Site Scanning

Collecting automated real-time data about federal government websites

The data is the product

The numbers

- **27,655** websites
- **1,322** domains
- **148** agencies
- **70** data points per website

The analysis

- Website/Service Inventory
- US Web Design System
- Digital Analytics Program
- Third Party Services
- 21st Century IDEA compliance
- IPv6, HTTPS, cloud

GSA

How the scanning works

The Process

- Hypothesis
- Research
- MVP
- Feedback/iterate
- Validate
- Reliability

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

Example Use Cases

- **Digital Analytics Program** - User adoption; Analysis of implementation

- **TTS Digital Council** - Utilization of TTS programs by federal agencies, Market research

- **GSA Customer Experience Team** - Inventory of GSA websites, DAP/USWDS compliance

- **GSA IT** - Adoption of IPv6

- **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

GSA

Data 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

GSA