

PDF Tagger Baseline Test Document

*Meridian Instruments Ltd. — 2025 Annual Product & Compliance Report
(Fictional)*

This document is a **baseline test fixture** for an automated PDF accessibility tagger. Every section exercises a specific structure type, and the expected tag outcome is noted in a shaded box at the start of each section. Two builds of this file exist: an *untagged* input and a *tagged* reference produced from the same source.

Scope: headings H1-H3, paragraphs, inline styles, links, footnotes, bulleted and nested lists, numbered lists, figures with/without alt text, decorative images, a simple table, a complex table with merged cells, a full AcroForm field set, multi-column text, rotated text, and an image-only “scanned” page.

Contents

1. [Section 1 — Headings, Text and Links](#)
2. [Section 2 — Lists](#)
3. [Section 3 — Images and Figures](#)
4. [Section 4 — Tables](#)
5. [Section 5 — Interactive Form Fields](#)
6. [Section 6 — Layout Edge Cases](#)
7. [Section 7 — Scanned \(Image-Only\) Page](#)

Section 1 — Headings, Text and Links

Expected tags: H1/H2/H3 hierarchy, P, Link (internal and external), BlockQuote, Span-level language change, Note/footnote.

Meridian Instruments designs precision measurement equipment for industrial laboratories. In fiscal year 2025 the company shipped **14,200 units** across four product lines, an increase of 31% over the prior year. Inline formatting in this paragraph includes **bold**, *italic*, underlined, `monospace code`, a superscript reference¹, and a chemical formula with subscript: H₂O.

1.1 Hyperlinks

External link: the full compliance standard is published at www.example.com/standards/pdf-ua. Internal link: completed order forms are described in [Section 5 of this document](#).

1.2 Quotations and language changes

“Accessibility is not a feature to be bolted on at the end; it is a property of a document made correctly from the start.” — Fictional Standards Committee, 2024

A span-level language change follows, which should receive a Lang attribute: our Lyon office reports « *la qualité avant tout, la précision toujours* » as its operating motto.¹

1.2.1 A third-level heading

This short paragraph exists only so the document contains an H3 beneath an H2, letting the tagger demonstrate correct heading nesting rather than flattening everything to a single level.

¹This is footnote one. A tagger should produce a Note structure element linked from the superscript citation in the body text.

Section 2 — Lists

Expected tags: L, LI, Lbl, LBody; nested L inside LI for sublists; correct list continuation across items.

2.1 Bulleted list with nesting (three levels)

- Laboratory instruments
 - Spectrometers
 - Model SP-100 (benchtop)
 - Model SP-200 (portable)
 - Calibration rigs
- Field equipment
- Consumables and reagents

2.2 Numbered list with sub-items

1. Unpack the instrument and inspect for shipping damage.
2. Connect the power supply.
 - a) Verify the voltage selector matches your region.
 - b) Use only the supplied grounded cable.
3. Run the self-calibration routine before first use.

Section 3 — Images and Figures

Expected tags: Figure with Alt text (3.1), Figure with MISSING alt text that the tagger must flag or fix (3.2), decorative image tagged as Artifact (3.3), Caption paragraphs associated with their figures.

3.1 Figure with alternative text

Quarterly Revenue 2025 (\$M)

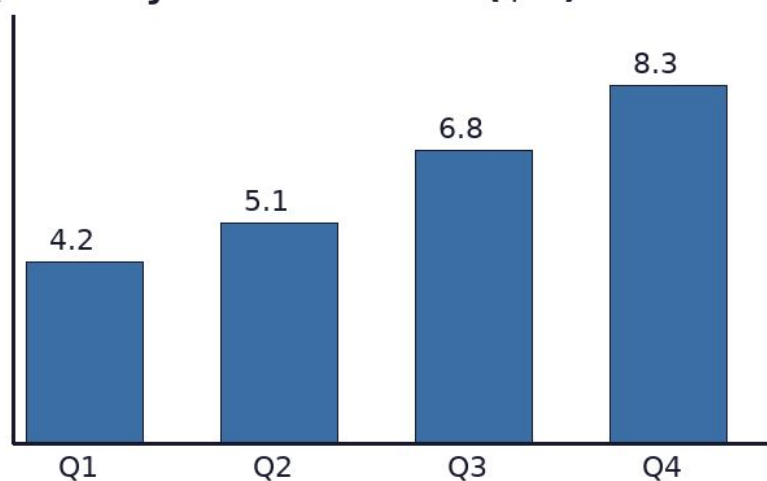


Figure 1: Quarterly revenue, 2025. This image HAS alt text in the source.

3.2 Figure without alternative text (deliberate failure)



Figure 2: Landscape photograph. This image deliberately has NO alt text — the tagger should detect and remediate this.

3.3 Decorative image

The ornamental divider below is purely decorative and is marked decorative in the source; it should be tagged as an Artifact, not a Figure.



Section 4 — Tables

Expected tags: Table, TR, TH (with Scope), TD; THead via table header rows; merged cells with ColSpan/RowSpan; a total row; rotated header text in 4.3.

4.1 Simple table with one header row

Region	Regional Manager	Field Offices
Mountain West	D. Okafor	Denver, Boise, Salt Lake City
Pacific Northwest	R. Lindqvist	Seattle, Portland
Southwest	M. Delgado	Phoenix, Albuquerque, El Paso

Table 1: Simple 3×4 table — single header row, no merged cells.

4.2 Complex table: multi-row header, merged cells, total row

Product Line	FY 2024		FY 2025	
	Units	Revenue (\$M)	Units	Revenue (\$M)
Spectrometers	3,100	6.8	4,900	10.9
Calibration rigs	1,450	4.1	2,200	6.3
Legacy meters	Discontinued in March 2025 — merged cell spanning four columns (ColSpan=4)			
Field kits	5,200	3.9	7,100	7.2
TOTAL	9,750	14.8	14,200	24.4

Table 2: Complex table — header cell with RowSpan=2, two header cells with ColSpan=2, a body cell with ColSpan=4, and a total row.

4.3 Rotated text inside table cells

ROTATED 90°	The label to the left is rotated 90 degrees. Reading order and extraction should still be correct.
--------------------	--

Section 5 — Interactive Form Fields

Expected: Form structure elements wrapping each widget annotation; tooltips (/TU) present on all fields EXCEPT the email field (deliberate failure); a radio group sharing one field name; a pre-checked checkbox; dropdown (combo) and multi-select list boxes; a push button; a digital signature field (added post-export).

Full name (text field, has tooltip)

Email address (text field, NO tooltip — deliberate failure)

Inspection date (date field)

Comments (multi-line text area)

Checkboxes

☐ Subscribe to the quarterly newsletter

☐ I have read the safety guidelines (pre-checked)

Preferred contact method (radio group — one field, three widgets)

☐ Email

☐ Telephone

☐ Postal mail

Sales region (dropdown / combo box)

Certifications held (multi-select list box)

Submit (push button)

Submit Report

Digital signature (signature field placed in the box below after export)

Sign in the area below this line:

Section 6 — Layout Edge Cases

Expected: correct reading order across two newspaper columns; footnote at page bottom tagged as Note; hyphenation and column breaks must not corrupt word order.

6.1 Two-column text flow

This section is set in two newspaper-style columns. A correct tagger must produce a logical reading order that runs down the first column completely before continuing at the top of the second column, rather than interleaving lines horizontally across the page. Automated taggers that rely purely on geometric y-position sorting typically fail this test.

The second paragraph continues the flow. Meridian's field service division completed 1,900 on-site calibrations during the reporting year², maintaining a first-visit

resolution rate of 94 percent. Response-time commitments were met in every region except the Great Plains, where winter storms caused a twelve-day service backlog in January.

A final short paragraph closes the columned section so that both columns contain a comparable amount of text and the column break lands mid-sentence somewhere above, which is exactly the hard case a reading-order algorithm needs to survive.

²Footnote two, cited from inside a two-column section. The note body renders at the bottom of the page, outside the column flow.

Section 7 — Scanned (Image-Only) Page

The next page is a single full-page image with NO text layer. Expected: the tagger flags it for OCR, or tags the page image as a Figure with alt text. The image below has alt text in the tagged reference.

MEMORANDUM

TO: All Department Heads
FROM: Records Management Office
DATE: March 14, 1987
RE: Archival of Legacy Files

This page simulates a scanned document. It contains no machine-readable text layer whatsoever. An accessibility tagger should detect that this page is image-only and either flag it for OCR or tag the whole page as a Figure with an appropriate alternative description.

All physical records dated prior to January 1, 1980 shall be transferred to the basement archive no later than the last business day of this quarter. Department heads are responsible for completing transfer form RM-104 for each box submitted.

Failure to comply may result in loss of records during the upcoming office relocation.

A handwritten signature in blue ink, consisting of a series of connected diagonal lines forming a stylized, jagged shape.

J. Whitfield, Records Manager