

STAFF HOWTO Website Image Licensing Audit (MVP)

Purpose

Quickly scan a site you manage for images that might need licensing review.

Access

Go to the Streamlit app URL.

If prompted, enter the passcode (ask your admin if you dont have it).

Run a Scan

1) Paste the Start URL (usually the homepage).

2) Scope & Depth:

Leave Include subdomains OFF at first (enable if the site uses multiple subs).

Set Max crawl depth to 23 for typical sites.

3) Limits (safety valves):

Max pages / Max images: caps how far the run can go.

Perpage image cap: stop collecting more than N images per page.

Perimage size cap (MB) & Total download cap (MB): limits data fetched for EXIF/thumbnails.

4) Fetch Policy:

Use moderate Concurrency (e.g., 5) and a Base delay (e.g., 300 ms) to be polite.

5) Features:

Attempt EXIF/IPTC: ON (reads EXIF Artist when available).

Show thumbnails: optional; ON uses more data (respects caps). When ON, EXIF is autoenabled.

Capture CSS background images: OFF by default; turn ON only if you need to catch hero/section backgrounds (may include fonts/gradients).

6) Click Run Audit.

Review Results

Use Show likely stock/library only to focus on stockhosted images.

Use Show rows with risk flags to prioritize review.

Use the Search box to filter by URL, page, alt text, or domain.

Click Google Images / TinEye links in each row to perform a quick provenance check.

Columns include: Page, Image URL, Source Type (IMG/CSS), Alt Text, Domain, Guessed Source,

ContentType, Estimated Bytes, EXIF Artist, Width/Height, Google Images, TinEye, Notes, Risk Flags.

Risk Flags (what they mean)

Stock source ensure license: hosted on a known stock domain (e.g., Shutterstock, Getty).

No alt text (check provenance): missing alt text on ; not proof of misuse but worth verifying.

Very large file (N MB): may be a licensed asset or hero image doublecheck rights.

Very large dimensions (N px on either side): likely a hero/banner doublecheck rights.

Suspicious stock ID in filename: filename includes patterns like shutterstock_12345.

Offdomain asset (hotlink): file is hosted on a different registrable domain than the start URL.

Brand term match: URL/alt contains one of your brand/trademark terms; may indicate a