

PDF-to-HTML Checklist

Factors to consider when converting existing long-form reports and toolkits from a single large PDF to interactive HTML web content.

For more tools and ideas visit reportkitchen.com

Make it friendly for human readers

- Use a comfortable text width and readable type size.
- Break long pages into meaningful pages and sections. You definitely don't want one web page for every page in the PDF, nor do you want one single web page for a 75 page document.
- Add an easy-to-use navigation menu that always orients the reader to where they are in the document.
- Make headings, lists, tables, charts, and callouts easy to scan.
- Ensure the layout works on phones without zooming or horizontal scrolling.

Make it accessible

Please note: This doesn't cover everything you need to do to make your content accessible, just issues likely to arise in a conversion like this.

- Use semantic HTML elements for headings, lists, tables, quotations, figures, and navigation.
- Provide meaningful alt text for informative images.
- Give charts and diagrams text alternatives that communicate their key information.
- Make tables understandable with proper row and column headers.

Improve search and AI understanding

- Give each substantial page a clear, specific title and primary heading.
- Use descriptive section headings that reflect the questions or subjects being addressed.
- Keep important findings in live text, not only inside images or charts.
- Make authorship, publication date, organization, and source information explicit.
- Add a concise summary that states what the report covers and why it matters.
- Use stable URLs for major sections when they deserve separate pages.
- Link related sections together so their relationships are explicit.
- Add appropriate metadata and structured data without treating either as a substitute for clear content.

Handle charts, tables, and figures deliberately

- Do not publish important charts as unexplained images.
- Include the chart title, key takeaway, labels, source, and underlying data where practical.
- Rebuild simple tables as real HTML tables.
- Keep captions and source notes attached to the correct figure.
- Preserve important visual distinctions without forcing readers to interpret the original page design.

Edit for the medium

- Remove page numbers, running headers, print-only instructions, and references such as “see opposite page.”
- Replace page-based navigation with section-based navigation.
- Rewrite links and cross-references so they work online.

- Decide whether repeated introductory material is still useful on every page.
- Add context where content previously depended on physical proximity within a spread.

Keep the PDF, but define its role

- Offer the original PDF for download when it remains useful as the official or printable edition.
- Make the HTML version the primary readable and discoverable version.
- Clearly connect the PDF and HTML editions.
- Avoid publishing several near-identical versions without a clear canonical source.

Review before publishing

- Compare the HTML against the original section by section.
- Check headings, reading order, citations, callouts, tables, charts, and footnotes.
- Test on desktop and mobile.
- Run accessibility checks, then perform manual review.
- Confirm that a reader can find, understand, quote, and link directly to the report's most important ideas.
- Have someone who knows the publication approve the final structure and meaning.