Open Source America's Operating System
"It's Not Just A Good Idea—It's The Law!"

An Audit of .Gov
Statement of Work - Phase I

The Federal Government operates a large number of web sites, FTP sites, and other points of presence on the Internet. This project suggests a number of Phase I activities that can lead to a full audit of these Internet sites with a view of providing feedback to policy makers on the relative strengths and weaknesses of various agencies.

An audit of a web site needs to take into account a variety of factors. We propose an extensible framework that identifies a target domain, such as a government agency, and conducts a number of different measurement exercises at different points in time.

- For example, one can measure the total number of HTML pages on a web site as well as the number of broken links. This measurement can be conducted weekly or on–demand.

- While the broken link analysis is based on an automated script, another measurement exercise might be to poll a set of usability experts for their subjective opinions about the usability of a particular site.

The measurement activity is thus somewhat similar to the process of generating statistics for a web site, where two activities coexist. The first activity is the collection of various statistics, in the form of log files in the case of a stats package for a web site. The second, independent from the first, is the periodic generation of various reports.

Our goal is to be able to generate reports at the level of major federal agencies, such as the Department of Commerce or the Securities and Exchange Commission. It is also our goal to be able to generate reports within departments, such as comparing the relative performance of Immigration and FEMA within the Department of Homeland Security.

In Phase I, we propose that the initial investigation be applied on the Internet Archive's .gov crawl. This archive of government web sites was performed in response to the cessation of those activities by the current government. We hope to demonstrate that an active archive of government web sites is needed not only for posterity, but also provides useful immediate policy guidance.

In Phase I, there are 3 sets of deliverables:

1. A data directory consisting of statistics gathered during this pilot phase. At a minimum, this directory should contain descriptive information (e.g., names of agencies) and an analysis of broken links, HTML validity, CSS validity, and a simple section 508 accessibility analysis.

2. An architecture document that spells out the format of data, the software tools that were used, any lessons from Phase I, and a proposal for providing this as a comprehensive production service in Phase II.

3. A set of sample reports generated from the data gathered.

Phase I is anticipated to last 60–90 days and will cost $25,000.