

Technical Specification

This document describes the Omni Suite architecture and the role of each component in the localization pipeline.

Architecture Overview

The pipeline consists of three stages: extraction, translation, and backfill. Each stage is implemented as a standalone module with its own CLI and MCP server.

Stage	Module	Output
1	OPP	MD + XLIFF + skeleton
2	OL	Translated MD/XLIFF
3	ORF	DOCX/PPTX/PDF/...

Component Details

OPP — Omni Pre Processor

Extracts content from 12 input formats including DOCX, PPTX, PDF, XLSX, HTML, EML, and others.

OL — Omni Localizer

Translates MD or XLIFF intermediates using configurable LLM backends with shield/repair/unshield pipeline.

ORF — Omni Re Formatter

Backfills translated content into 16 target formats using pandoc, weasyprint, and pure-Python renderers.

- Full format matrix verified
- MCP and CLI parity
- Fidelity-tested on real corpus

Performance Metrics

Metric	Value
Avg cell runtime (CLI)	15s
Avg cell runtime (MCP)	18s
Text preservation	99%
Table preservation	100%

Conclusion

The system is production-ready for the verified path set.