pages scanned past the 2GB file size limit to become lost, it is recommended that scanning to multipage TIFF be done conservatively, and that users err on the side of fewer rather than more pages in a multipage TIFF.

Textual grey literature

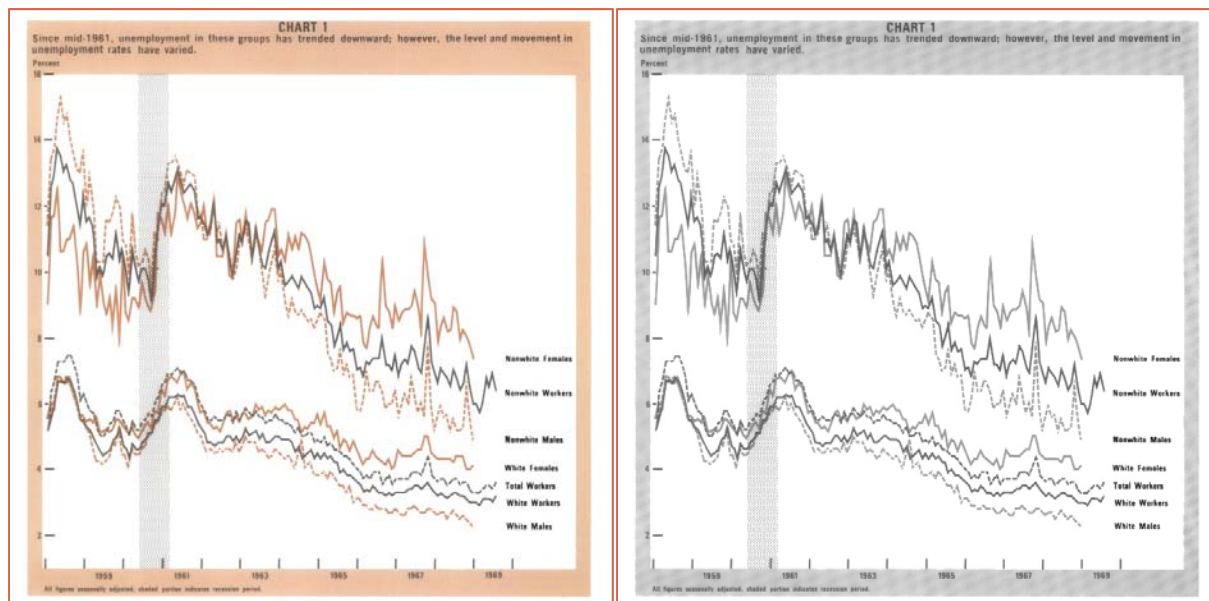
Textual grey literature consists of those items produced by institutions whose primary focus is not publishing, including but not limited to the Federal Reserve and the U.S. government, that are predominantly made up of textual information. Examples include working papers, scripts for or transcripts of speeches that are made available to the public, printed hearings and legislation, and annual reports. The standard for FRASER scanning of these documents varies by the content of the items.

Data-intensive or tabular grey literature

Data-intensive or tabular grey literature consists of those items produced by institutions whose primary focus is not publishing, including but not limited to the Federal Reserve and the U.S.

government, that are predominantly made up of numerical information, particularly that in table form. The base standard for FRASER scanning of these documents is the same as that for common publications, with the following caveat: Two-color graphs or charts may appear to be accurately scannable as a bitonal or grayscale image but lose crucial information when color differentiation is removed (see illustration below). When in doubt, it is recommended that these images be scanned in 400-dpi 24-bit color.

The following example shows the difference between color and grayscale versions of the same image. Because the color, gray, and black lines are of similar weight, the image becomes ambiguous in grayscale.



Photographs

FRASER scanning standards for photographs are:

- 600-dpi 24-bit color single-page TIFF. Photographs are always scanned in color regardless of the color profile (full color, black-and-white, sepia) of the original.
- Any photograph with original writing or other meaningful markings on the back side should also have the back side captured.
- If the photograph is from a collection or archive and has identifying information either on a label on the back of the photograph or on a photograph sleeve, that information should be stored in the file metadata, but this information does not require image capture.

Born-digital documents

The FRASER collection includes born-digital documents from a variety of sources. Documents produced by the Federal Reserve Bank of St. Louis are saved in their original format (PDF, Microsoft Word document, HTML file, and so on) and a copy is converted into 300-dpi resolution PDF files with FRASER-standard filenames ([see below](#)). Documents produced externally are converted to 300-dpi resolution PDF with FRASER-standard filenames. Digitally secured or verified PDF files are left as-is but their filenames may be changed to conform to FRASER standard.

Image processing standards

After the document has been scanned, the image may need additional processing before posting. Ideally, most resources should be scanned at a level that requires the least amount of processing; however, when it is required, the following guidelines should be followed:

Documents

For common publications and some grey literature, FRASER standard is to strive for a born-digital look. To this end, the following modifications are made to page images:

- Correct page orientation
- Deskew
- Remove page borders
- Remove stray marks
- Center image
- Standardize page size (by crop or expansion)
- For color and grayscale images, background colors are digitally “dropped out” where it does not interfere with the readability of the text (see Appendix II, example 1)
- Some projects are scanned in grayscale or color and then converted to bitonal in editing, depending on the particular project needs

For archival items, little to no editing is done, though page orientation is always corrected.

Rare books may be more or less processed, depending on the particular volume and project needs.

Born-digital documents are left unedited.

See Appendix II: Image editing examples for examples of edited page images.

Photographs

Photographs scanned carefully should require minimal editing. FRASER standard is to only crop edges, straighten/deskew, and enhance color (using Adobe Photoshop’s automatic correction features) where necessary.

Metadata standards

Filenames

FRASER has historically used a wide variety of filenames depending on the type and origin of the scanned document. Current practice for newly scanned documents that continue a legacy periodical, series, or multipart already on FRASER is to match the existing filename pattern. Certain document types may have unusual or particular standards; these are documented separately.

Scanned documents are given a filename that represents the original item scanned, with date, volume, and/or other unique identifier appended as necessary, separated by underscores.

Standards for current document naming are:

- No spaces or special characters except underscores and dashes; underscores are preferred
- Use all lowercase letters.
- Separate dates with underscore and formation YYYYMMDD where possible (e.g., filename_20151228.pdf)
- Spell out abbreviations when possible to eliminate confusion or jargon.
- If a filename within a folder will be the same as another, differentiate with an underscore and a lowercase a, b, c, d, and so on. This is most common for scans of very long books that need to be split into multiple TIFFs and for items issued on the same date that have no other unique identifier.
- Periodical issues should always have a date in the filename (e.g., the FDIC annual report for 1936 will be named fdic_ar_1936.tif)
- Books should be named [author's surname]_[short version of title, separated by underscores]_year of publication (e.g., cannon_clearing_house_loans_1910.pdf)
- Legislation should be named with a shortened version of the law's title, with year if necessary: real-estate-settlement-procedures-1974.pdf; revenue-act-1861.pdf; securities-exchange-act.pdf (Note that legacy legislation files used dashes instead of underscores, so this practice is continued.)

Generally, a scanned image will produce a minimum of three files: the highest-quality, unprocessed TIFF image file, which is given the suffix “_orig”; the processed or edited TIFF image file, which is saved without the suffix; and the PDF, which should match the filename of the edited TIFF. For instance, scanning the 302nd *Bulletin of the Bureau of Labor Statistics*, published in 1922, produces three files:

- bls_0302_1922_orig.tif
- bls_0302_1922.tif
- bls_0302_1922.pdf

In this example, because there are more than 999 *Bulletins* total, a four-digit number is used instead of the base three-digit “302.”

PDF metadata

Basic PDF metadata used by Adobe Acrobat includes four editable fields: Title, Author, Subject, and Keywords. FRASER practice is always to include Title and Author; Subject and Keywords may or not be used, depending on project needs.

For periodicals and multipart items, we strive to include the unique identifier (such as issue date or number) in the metadata title. If this is not feasible, the parent title suffices for the issue title.

Photograph metadata

Using Adobe Bridge allows for the automatic capture of technical metadata; date, file size, dimensions, resolution, bit depth, color mode, and filename should always be recorded in the file itself. Because of differences in available descriptive metadata, FRASER has no overarching standards for descriptive metadata. Where possible, capture as many as possible of the following: subject, photographer, rights holder, date, size, and type of original.

FRASER website metadata

Metadata for the FRASER website (<https://fraser.stlouisfed.org>) uses an implementation of the Library of Congress MODS metadata standard, with descriptive, structural, and administrative metadata all attached to each record. For further information on the FRASER metadata structure, see the FRASER Metadata Guide.

File storage standards

As mentioned previously, multiple files are produced from each scan. Until the completion of each project, its files are retained on local networked storage drives, which are backed up nightly. Once the project has been closed, its files are moved to a “completed” network storage drive, which is backed up only when changes are made.

At any point, if a folder or group of files is relocated from the project folder, the folder is retained (even if it would otherwise be empty), and a readme.txt created and left in the folder to explain where the files are, and when and why they were moved.

If a completed project needs to be moved to a non-networked storage option, a folder should be created on the usual “completed” drive with a readme.txt explaining the permanent location of the files for that project. External drives are labeled with the project name, the extent of the files contained (e.g., date, reel number, or archival box number range), and the FRASER title ID number. The location of the external drive is also noted in a private note in the title record in FRASER.

Files are generally organized by producer (government office, Federal Reserve entity, publisher, and so on) and then by project.

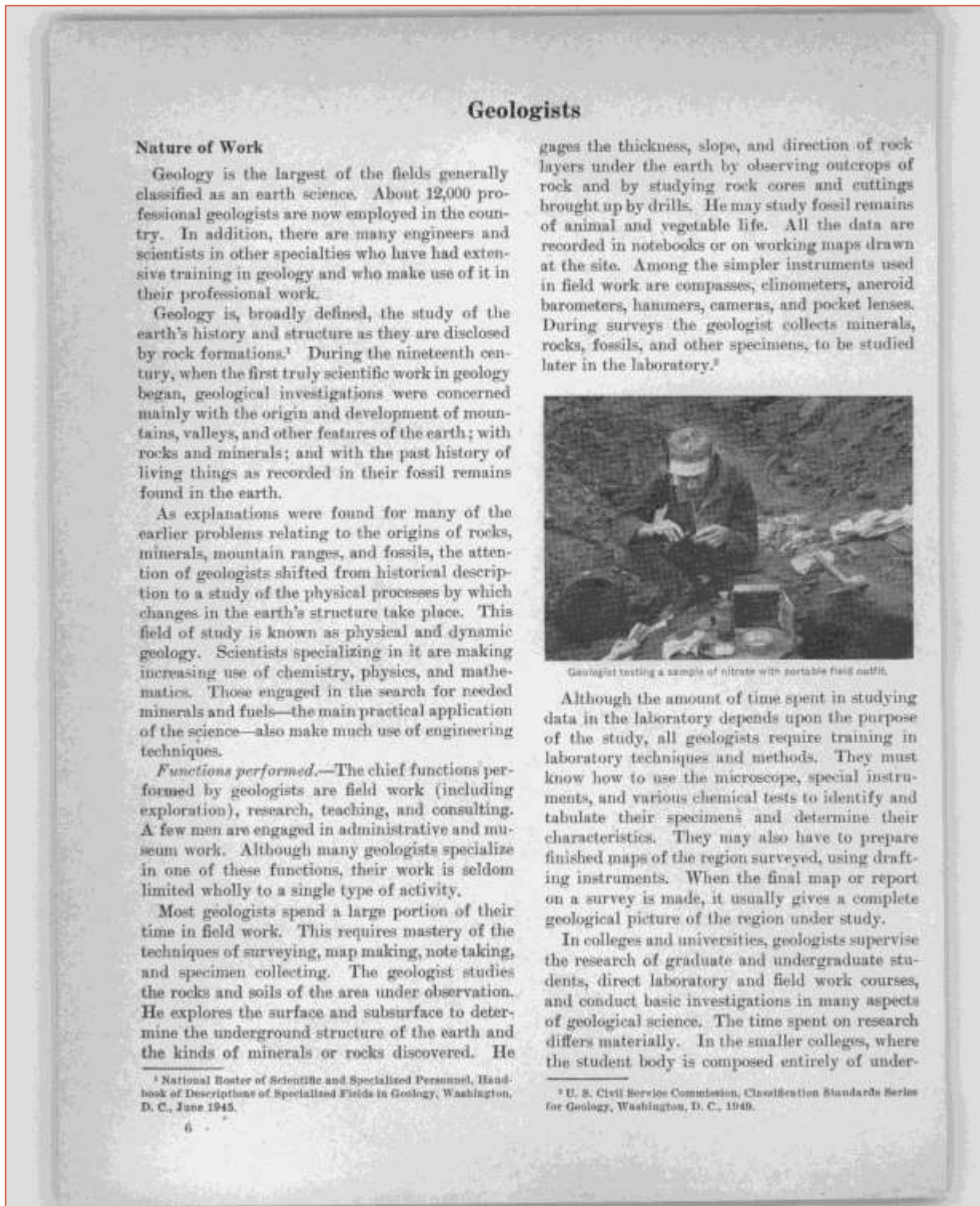
Appendix I: FRASER Workflow

Note: FRASER workflows are highly dependent on project needs and may vary from what is outlined below. In general, however, this is the standard process for an in-house digitization project.

1. Project manager reviews paper documents: This includes assessing the quality of originals and noting any problems (tears, missing pages, missing issues) in the project notes.
2. Project manager establishes scanning procedures: type of scanner to be used, color profile, filenames, and data entry and problem documentation procedures.
3. Technician scans documents.
4. Third party (project manager or second technician) performs a quality check: scanned image files are checked against the original on a page-by-page basis; missing pages, problems, or concerns about the image quality are noted so they can be corrected by the scanning technician.
5. Technician performs image processing.
6. Technician creates PDF copies of the scanned image using optical character recognition software. PDF files are then spot-checked for completeness and inclusion of recognized text (a few—no more than 5 percent—of PDF files should be opened to verify the files show appropriate quality).
7. Technician completes data entry using FRASER batchload spreadsheet template.
8. Third party (project manager or second technician) performs a quality check: metadata is checked against the scans on a document-by-document basis; issues are noted so they can be corrected by the technician.
9. Technician applies metadata to the PDF files and applies the digital FRASER watermark.
10. Project manager transfers PDF files to FRASER web server.
11. Technician loads metadata to the FRASER content management site (FRASER Admin) with non-public/hidden status.
12. Project manager reviews the metadata and sets the posted items to “public” status, making them available on FRASER.
13. Technician boxes, ties with archival tape, or otherwise arranges for safe storage of the originals and transfers originals to be kept with other scanned originals in an archival storage room.
14. Project manager transfers project files (TIFF, PDF, data entry Excel files, and any other files and documentation) to the “completed” project storage drive and sets the project status in project management system to Complete.

Appendix II: Image editing examples

Example 1A: Unedited grayscale image from a predominantly black-and white-publication:



Geologists

Nature of Work

Geology is the largest of the fields generally classified as an earth science. About 12,000 professional geologists are now employed in the country. In addition, there are many engineers and scientists in other specialties who have had extensive training in geology and who make use of it in their professional work.

Geology is, broadly defined, the study of the earth's history and structure as they are disclosed by rock formations.¹ During the nineteenth century, when the first truly scientific work in geology began, geological investigations were concerned mainly with the origin and development of mountains, valleys, and other features of the earth; with rocks and minerals; and with the past history of living things as recorded in their fossil remains found in the earth.

As explanations were found for many of the earlier problems relating to the origins of rocks, minerals, mountain ranges, and fossils, the attention of geologists shifted from historical description to a study of the physical processes by which changes in the earth's structure take place. This field of study is known as physical and dynamic geology. Scientists specializing in it are making increasing use of chemistry, physics, and mathematics. Those engaged in the search for needed minerals and fuels—the main practical application of the science—also make much use of engineering techniques.

Functions performed.—The chief functions performed by geologists are field work (including exploration), research, teaching, and consulting. A few men are engaged in administrative and museum work. Although many geologists specialize in one of these functions, their work is seldom limited wholly to a single type of activity.

Most geologists spend a large portion of their time in field work. This requires mastery of the techniques of surveying, map making, note taking, and specimen collecting. The geologist studies the rocks and soils of the area under observation. He explores the surface and subsurface to determine the underground structure of the earth and the kinds of minerals or rocks discovered. He

gages the thickness, slope, and direction of rock layers under the earth by observing outcrops of rock and by studying rock cores and cuttings brought up by drills. He may study fossil remains of animal and vegetable life. All the data are recorded in notebooks or on working maps drawn at the site. Among the simpler instruments used in field work are compasses, clinometers, aneroid barometers, hammers, cameras, and pocket lenses. During surveys the geologist collects minerals, rocks, fossils, and other specimens, to be studied later in the laboratory.²



Geologist testing a sample of nitrate with portable field outfit.

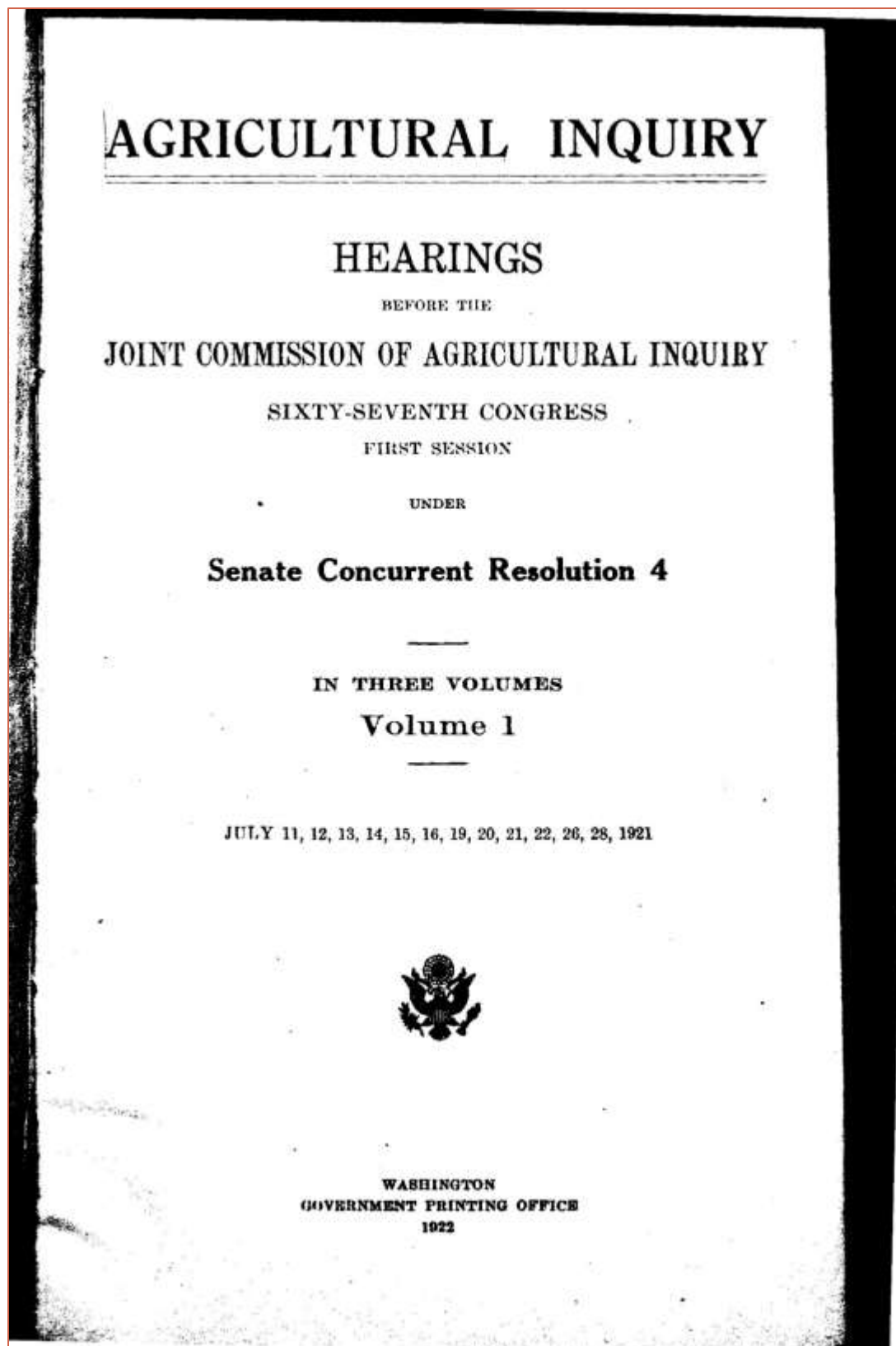
Although the amount of time spent in studying data in the laboratory depends upon the purpose of the study, all geologists require training in laboratory techniques and methods. They must know how to use the microscope, special instruments, and various chemical tests to identify and tabulate their specimens and determine their characteristics. They may also have to prepare finished maps of the region surveyed, using drafting instruments. When the final map or report on a survey is made, it usually gives a complete geological picture of the region under study.

In colleges and universities, geologists supervise the research of graduate and undergraduate students, direct laboratory and field work courses, and conduct basic investigations in many aspects of geological science. The time spent on research differs materially. In the smaller colleges, where the student body is composed entirely of under-

¹ National Roster of Scientific and Specialized Personnel, Handbook of Descriptions of Specialized Fields in Geology, Washington, D. C., June 1945.

² U. S. Civil Service Commission, Classification Standards Series for Geology, Washington, D. C., 1949.

Example 2A: A black-and-white page from an original of average quality:



Example 2B: The same page, edited:

AGRICULTURAL INQUIRY

HEARINGS

BEFORE THE

JOINT COMMISSION OF AGRICULTURAL INQUIRY

SIXTY-SEVENTH CONGRESS

FIRST SESSION

UNDER

Senate Concurrent Resolution 4

—
IN THREE VOLUMES

Volume 1
—

JULY 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 26, 28, 1921



WASHINGTON
GOVERNMENT PRINTING OFFICE
1922

Example 3A: A black-and-white page from a low-quality original:

Library
Confidential
 FEDERAL RESERVE SYSTEM

FEDERAL RESERVE SYSTEM
 FEDERAL RESERVE SYSTEM

MEMBER BANKS - ASSETS AND LIABILITIES BY CLASS OF BANK
 Close of business last Wednesday of July, 1947
 (Partly estimated. In millions of dollars)

	All member banks	Central reserve city member banks		Reserve city member banks	Country member banks
		New York	Chicago		
ASSETS					
Loans and investments	95,384	20,279	4,864	35,017	35,224
Loans and discounts	28,950	6,393	1,631	11,695	9,211
U. S. Gov't. obligations	59,350	12,595	2,876	20,962	22,917
Other securities	7,104	1,291	357	2,360	3,096
Reserves, cash, and bank balances	27,985	5,943	1,510	10,932	9,600
Reserve with F. R. Bank	16,282	4,277	977	6,366	4,662
Cash in vault	1,529	131	32	526	840
Balances with banks in U. S. 1/	5,394	47	158	1,743	3,146
Balances with banks in foreign countries 1/	38	19	2	11	6
Cash items in process of collection	4,742	1,469	341	2,286	646
Other assets	1,385	327	42	551	465
Total assets	124,754	26,549	6,416	46,500	45,289
LIABILITIES AND CAPITAL					
Gross demand deposits	87,081	22,218	5,076	32,044	27,743
Deposits of banks	10,944	3,975	1,104	4,971	894
Other demand deposits 2/	76,137	18,243	3,972	27,073	26,849
Time deposits	28,096	1,158	886	11,252	14,500
TOTAL DEPOSITS	115,177	23,676	5,962	43,296	42,243
Borrowings 3/	238	138	--	55	45
Other liabilities 3/	1,037	504	40	349	144
Total capital accounts	8,302	2,231	414	2,800	2,857
Total liabilities and capital accounts	124,754	26,549	6,416	46,500	45,289
Demand deposits adjusted	70,433	16,623	3,524	24,484	25,802

1/ Slight revision in composition from that shown in previous monthly reports. Previously the 8th and 9th items were "Demand balances with banks in U.S." and "Other bank balances", respectively.

2/ Includes "War loan accounts" previously shown as a separate item.

3/ Borrowings other than from F. R. Bank were previously included in "other liabilities".

Example 3B: The same page, edited. Note that the library stamp and most of the bleed-through from the back page of the original have been kept to show the cause of any interference with the text.

Library
Confidential
 FEDERAL RESERVE SYSTEM

MEMBER BANKS - ASSETS AND LIABILITIES BY CLASS OF BANK
 Close of business last Wednesday of July, 1947
 (Partly estimated. In millions of dollars)

	All member banks	Central reserve city member banks		Reserve city member banks	Country member banks
		New York	Chicago		
ASSETS					
Loans and investments	95,384	20,279	4,864	35,017	35,224
Loans and discounts	28,930	6,393	1,631	11,695	9,211
U. S. Gov't. obligations	59,350	12,595	2,876	20,962	22,917
Other securities	7,104	1,291	357	2,360	3,096
Reserves, cash, and bank balances	27,985	5,943	1,510	10,932	9,600
Reserve with F. R. Bank	16,282	4,277	977	6,366	4,662
Cash in vault	1,529	131	32	526	840
Balances with banks in U. S. ^{1/}	5,394	47	158	1,743	3,146
Balances with banks in foreign countries ^{1/}	38	19	2	11	6
Cash items in process of collection	4,742	1,469	341	2,286	646
Other assets	1,385	327	42	551	465
Total assets	124,754	26,549	6,416	46,500	45,289
LIABILITIES AND CAPITAL					
Gross demand deposits:	87,081	22,218	5,076	32,044	27,743
Deposits of banks	10,944	3,975	1,104	4,971	894
Other demand deposits ^{2/}	76,137	18,243	3,972	27,073	26,849
Time deposits	28,096	1,458	886	11,252	14,500
TOTAL DEPOSITS	115,177	23,676	5,962	43,296	42,243
Borrowings ^{3/}	238	138	--	55	45
Other liabilities ^{3/}	1,037	504	40	349	144
Total capital accounts	8,302	2,231	414	2,800	2,857
Total liabilities and capital accounts	124,754	26,549	6,416	46,500	45,289
Demand deposits adjusted	70,433	16,623	3,524	24,484	25,802

^{1/} Slight revision in composition from that shown in previous monthly reports. Previously the 8th and 9th items were "Demand balances with banks in U.S." and "Other bank balances", respectively.

^{2/} Includes "War loan accounts" previously shown as a separate item.

^{3/} Borrowings other than from F. R. Bank were previously included in "other liabilities".