NPRHA Digitization Standards

Summary

This document recommends scanning and file naming conventions of the NPRHA to serve its image quality and easy identification needs.

Implementing these standards has several benefits:

1. **Accuracy** – sharing standard files online (AtoM or SharePoint) or in digital products like the Vista Dome Package ensures quality and reference accuracy for users such as researchers, manufacturers and publishers.
2. **Consistency** – standards for image quality and naming ensure that all archiving volunteers understand how to do their work and integrate it with others. Every volunteer inventing their own conventions would be chaos.
3. **Conservation** – following uniform standards helps conserve resources (such as computer storage, network bandwidth, etc.) by reducing wasted rework, time to find or transfer, or future compatibility with online library archive systems. We also protect our assets by digitizing them right the first time without unnecessary re-handling.

Scanning Standards

Image File Format

The NPRHA has **standardized on the TIFF standard for initial image capture**. The TIFF image format is recognized as a universally sustainable and supportable archive format by organizations such as the Library of Congress.

When scanning, select the **TIFF format** for saving files. When possible, also select **LZW Compression** as a TIFF format option. LZW compression within a TIFF can sometimes greatly reduce file size without changing any original image quality (also known as Lossless Compression).

Workflow is a sequence of actions that result in a desired outcome. Consider these workflow steps:

- **Capture** (required) – the act of digital image scanning that creates a file
- **Working** (optional) – an optional step that may change the quality or naming of a captured file (clean, crop, brighten, rename, etc)
- **Delivery** (optional) – an optional step of collecting, organizing, or sharing one or more files
- **Archive** (required) – the act of copying digital files to proper library storage

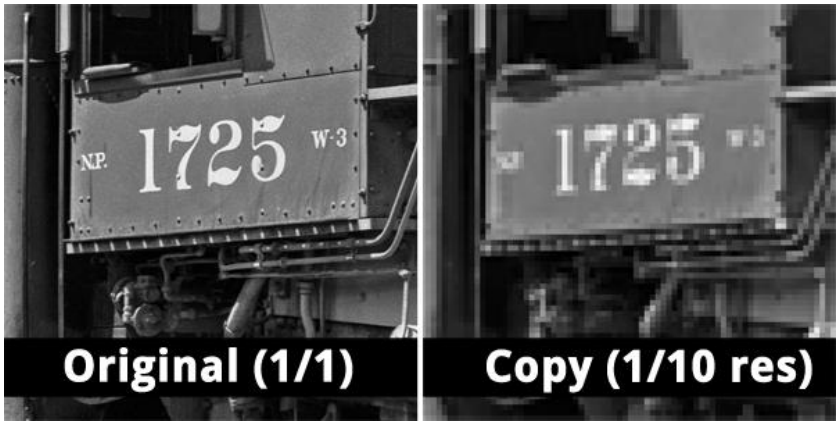
Scanning images and sharing them is a basic workflow for archive's mission. The TIFF file format provides a robust and very broadly adopted standard for the Capture and Archive steps.

This table provides NPRHA guidance, based on standards from the federal government digital archive guidelines. For further research, see http://www.digitizationguidelines.gov/

Format/Use	**Capture**	**Working**	**Electronic Delivery**	**Print Delivery**	**Archive**
TIFF	Yes	Yes	Maybe	Yes	Yes
JPEG	No	No	Yes	No	No
PSD	No	Yes	No	Maybe	No

Image Resolution for Initial Capture

Proper scanning resolution ensures that valuable historical data is not lost when the original artifact is unavailable. For example, the 1/10th resolution scan of the same image below blurs small historical details such as the engine class lettering for this W-3 locomotive:



Without access to the original image, unnecessary ambiguity is introduced and may require rescanning to recover the original detail. For scanning resolution, the smaller the size the media, the higher the resolution. The larger, the lower the resolution.

Film Formats	**Examples**	**Resolution/DPI <u>Standard</u>**
35mm format	35mm slide, 110	4800
Medium format	616, 645, 6X6, 6x7	2400
Large format	4x5, 5x7	1200
Very large format	8x10	600
8x10 print, other printed photos	Various	600, occasionally 1200 for small prints

As a double check on your scanning resolution, color films and slides saved as TIFF should be in the 50MB to 80MB range uncompressed, 25MB to 35MB with LZW compression.

Document File Format

The NPRHA has **standardized on the PDF standard for multi-page document capture**. Examples include AFEs, annual reports, timetables, memos and correspondence.

The NPRHA has **standardized on the TIFF standard for single-page documents**, in particular for documents that are not 8.5x11 or standard paper sizes. Examples include maps, technical drawings, postcards, train sheets, etc.

PDFs should be in version 1.6 (Acrobat 7.x) or higher.

Printed Document Resolution for Initial Capture

The NPRHA scans **virtually all printed documents at 300 DPI**, unless there is an extraordinary reason to jump up to 600 DPI for an extra special document.

Printed/Reflective Formats	**Resolution/DPI**
Documents, 8.5 x 11 pages, drawings	300
Postcards, forms, etc.	300

One reason to scan at 600 DPI is to de-screen a printed image. Printed images are created at the print shop with a screen (very, very, tiny dots of various sizes). Once scanned, the screen-printed images can look poor on the computer. De-screening software can greatly reduce the effect. Example below from a 25.5 x 42 inch NP calendar, using the Photoshop plug-in *Sattva*. It requires 600 dpi to de-screen. It's also true that some initial-scan packages like VueScan can de-screen initial scans when done at 300 dpi.

When In Doubt

When in doubt, it's always better to scan at a higher resolution. Down-sampling an image is easy but scanning at higher detail requires a physical rescan which further micro-damages the document due to handling. <u>Server storage space is cheaper than the labor to re-scan.</u>

File Naming Standards

The NPRHA strives to have consistent file naming standards for anything we digitize, whether it be photos or documents.

The table below provides a summary, with further sections below for each specific row in the table.

Archived Item	Archive File Format	Electronic Sharing Format	Naming Convention for Archive File
Images/Photos/Pictures	TIF	JPG	CCCC-SS-XX-YYY.TIF
Mechanical Dept Drawings	TIF	JPG	FIRST TITLE - SECOND TITLE SERIES# DRAWING# MM-DD-YYYY.TIF
Maps	TIF	JPG	STATION NAME – STATE INITIALS – MM-DD-YYYY.TIF
Forms	PDF	PDF	FORM NUMBER_FORM TITLE_YYYY-MM_FILLED.PDF
Postcards	TIF	JPG	STATION NAME or MAIN SUBJECT – STATE INITIALS – UNIQUE NUMBER.TIF
Annual or Financial Reports	PDF	PDF	YYYY – TITLE – SIZE.PDF
Timetables	PDF	PDF	RR Initial(s) YYYYMMDD XTT Location Descriptor.PDF
AFEs	PDF	PDF	RR-YYYY-AFEnumber.PDF
Condensed Track Profiles	TIF	JPG	Number Station to Station YYYY.TIF
Long Track Profiles	TIF	JPG	Division – Sub Division – Station to Station – MP XX to XX - YYYY.TIF
Marketing Brochures/Booklets	PDF	PDF	MAIN TITLE ON ITEM - YYYY.PDF
Train Sheets	TIF	Zoomify	RR-Initial_DivisionXX-OTHER_DATE_pX-YYYY-MM-DD.TIF
Structure Drawings	TIF	JPG	PlanNumber.InitialYear.revYear-Plan Title.Collection.TIF
Aerial Photos	TIF	JPG	TBD
Large Calendars	TIF	JPG	YYYY.TIF
Posters	TIF	JPG	FIRST TITLE - SECOND TITLE YYYY.TIF
Miscellaneous Anything – Single Page or Image	TIF	JPG	MAIN TITLE or SUBJECT OF ITEM - YYYY.TIF
Miscellaneous Anything – Multi Page	PDF	PDF	MAIN TITLE or SUBJECT OF ITEM - YYYY.TIF

NOTE – the longest file names accepted is 128 characters, but 80% of file names that follow the spec will be much shorter, in the 30-60 character range.

Images/Photos/Pictures

The following describes the file naming standard for slides, negatives, prints or any image that is scanned and given a digital file name.

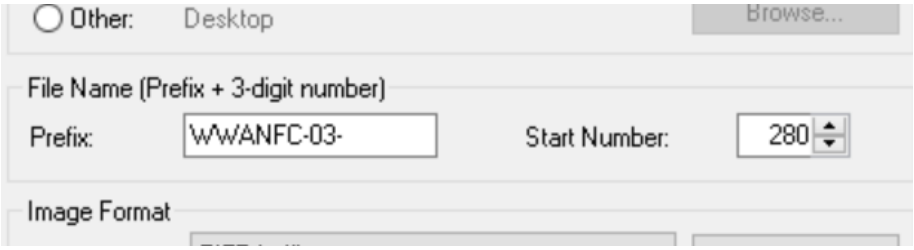
CCCC-SS-XX-YYY.TIF

CCCC – 3-4 Alpha characters that indicates the COLLECTION these images are part of. All slides, negatives, prints and other images are donated to the historical association as part of named collection that is tracked at the archive. The historical association should assign a 3-4 alpha character set to identify each and every collection.

SS – 2 Alpha characters that indicate the SUB COLLECTION of the overall collection. For instance, if Walt Ainsworth donated 10,000 slides with one quarter freight cars, one quarter passenger, one quarter diesel locomotives, and one quarter miscellaneous images, the sub collection alpha characters might be FC, PA, DL, MS. Its up to the archivist who starts scanning the collection to come up with reasonable initials for the Sub Collections.

XX – 2 Numbers that represent the container number the images are stored in. This will typically represent a slide binder or carousel of slides. The binder or carousel should be given a number (even if just a simple piece of blue tape on the cover with a number), and then all images scanned from that container should be given this number. Since the next element YYY can handle UP TO a thousand images, your container XX should be a set count in the 30-999 range.

YYY – The individual item number of the image. This is typically assigned automatically by the scanning software. For instance, **Epson Scan** will automatically add and increment a number at the end of batches of images.



.TIF – The file extension that also represents the format of the file. In virtually all scenarios, original scans should be saved to TIFF file format, with **LZW** lossless compression turned on.

Examples – Good, not perfect	**Examples – Early, not ideal**
WWA-NPD-004.tif	JMF00-00024.0A.tif
WWANFC-04-024.tif	W2015-56-011.tif
	Moreau-004-R.tif
	JC10126.tif

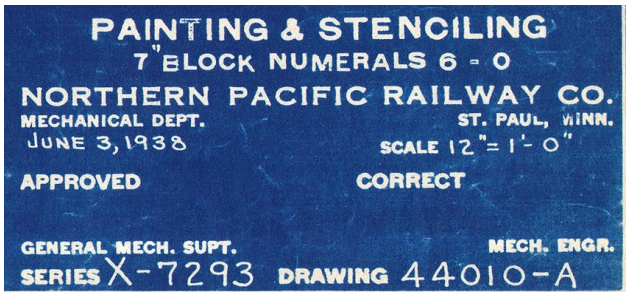
Mechanical Department Drawings

The following describes the file naming standard for blueprints, sketches, or plans from the Mechanical Department that is scanned and given a digital file name.

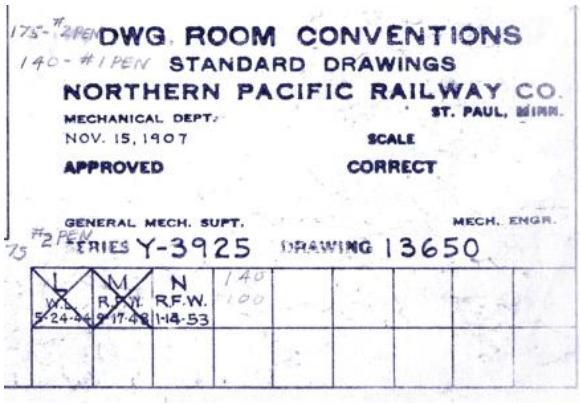
Generally files should not be longer than 75 characters in length, including spaces and extension.

FIRST TITLE - SECOND TITLE SERIES# DRAWING# MM-DD-YYYY.TIF

Note – there is a dash between the first name and second name of the drawing title.



FIRST NAME ON DRAWING – Mechanical Department drawings have a **two-line title**, 99% of the time in the lower right corner. Below is the standard convention. From the example above "Painting & Stenciling" is the first title. It is strongly recommended to change special characters like **& " / \ .** and other characters to words or drop them, so hosting computer systems don't complain or transform the unexpected, unusual characters.

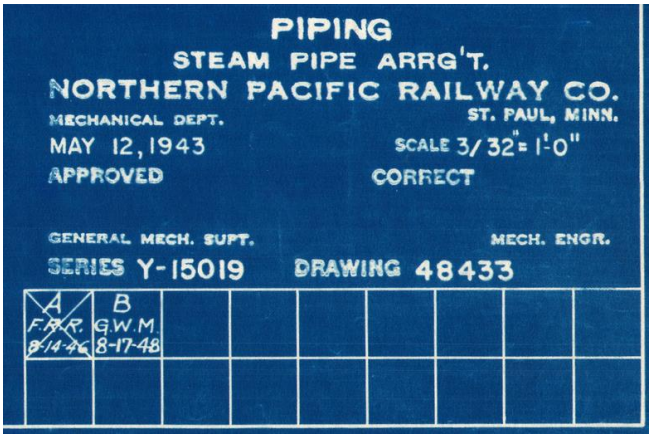


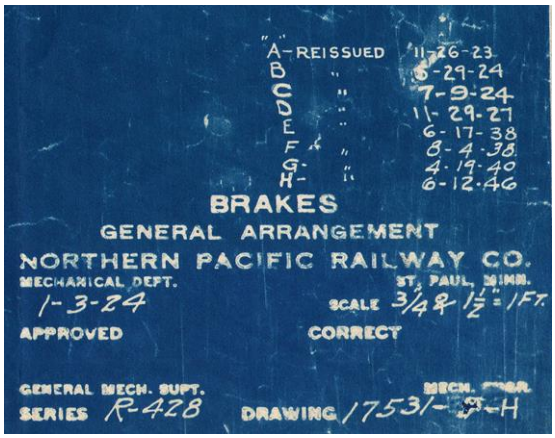
SECOND TITLE ON DRAWING – The second line is below the first, usually in smaller type. In the example above, the second title is "7" BLOCK NUMERALS 6-0". In some rare cases, there is only one title, so skip the second title.

SERIES NUMBER – Each drawing has a SERIES NUMBER, which starts with either a Z, Y, X, W, T, S R or Q. Typically is in the title cluster in the lower right, although its sometimes on the right edge or bottom of the drawing. The first letter indicates the size of the drawing, see DWG Room Conventions for sizes.

DRAWING NUMBER – Each drawing has a DRAWING NUMBER, which is a sequential release number for the drawing, and often includes a revision number, like -A, -G or something like that.

MM-DD-YYYY – The LAST REISSUE DATE that can be found on the drawing. Drawings will often have a listing of reissues, either in a reissue table or as a simple list.





Above left, drawing 48433 is at revision -B with a LAST REISSUE of 08-17-1948

Above Right, drawing 17531 is at revision H, with a LAST REISSUE of 06-12-1946

.TIF – The file extension that also represents the format of the file. In virtually all scenarios, original scans should be saved to TIFF file format, with **LZW** lossless compression turned on.

Examples:

Air Brake Arrangement W-1314 11-20-1942.TIF

Brakes - General Arrangement Q-1141 48066-B 6-11-1948.TIF

End Door Arrangement - Automobile Car R-635 22281-C 1-24-1917.TIF

General Plan - Arrangement Safety Appliances X-3057 19248 9-2-1914.TIF

Window and Details - Cupola end Window W-5781 54416 4-25-1951.TIF

Platforms and Steps - Application Caboose Steps W-4792 49549 8-21-1944.TIF

Painting and Stenciling - 1 Inch Numerals Z-11743 45128-A 10-25-1948

Maps

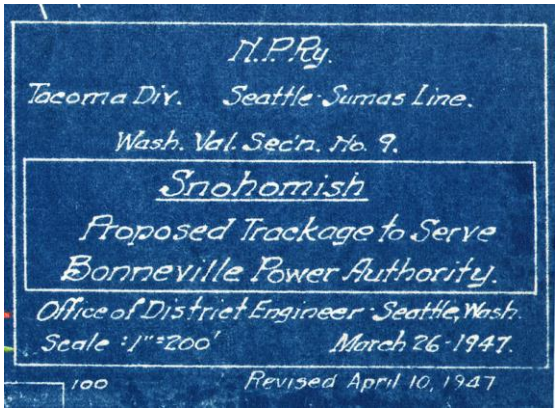
The following describes the file naming standard for maps, including:

- General Building and Bridge Department maps
- Station Maps
- Other 1
- Other 2

This does not include Long or Condensed Track Profiles, which are described below.

STATION NAME – STATE INITIALS – MM-DD-YYYY.TIF

FIRST NAME ON PLAN – Most maps have a **station or location title**, 99% of the time in the lower right corner. Below is the standard convention. From the example below "BIG TIMBER" is the station. It is strongly recommended to change special characters like **& " / \ .** and other characters to words or drop them, so hosting computer systems don't complain or transform the unusual characters.



STATE INITIAL – The second element of the map title is the state initial, in the case above MT. Valid state initials include MN, WI, ND, MT, ID, WA, OR, and MB for Manitoba.

MM-DD-YYYY – The LAST REV or LAST REISSUE DATE that can be found on the drawing.

.TIF – The file extension that also represents the format of the file. In virtually all scenarios, original scans should be saved to TIFF file format, with **LZW** lossless compression turned on.

Examples – Good	Examples – Early, not ideal
Big Timber – MT – 05-21-1964.tif	NP Bozeman low line detail 4.1966.Box26.tif
Snohomish – WA – 04-10-1947.tif	Buckley.1924.tif
	NP Drummond.1960.Box7.tif
	Tacoma.PacificAv to TacomaAv.1927.1.tif
	NP Vincent Section 4.1949.Box21.tif
	NP Old Yegen (Siding #1).1979.Box8.tif

OR, should map titles be more descriptive, like:

Bonneville Power Trackage - Snohomish WA 4-10-1947.TIF

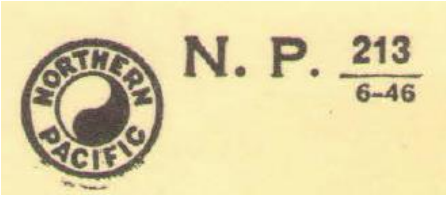
Forms

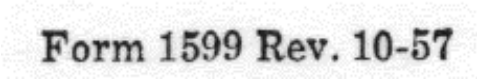
The following describes the file naming standard for Northern Pacific forms. Forms always have a number associated with them, so form file names start with that number. Forms are digitized to PDFs.

This file name standard makes heavy use of underscores, see examples.

FORM NUMBER_FORM TITLE_YYYY-MM_FILLED.PDF

FORM NUMBER – Form numbers are listed in the NPR Stationery Catalog and are typically at a corner of the form.





FORM TITLE – The second element of the form name is the title, again from the Stationary Catalog. Spaces are replaced with Underscores.

YYYY-MM – The year and the month of the form, which is usually very small numbers near or just after the form number at the bottom of the page.

FILLED – FILLED is optional, and indicates that the scanned copy of the form is actually filled out. Filled-out forms are used because a clean original is not available. Or the filled-out version might ad clarity to the usage of the form.

.PDF – The file format and extension for Forms is PDF.

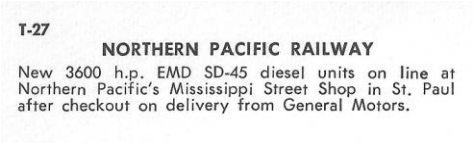
Examples – Good
0214.5_Shipping_Contract_(small)_1897-06_filled.PDF
0220_Report_of_Over,_Short,_and_Damaged_Freight_1924-06.PDF
0277_Commodity_Order.PDF
0278_Order_for_Local_and_Coupon_Tickets_1956-02.PDF
0281_Shipping_Bill_1896-06-15_filled.PDF

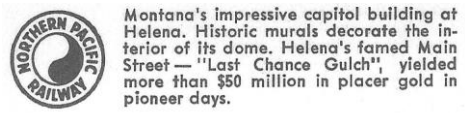
Post Cards

The following describes the file naming standard for postcards showing NP locations or equipment.

STATION NAME or MAIN SUBJECT – **STATE INITIALS** – **UNIQUE NUMBER.TIF**

STATION NAME or MAIN SUBJECT – Most port cards show locations, like stations. But certainly many show specific subjects like a circus train or new locomotives in a yard. Use your discretion in naming the postcard, but Station Name is preferred.





STATE INITIAL – The second element of the map title is the state initial, in the case above MT. Valid state initials include MN, WI, ND, MT, ID, WA, OR, and MB for Manitoba.

UNIQUE NUMBER – Because there are many post cards for given stations or locations, look for any date or code number for the postcard that can make the file name unique.

.TIF – The file extension that also represents the format of the file. In virtually all scenarios, original scans should be saved to TIFF file format, with **LZW** lossless compression turned on.

Examples – Good, from above
New NP SD-45s – J5159.tif
Capitol Building at Helena – MT – NP6019.tif
Tacoma Depot – WA – 05-1924.tif
Morehead – MN – 0099.tif

Annual Reports

The following describes the file naming standard for NP annual reports.

YYYY_TITLE.PDF

YYYY – Year the annual report was published.

TITLE – The title of the publication.

.PDF – The file format and extension for NP corporate publications like annual reports is PDF.

Examples – Good
1943_NPR_ANNUAL_REPORT.PDF
1952_NPR_ANNUAL_REPORT.PDF
1969_NPR_ANNUAL_REPORT.PDF

Time Tables

The following describes the file naming standard for employee and public timetables.

RR Initial(s) YYYYMMDD XTT Location Descriptor.PDF

RR Initial(s) – the reporting marks of the Railroads that published the timetable. For instance "NP" or "GN-NP-UP" or "MILW"

YYYYMMDD – the year (YYYY), month (MM) and date (DD) the timetable was published. If month or date is not stipulated, just omit.

XTT – either ETT for Employee Time Table, or PTT for Public Time Table. Are their others?

Location Descriptor – a short description of the location covered by the timetable. Less than 32 characters, ideally short as possible.

Examples – Good
NP 19650907 PTT SYSTEM.pdf
GN-NP-UP 19651031 PTT LOCAL Portland-Seattle Pool.pdf
NP 19661030 PTT System.pdf
NP 19680818 PTT System.pdf
GN-NP-UP 19681027 PTT LOCAL Portland-Seattle Pool.pdf

Authority for Expenditure (AFEs)

The following describes the file naming standard for Authority for Expenditure.

RR-YYYY-AFEnumber.PDF

RR Initial – the reporting marks of the railroad that published the AFE.

YYYY – The year the AFE was issued, four digits.

AFE number on the sheet – typically a **FOUR-TO-SEVEN** place alphanumeric, with the actual AFE number padded with zeros so it equals five digits (i.e. not 49, but 00049).

Examples
NP-1920-D1234.pdf
GN-1944-00003.pdf
CP-1950-00800.pdf
NP-1897-RM0016.pdf
NP-1920-5081a.pdf

Condensed Track Profiles

The following describes the file naming standards for Condensed Track Profiles.

Number Station to Station YYYY.TIF

Number – Condensed profiles are numbered, East to West.

Station to Station – Condensed profiles are also titled, station to station.

YYYY – Condensed profiles are published for a year.

Examples – Condensed Track Profiles
7 Manitoba Jct to E Grand Forks 1928.tif
4 St Paul to Staples 1969.tif
27 Tacoma to Tenino 1969.tif

Long Track Profiles

The following describes the file naming standards for regular or "Long" track profiles. This naming convention is long, but since there are so few long profiles and most have been scanned to this convention – this is the convention.

Division – Sub Division – Station to Station – MP XX to XX - YYYY.TIF

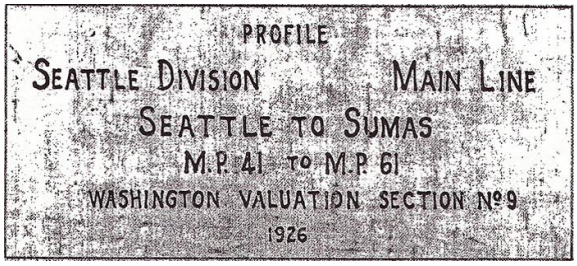
Division – Overall division for which the track existed. This comes directly from the profile title.

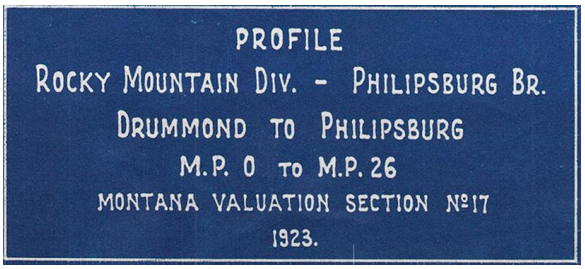
Sub Division – Sub Divisions are pulled from the 1969 sub division listing in the Station Agents book.

Station to Station – The profile will clearly list the station to station covered by the profile in the title box.

MP XX to XX – The profile will also list the milepost to milepost (MP) for each profile. This can get complicated when the profile crosses MP number transitions, see existing examples for guidance.

YYYY – Long profiles are published for a year.





Examples – Long Track Profiles
Rocky Mountain Div - Third Subdivision - Helena to Missoula - MP 22 to 42 – 1924.tif
Yellowstone Div - Fifth Subdivision - Laurel to Livingston - MP 100 to 120 - 19XX.tif
Seattle Div - Fifth Subdivision - Seattle to Sumas - MP 41 to 61 – 1926.tif
Tacoma Div - Fifteenth Division - Wickersham to Bellingham - MP 0 to End – 1904.tif
Camas Prairie - Fourth Division - Oro Fino to Headquarters - MP 0 to 20 – 1928.tif

Marketing Brochures and Booklets

Over the decades, the NP published thousands of brochures (no binding, just folded) and booklets (with binding).

MAIN TITLE ON ITEM - YYYY.PDF

MAIN TITLE ON ITEM – Overall division for which the track existed. This comes directly from the profile title.

YYYY – The year the brochure or booklet published for a year. Where do you find the year? Sometimes its right in the title. But frequently, its in very small type on the info page of the item, where the form number is also listed. If after checking this location, you may be left listing the rough decade, like – 193X, or consulting with others or similar artifacts for the best date to put in the file name.

A. M. CLELAND, General Passenger Agent St. Paul, Minn. 1917	Published 1917
City_____ Zone____ State____ F. 6935 Printed in U.S.A. 12-64 14	Published December, 1964
F6881—PRINTED IN U. S. A.—4-50 6	Published April, 1950
FORM 6859—51 6	Published 1951

Examples – Marketing Brochures and Booklets
Jim Bridgers Yarns of the Yellowstone - 1931.pdf
The Yellowstone National Park via NPR - 1892.pdf
Magic Yellowstone - 1931.pdf
Lewis and Clark Centennial Exposition - 1905.pdf
North Pacific Coast Events - 1909.pdf
San Francisco Worlds Fair - 1939.pdf
Scenic Route Across America - 1951.pdf

Train Sheets – Need finish-up from Kent

The following describes the file naming standards for Train Sheets.

RR-Initial_DivisionXX-OTHER_DATE_pX-YYYY-MM-DD.TIF

RR Initial – the reporting marks of the railroad that published the train sheet.

DivisionXX- - Kent?

OTHER_DATE_ - Kent?

pX- - Kent?

YYYY-MM-DD – The specific date stamped on the Train Sheet.

Examples
NP_Tacoma07-08-10-19_p2-1952-08-13.tif
NP_Tacoma07-08-10-19_p2-1951-06-20.tif

Structure Plans

The following describes the file naming for structure plans, typically as large, stand-alone sheets of paper (not simply pages in a Standard Plan book).

PlanNumber.InitialYear.revYear-Plan Title.Collection.TIF

Plan Number – The plan number on the sheet, usually in the upper right corner. Examples include S-26-16, 5-4-1, etc.

InitialYear – Typically plans have the initial year the plan was published, just below the title.

revYear – The plans will also include a listing of revisions, including data. So this would be the last date on the sheet the plan was revised.

Plan Title – The first title from the plan, and if a secondary title is included, some or all of the secondary title. They key is to keep the file name reasonably short, so the title should <u>not be longer than 50 characters.</u>

Collection – The name of the collection that the plan comes from, shortened to 5 or 6 characters.

.TIF – The file extension that also represents the format of the file. In virtually all scenarios, original scans should be saved to TIFF file format, with **LZW** lossless compression turned on.

Examples
CHUCK SCHEME
S-29-2.1889.16-rm Boarding House.Schr.tif
M-41-15.1902.rev1904.Coal Shed.Iszley.tif
S-26-16.1900.3rd Class Combo.Iszley.tif
S-26-17.1900.rev1904.3rd Class Combo BillMtl.Schr.tif
DEAN/OTHER SCHEMES
Standard Plans of 48,000 Gal Tank - Low Frame – Modern.tif
Standard Plan for 100000 Gallon Water Tanks - Sheet 1.tif
NP Standard Plan Blow off Basin 1912-04-10 Drawer 4-US.tif

NOTE – Would like to propose this as the new standard going forward:

PlanNumber InitialYear Plan Title.TIF

Or, to align with Mechanical Drawings (consolidate to one date, and drop collection).

FIRST TITLE - **SECOND TITLE PlanNumber MM-DD-YYYY.TIF**

Aerial Photos

This section TBD, need to document scheme adopted by Chuck Soule.

Large Calendars

The following describes the file naming standards for large **40 x 25.5** and medium **25.75 x 20.5** NP wall calendars.

YYYY.TIF

YYYY – The year the calendar presents.

We could consider prepending the "Title" of the calendar, but they often are very long and its not always clear what is title and what is commentary. And often they did not change for multi-year stretches. "Ship and Travel via NORTHERN PACIFIC RAILWAY" for instance.

Year is short and sweat.

Posters

The following describes the file naming standards for large **40 x 25.5** and medium **25.75 x 20.5** NP wall calendars.

FIRST TITLE - **SECOND TITLE YYYY.TIF**

FIRST TITLE – The first title of the poster.

SECOND TITLE – If there is a secondary tile to the poster, or a tag line that is short, consider adding this. Its OPTIONAL. Also, if you know of multiple posters of the same exact subject, consider using the SECOND TITLE to differentiate the two (Portrait, Landscape for the Grand Canyon of Yellowstone).

YYYY – The year poster was published, if available in the fine print.

Examples – Posters
Beef Cattle.tif
Paradise Valley.tif
FT 6000 - ONE OF A FLEET.tif
U33Cs on River.tif
Streamlined NCL in Rockies.tif
North Coast Limited by Ragan.tif
Grand Canyon of Yellowstone – Landscape.tif