

Scaling  Internationalization



Project goals & discovery questions

Background

- Our work began as the new [REDACTED] landing pages were completed, with the intention to explore expansion of that work

Goals

- We want to better serve LEP (Low English Proficiency) consumers.
- We want to develop a content management process that is sustainable by staff and ensures accurate content for LEP consumers.

Discovery questions

- What options are available to scale up beyond current capacity?
- What would it take to make those options happen?

Recommendation

We recommend conducting a proxy translation pilot.

What is proxy translation?

A third-party vendor would crawl [REDACTED] to produce a mirror that is then translated, and the translations would be served to [REDACTED] visitors through their browser. Proxy translation provides options for 100% human translation, machine translation, or a combination of the two.

Why are we recommending it?

Based on the options we explored, proxy translation provides the greatest potential for [REDACTED] to offer a large amount of quality translated content to LEP consumers. The method would also enable us to avoid the costs and concerns of a [REDACTED] integration for LEP content.

Current state

Our current approach to translation – from a content management perspective and a web platform development perspective – **will not scale**.

- **Ownership:** No clear ownership of larger strategy beyond ad hoc translation work.
- **Maintenance:** Outside of certain products, non-English content is not consistently synchronous with English content. (e.g. ██████)
- **Publishing:** Building non-English pages in ██████'s CMS is inordinately cumbersome and prone to error. Outside of special projects, a single ██████ staffer (██████) is solely responsible for most/all daily web implementation and publishing.

Our current offerings are only valuable to LEP audiences because of **individual, non-scalable projects** like ██████'s work on the landing pages, or the ██████.

Our CMS's current, custom-built LEP functionality needs to be untangled **before** any other approach can be implemented in its place.

Discovery: Translation options

Human translation

■ Research insights

- [REDACTED] has robust resources, including an in-house linguistic staff of 13, a \$[REDACTED] external translation contract, and [REDACTED] dedicated website publishers.
 - **Despite this**, they translate only 19% of their web content into Spanish and under 5% into any other language.
- Proxy translation (detailed on slide 8) provides an option for human translation through a third-party vendor.

■ Key takeaways

- [REDACTED] in-house staff approach is beneficial in that linguists with subject matter expertise can be employed. But it is also expensive, highly manual, and requires significant coordination/tracking.
- Human translation through proxy translation appears promising.

Discovery: Translation options, cont.

Machine translation

- **Research insights:** Some [REDACTED] we talked to use machine translation, such as the Google Translate API, without human review.
- **Key takeaway:** This approach provides a quick translation workflow and significant cost savings for initial translation. But we have a [REDACTED] requirement for human review, and this approach would require significant quality control and review. Human review would also impact the publishing workflow (detailed on slide 7).



Human review is
required for all content
translated via machine,
per [REDACTED]

Discovery: Publishing options

Working in the CMS (i.e. expanding existing [REDACTED] implementation)

- **Research insights:** To scale up, we would need to use [REDACTED] core packages and rework non-English content into a 1:1 structure, with parallel URL paths. This approach requires significant additional staff: content editors (ideally with i18n experience) to publish translated content, and [REDACTED] specialists to lead strategy and support implementation.
- **Key takeaway:** In addition to technical and resourcing needs, this approach would be highly manual and would not remedy current quality control issues.

Machine translation with a non-CMS implementation

- **Research insights:** The machine translations of each [REDACTED] we spoke to are hosted outside of their site's CMS. This enables near-instant, automatic publishing and no substantial manual workflow needs involving their CMS.
- **Key takeaway:** Because of the [REDACTED] mandate, human review by someone who speaks the language would need to be built into the workflow and, depending on the amount of content, could add significant time, coordination, and cost.

Discovery: Publishing options, cont.

Proxy translation services

- **Research insights:** A third-party vendor crawls and stores current site content to produce a mirror, which is then translated. The translated site(s) appear seamlessly connected to [REDACTED] to the public, but are hosted on a separate server and do not interact with [REDACTED] or internal publishing workflows.
 - Can rely upon existing glossaries and key terms; [REDACTED] staff have access and control to change translations if desired
 - Choice of translation method: human, machine, or a hybrid. One major vendor, [REDACTED], is rolling out an “adaptive” translation process, where human translation is supplemented with a machine-learning model, trained on existing human translations.
- **Key takeaways:**
 - Potential major cost savings; requires minimal technical implementation; has no/minimal [REDACTED] functionality.
 - Possible technical hurdles for provisioning and setting up subdomains.
 - Could benefit from more discovery into technical details as well as other [REDACTED]' experiences with [REDACTED] or other vendors.

Discovery: Content management needs

Before beginning a translation/publishing approach:

- The question of what should be considered "consumer content" on [REDACTED] should be definitively answered with a site audit, in consultation with content owners and divisional stakeholders.
- A prioritization model should also be created to assess which content is most important to provide to LEP audiences, to assist in any rolling efforts to expand translation services.

Discovery: Options for scope of translation

Potential scopes of work:

1. Microsite (~20 pages/language). Expansion of existing LEP landing page offerings, offered as a navigationally complete and separate section of site content from English. Curated and selected based on existing English and non-English [REDACTED] content.
 - Significant potential trade-offs in user experience and potential content parity with this approach vs. translating existing English content
 - **We recommend starting with a microsite as an "MVP" approach to, at minimum, better test options for expanding internationalization offerings.**
2. Limited consumer content (~500 pages/language). Expansion of translation to include some [REDACTED] content based on prioritization and assessment of English content.
3. Most consumer content (~1500 pages/language). Expansion of all translation to match scope of Spanish site translation, i.e. all [REDACTED].
4. Expansive consumer and site content (~5000 pages/language). Same as above, plus inclusion of more [REDACTED].
5. Complete translation (~15,000 pages/language, plus thousands? of print materials). Complete site mirror in all available languages.

Initial recommendations

Current state

- **Unsustainable.** To maintain or expand translation and internationalization efforts on ██████, we must develop new process and support it with proper resourcing.

Translation options

- **100% in-house (or contract-based) human translation** — **Do not recommend.** Too costly and burdensome on existing or new staff and has heavy publishing implications.
- **Machine translation with human review** — **Do not recommend** except as part of a limited pilot study to understand quality and feasibility.

Publishing options

- **Working in the CMS** — **Do not recommend.** ██████ is not equipped handle to the additional strain.
- **Proxy translation (with human or adaptive translation)** — **Recommend pilot**
 - Already have established contact with ██████, a major vendor
 - Tools promise to provide human-level translation support for full-site translation without the costs or concerns of ██████ integration.

Recommendations

	Microsite (Recommended as initial scope)	Limited consumer content	Most consumer content	Expansive consumer and site content	Complete 1:1 translation
Manual translation within existing CMS	Feasible, but expensive	Not recommended	Not recommended	Not recommended	Not recommended
Machine translation with human review	Not recommended	Not recommended	Not recommended	Not recommended	Not recommended
Proxy service using 100% human translation	Exploration recommended	Exploration recommended	Exploration recommended	Feasible, but expensive	Not recommended
Proxy service using adaptive translation	Exploration recommended	Exploration recommended	Exploration recommended	Exploration recommended	Not recommended