

Site Scanning

*Collecting automated real-time data
about federal government websites*



Office of Solutions | Data & Analytics Portfolio
Technology Transformation Services

The data is the product

The numbers

- 25,282 websites
- 1,272 domains
- 143 agencies
- 54 data points per website

The analysis

- Website Mapping/Inventory
- US Web Design System
- Digital Analytics
- SEO
- Third Party Services

How the scanning works



The Process

- Hypothesis
- Research
- MVP
- Feedback/iterate
- Validate
- Stabilize and ensure



The Scan Itself

- Combine .gov registry + open datasets to find all known websites
- Run daily automated scans of the target list
- Makes scan data available via API and bulk (CSV) download



Principles

- Open sources
- What is public on the internet
- Open data
- Real user need

Major Users

- Digital Analytics Program - User adoption; Analysis of implementation
- Search.gov - Preparatory work for new implementations; State of .gov SEO
- US Web Design System - User adoption; Analysis of implementation
- DigitalDashboard.gov - Inventory of Federal Websites; USWDS compliance
- Cybersecurity and Infrastructure Security Agency - .gov Domain

Available Now

- Program website - digital.gov/site-scanning
- Data downloads/API documentation - open.gsa.gov/api/site-scanning
- Contact the team - site-scanning@gsa.gov

