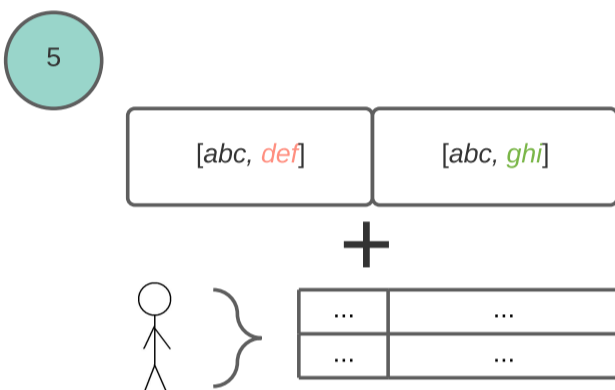# EDGI Environmental Data & Governance Initiative

**Step # 1**

EDGI Web Monitoring chooses which webpages to monitor

- epa.gov/climatechange
- doi.gov/scientific integrity
- niehs.nih.gov/research
...

**EDGI Web Monitoring In Action**

7 — The Intercept_ / CLIMATE CENTRAL / E&E NEWS

6 — ? ✓

5 — [*abc, def*]  [*abc, ghi*] + ... ... / ... ...

2 — Climate.gov

3 — Wayback and/or Versionista → AWS

4 —
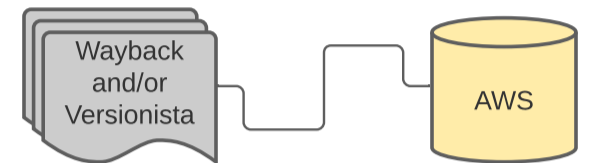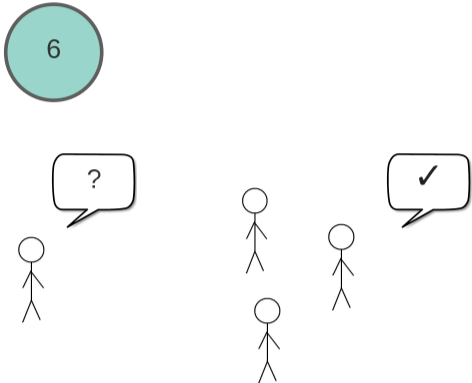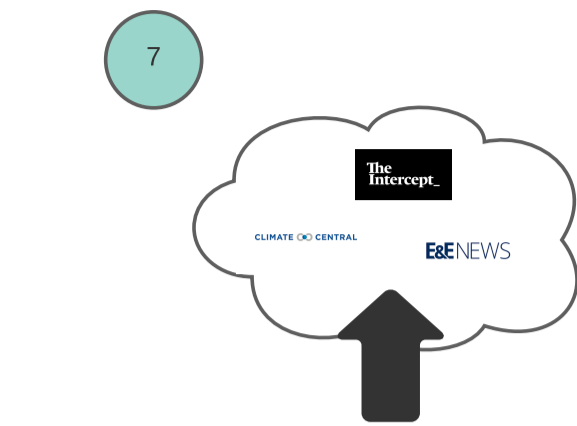
| PageID | URL | Information about change |
|--------|-----|--------------------------|
| PageID | URL | Information about change |
| PageID | URL | Information about change |

---

Key terms:

- **AWS** - Amazon Web Services
- **Wayback Machine** - Digital archive of the web
- **Versionista** - Web change tracking software

■ **Step 1: Choosing webpages to monitor**

The EDGI Web Monitoring team creates and maintains a list of environment and climate related government websites to monitor, consulting analysts based on their monitoring experience as well as their subject expertise. The team also monitors webpages of special interest based on regulatory proceedings and current events. Journalists and other EDGI team members also suggest new webpages to monitor.

■ **Step 2: Archiving**

EDGI pulls data from web services, such as the Internet Archive's Wayback Machine, that archives versions of webpages.

■ **Step 3: Comparison**

Tools created by developers at EDGI then compare the source code of different versions of a webpage. These versions, which were archived at different times, oftentimes contain differences in their code which might be significant. Information about changes a webpage has undergone, represented by a earlier and later version of the same webpage, are exported into datasets for further analysis.

■ **Step 4: Metadata creation**

A priority algorithm ranks webpages that underwent changes by their importance, using criteria such as changes to certain parts of a webpage or the number of words that were added or removed. This and other metadata are made available to analysts via Google Sheets.

■ **Step 5: Rendering for analysis**

Webpage comparison tools created by EDGI developers visualize webpage changes through different renderings of the two versions' HTML code. Without these tools, EDGI's web monitoring work would be nearly impossible.

Using these tools, EDGI WM analysts review changes to webpages to further determine if they are significant. They annotate the metadata about these changes they believe deserve further analysis and share their findings with their peers to discuss further.

■ **Step 6: Meeting and Discussion**

After discussion and further investigation, members of EDGI's Web Monitoring team decide whether to conduct further research about a change or set of changes.

These endeavors might either lead to a brief report outlining observed changes or a longer research publication covering a larger set of changes observed accross multiple domains.

■ **Step 7: Publishing reports**

EDGI publishes and distributes its reports, eliciting feedback from a broader community of journalists, concerned citizens, and academics. Reaction to EDGI's reporting informs its methods and which webpages it decides to add to the list it monitors.