Determining the true value of a website: A GSA case study

It’s more complicated than you might think

Aaron Meyers
GSA

Ana Monroe
GSA

Apr 16, 2024

Join a Community

Join the Communicators
Join the Contact Center
Join the Multilingual
Join the Plain Language
Join the Social Media
Join the User Experience
Join the Web Analytics
Join the Web Managers
See all 34 communities →

Related Topics

Analytics
Customer experience
Usability

Email Print

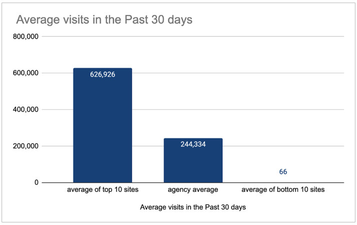
Cleaning up: A hypothetical scenario

Consider this scenario: you’ve been told to clean up a giant room full of *Things Your Agency Has Made in the Past and Now Maintains for Public Use*. This means disposing of the *Things* that no longer add value, and sprucing up the *Things* that are still useful. How do you determine which *Things* belong in which category, especially when all the *Things* in that giant room have been used by the public, and available for all to see?

When the “things” we’re talking about are websites, this determination is often much more complicated than it might appear on the surface. This scenario is one facing web teams across the government, including at the U.S. General Services Administration (GSA), every single day. If you’re in this situation, consider all the ways you might begin to tackle this cleanup job.

Evaluating by visits

You decide to start by determining how many people visit each website each month. Delighted, you pull those numbers together and produce a chart that looks something like this:



The chart states that the 10 least-visited GSA websites had only about 66 visits in the past 30 days, whereas the top 10 websites averaged over 629,000 visits, and the agency average websites averaged over 244,000 monthly visits. So there you have it: clearly, it appears the websites with only 66 visits are the least useful and should be decommissioned. (Note that the low-traffic websites all show 66 visits because of the analytics tool’s statistical sampling methodology.)

However, you stop to examine one of the low-traffic sites. In studying it, you realize that it was never designed to have many visitors. Instead, it was designed to support a very small audience that only appears at random, unpredictable intervals; say, when a natural disaster strikes. Clearly, you don’t want to get rid of that website, since it’s meeting a specific need of a small but well-defined and important audience.

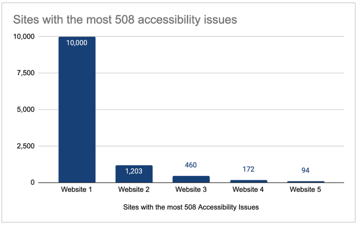
Through this consideration, you realize that using the number of visitors to determine the usefulness of a website incorrectly assumes:

1. Each visit across all your websites is of the same value.
2. Each audience, whether 66 people, or 629,000, have the same level of urgency and need for each website, even if one website is intended to serve a large, continuous audience, while another is designed to serve a small, irregular audience.

Since both of these assumptions are false, visitor numbers are not enough to determine the usefulness of a website. You need another evaluation tactic.

Evaluating by accessibility

After some consideration, you realize that all the websites have to be fully accessible to everyone, regardless of ability. You also have the tools and processes to help determine whether that standard has been reached. Excited, you start by assembling and running your automated accessibility tests.



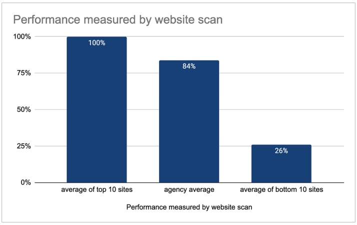
Five websites stand out as having the worst accessibility errors, according to your tests. Clearly, these websites must go. As you prepare to get rid of them, however, you notice that the vast majority of the errors in the worst website are identical and all seem to originate from the same part of the website. You look closer and realize that the problem causing all those errors is actually quite basic and can be fixed easily, taking the worst website out of the bottom ranking. Looking at the other websites in your list, you realize that other errors that have surfaced are only errors in an automatic test, not a human one. Many of them aren't on critical paths for the website's use, so while they should be addressed, they are not meaningfully blocking access to the website.

That throws your entire evaluation into question: how can you possibly batch and judge the usefulness of a website by accessibility, if the severity and impact of each accessibility error varies so much? Instead, you must pair automated accessibility tests with manual testing to reach conclusions on the least accessible websites. That won't help you quickly get rid of the lowest value websites, so yet another evaluation tactic is needed.

Evaluating by speed and performance

After considering the number of visits and the accessibility, you realize that an evaluation of usefulness needs to consider a basic question: is the performance and speed of the website reasonable? If a product is so frustratingly slow that people don't use it, then nothing else matters.

To figure out which websites are so slow as to be essentially non-functional, you find a free online tool that tests website performance. Additionally, you get smart based on your previous experiments: this tool tests for a few different parameters, not just one element of performance. It then compiles these parameters into a single index score, so its results are compelling.



This performance metric shows you that, on average, your websites perform at 84% of a perfect 100% score, and there are a few low-performing websites at 26% performance or lower. This works for you; you know you need to get rid of your agency's low-performing websites. As you're planning to decommission these sites, however, a user visits one of them to complete a task and provides some feedback.

Evaluating by customer research

The user waits while the website slowly loads. Then, they interact with the website and exit the page. To gauge their satisfaction, you prompt them to give you feedback on the page by asking, "Was this page helpful?" The user shares:

> "This website does work; it just works slowly. I'm willing to wait, though, because I need the information. There's nowhere else to get this information, so please don't get rid of this website; I have to come back and get information from it every month."

After taking this customer research into account, you realize that visits, accessibility, performance, and speed do not, on their own, fully reflect the website's value, so you still don't know which websites to decommission.

What next?

At this point, you've discovered that evaluating websites is a multidimensional problem — one that cannot be determined by a single, simple metric. Indeed, even when you consider several metrics, your conclusions lack a customer's perspective.

Determining the value of agency websites therefore must use an index that is not just composed of similar metrics (like the performance index) but is in fact a composite index of different datasets of different data types. This approach will allow you to evaluate the website's purpose, function, and ultimately, value, to your agency and your customers. This aggregation of dataset types is known as a composite indicator.

Methodology: The Enterprise Digital Experience composite indicator

This is the story of evaluating websites in GSA. Websites seem simple to evaluate: do they work or not? But in truth, they are a multidimensional problem. In taking on the definition and evaluation of GSA public-facing websites, the Service Design team in GSA's Office of Customer Experience researched and designed a composite indicator of multiple data sets of different types to evaluate the value of websites in GSA. Since 2021, we've been doing this by examining six things:

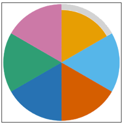
1. **Accessibility**, scored by our agency standard accessibility tool (*quantitative data, 21st Century IDEA Section 3A.1*)
2. **Customer-centricity**, scored by a human-centered design interview (*qualitative data, 21st Century IDEA Section 3A.6 and OMB Circular A-11 280.1 and 280.8*)
    - **Stated audience**: Can the website team succinctly and precisely name their website's primary audience?
    - **Stated purpose**: Can the website team succinctly and precisely name their website's primary purpose?
    - **Measurement of purpose**: Does the website have a replicable means to measure if the website's purpose is being achieved?
    - **Repeatable customer feedback mechanism**: Does the website team have a repeatable customer feedback mechanism in place, such as an embedded survey, or recurring, well-promoted and attended meetings, or focus groups with customers? (Receiving ad hoc feedback from customer call centers or email submissions does not meet this mark.)
    - **Ability to action**: Does the website team have a skillset that can contribute to rapidly improving the website based on feedback and need, such as human-centered design research, user experience, writing, or programming skills?
    - **Ability to measure impact**: Does the website team have the ability to measure the impact of the improvements they implement? Have they devised and implemented a measurement methodology specifically for their changes (an ability to measure impact) or do they rely solely on blanket measures such as Digital Analytics Program data (no ability to measure impact)?
3. **Performance and search engine optimization**, scored by Google Lighthouse (*quantitative data, 21st Century IDEA Section 3A.8*)
4. **Required links**, scored by the Site Scanning Program's website scan (*quantitative data, 21st Century IDEA Section 3A.1 & 3E*)
5. **User behavior, non-duplication**, scored by Google Analytics with related sites (*qualitative + quantitative data, 21st Century IDEA Section 3A.3*)
6. **U.S. Web Design System implementation**, scored by Site Scanning Program's website scan (*qualitative + quantitative data, 21st Century IDEA Section 3A.1 & 3E*)

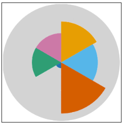
References

View all sections of the law and the circular mentioned above:

- 21st Century IDEA (Public Law No. 115-336)
- OMB Circular A-11 (PDF, 385 KB, 14 pages, 2023)

We visualize this evaluation in website maps, rendered as charts that are available internally to GSA employees. This helps us see examples of good performers, such as Website A (on the left), and not-so-good performers, like Website B (on the right.)





In addition, these charts, like all maps [1], contains some decisions that prioritize how the information is rendered. They include:

- **An equal weight to all datasets and data types, regardless of fidelity**. In the charts above, the slices spread out from 0 along even increments. Our measurement of customer-centricity gives equal weight to whether a site proactively listens to their customers, as well as to whether it has the resources to implement change.

- **A direct comparison by slice**. For example, our customer-centricity slice gives the same amount of distance from the center for listening to its customers as our required links slice gives for including information about privacy, regardless of the fact that customer listening is foundationally different (and more complicated) as an activity than including required links.

We made these decisions because to weight all of the metrics would be to travel down the coastline paradox [2], meaning: we had to identify a stopping point for measurement and comparison that is somewhat arbitrary because, paradoxically, the more closely we measure and compare, the less clear the GSA digital ecosystem would become. These measures are the baseline because, broadly, they are fair in their unfairness: some things are easier to do, and some things are harder, but what is "easier" and what is "harder" differs depending on the resources available to each website team.

But even in comparing websites using charts and maps containing multiple dataset types, we're missing some nuance. "Website A" is a simple, informational site, whereas "Website B" contains a pricing feature, which introduces additional complexities that are more difficult to manage than simple textual information. To give visibility to this nuance, the Service Design team uses these maps as part of a broader website evaluation package, which includes qualitative research interviews and subsequent evaluation write ups. These are sent to every website team within three weeks after we conduct the research interview. Taken together, the quantitative and qualitative data in the website evaluation packages allow GSA staff to consistently measure how digital properties are functioning, and what their impact is on customers.

Concluding which websites should exist

The reality is: value exists in dimensions, not in single data points, or even in single datasets. To further complicate things, the closer you look at single datasets, the more your decision-making process is complicated, rather than clarified. This is because each data type and each data point in complex systems can be broken down into infinitely smaller pieces, rendering decisions made based on these pieces more accurate, but also of smaller and smaller impact. [3]

None of the measures in the Enterprise Digital Experience composite indicator or their use as a whole pie results in an affirmation or denial of the value of a digital property to the agency or to the public; value will always exist as an interpretation of these datasets. The indicator can tell us how existing sites are doing, but not whether we should continue supporting them.

To understand whether a website is worth supporting and how to evolve it, the Service Design team pairs qualitative and quantitative data with mission and strategic priorities to evaluate which websites to improve, and which to stop supporting. To achieve this pairing, three elements must come together:

1. Technical evaluations
2. Regular dialogue with each website's customers, including internal stakeholders and leadership
3. Enterprise-level meta-analysis of a digital property's functions in comparison to other digital properties

Customer dialogue is the responsibility of each team, and technical evaluations are readily available, thanks to tools like the Digital Analytics Program (DAP), but enterprise-level meta-analyses require a cross-functional view. This view can be attained through matrixed initiatives like GSA's Service Design program, or cross-functional groups like GSA's Digital Council, in collaboration with program teams and leadership.

From an enterprise perspective, the next phase in our evaluation of GSA properties is to apply service categories to each website, to better understand how GSA is working along categorical lines, instead of businesses or brands. Taxonomical work like this is the domain of enterprise architecture. Our service category taxonomy was compiled by using the Federal Enterprise Architecture Framework (FEAF) [4] as a starting point, and crosswalks a website's designed function with its practical function, evaluated through general and agency use.

We're starting to leverage service categories, and working with teams to create a more coalesced view of website value as we do so.

What can I do next?

Review an introduction to analytics to learn how metrics and data can improve understanding of how people use your website.

If you work at a U.S. federal government agency, and would like to learn more about this work, reach out to GSA's Service Design team at customerexperience@gsa.gov.

Disclaimer: All references to specific brands, products, and/or companies are used only for illustrative purposes and do not imply endorsement by the U.S. federal government or any federal government agency.

Footnotes

1. Wu HY, Niedermann B, Takahashi S, Roberts MJ, Nöllenburg M. "A Survey on Transit Map Layout - from Design, Machine, and Human Perspectives." Comput Graph Forum.

2020 Jun;39(3):619-646. doi: 10.1111/cgf.14030. Epub 2020 Jul 18. PMID: 33041405; PMCID: PMC7539984 ↩

2. High resolution. 2012. What Is The Coastline Paradox? Hosted by Veritasium, YouTube. ↩
3. Stoa, Ryan, The Coastline Paradox (August 30, 2019). Rutgers University Law Review, Volume 72, Issue 2 (2020), Available at SSRN: or ↩
4. Various, including Tupper, Charles D, 2011. “Enterprise Architecture Frameworks and Methodologies.” ScienceDirect. ↩

Was this page helpful?

Join 18,000 others in government and subscribe to our newsletter — a round-up of the best digital news in government and across our field.

Email *

name@example.com

☐ By checking this box, you consent to our data privacy policy. *

Submit

Return to top ↑

Learn

- Blogs
- Events
- Resources
- Communities
- Guides
- Job board

Contribute

- Write for us
- Host an event
- Submit GitHub issue

About

- About Us
- Site Policies
- Contact Us
- Content schedule

Social

- X (Formerly Twitter)
- Facebook
- YouTube
- LinkedIn
- RSS

GSA Digital.gov

An official website of the U.S. General Services Administration

About GSA
Accessibility statement
FOIA requests
No FEAR Act data
Office of the Inspector General
Performance reports
Privacy policy

Looking for U.S. government information and services? **Visit USA.gov**