**USASpending.gov**

# Usability Test Plan

**1.0.1**

**Heidi McGuigan**

**02/08/2016**

## Table of Contents

[Document Overview 3](#_Toc359583313)

[Executive Summary 3](#_Toc359583314)

[Methodology 4](#_Toc359583315)

[Participants 4](#_Toc359583316)

[Training **Error! Bookmark not defined.**](#_Toc359583317)

[Procedure 4](#_Toc359583318)

[Roles 5](#_Toc359583319)

[Trainer **Error! Bookmark not defined.**](#_Toc359583320)

[Facilitator 5](#_Toc359583321)

[Data Logger **Error! Bookmark not defined.**](#_Toc359583322)

[Test Observers 5](#_Toc359583323)

[Ethics 5](#_Toc359583324)

[Usability Tasks 6](#_Toc359583325)

[Usability Metrics 6](#_Toc359583326)

[Scenario Completion 6](#_Toc359583327)

[Critical Errors 6](#_Toc359583328)

[Non-critical Errors 6](#_Toc359583329)

[Subjective Evaluations 7](#_Toc359583330)

[Scenario Completion Time (time on task) **Error! Bookmark not defined.**](#_Toc359583331)

[Usability Goals 7](#_Toc359583332)

[Completion Rate 7](#_Toc359583333)

[Error-free rate 7](#_Toc359583334)

[Time on Task (TOT) **Error! Bookmark not defined.**](#_Toc359583335)

[Subjective Measures 7](#_Toc359583336)

[Problem Severity 7](#_Toc359583337)

[Impact 8](#_Toc359583340)

[Frequency 8](#_Toc359583341)

[Problem Severity Classification 8](#_Toc359583342)

[Reporting Results 8](#_Toc359583343)

## Document Overview

This document describes a test plan for conducting recurring usability testing during the development of USASpending.gov. The goals of usability testing include establishing and validating user performance measures, and identifying potential design concerns to be addressed in order to improve the efficiency, productivity, and end-user satisfaction for both government agencies and public users.

The usability test objectives are:

* To determine design inconsistencies and usability problem areas within the user interface and content areas. Potential sources of error may include:
  + Navigation errors – failure to locate functions, excessive keystrokes to complete a function, failure to follow recommended screen flow.
  + Presentation errors – failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities.
  + Control usage problems – improper toolbar or entry field usage.
* Exercise the application or web site under controlled test conditions with representative users. Data will be used to access whether usability goals regarding an effective, efficient, and well-received user interface have been achieved.
* Establish baseline user performance and user-satisfaction levels of the user interface for future usability evaluations.

USASpending.gov has an array of user groups who are each associated with distinct goals and tasks; as outlined in the Personas developed for the project. The initial site audience to be tested is the government/Agency user who will upload their agency budgetary data into the Data Broker tool in order to comply with the DATA Act. As site development progresses, the focus of testing will shift to data consumers at all levels; from agency users monitoring for compliance to data aggregators (who analyze and re-package data), and from researchers/reporters to interested citizens. Usability testing will occur both in a lab and in realistic environments on people’s personal and work devices. Testing will repeat on a regular cadence that sync’s with the development cycles of our Agile development environment.

## Executive Summary

This usability test plan for USASpending.gov captures an intent to regularly validate what has been recently developed for the site, and to evaluate site designs that will soon go into development. Participants will test site content, and the results of these tests will inform user stories that address weaknesses identified through testing. To fully implement this plan the USASpending.gov UX team must create usability test scripts, criteria for evaluating task completion and success, and a means for collecting data from participants.

Upon review of this usability test plan, including usability goals for USASpending.gov, documented acceptance of the plan is expected.

## Methodology

### Participants

The demographic details of test participants will be dependent upon what has most recently been developed or is soon to go into development, and the expected end-users for each function within the site. Participants will loosely correspond the project personas in both skills and general use cases for the site. Due to the planned frequency of testing, the goal is to gain insights from at least 5-15 participants each month. Actual agency users with knowledge of the DATA Act and the schema will be required for the initial test group. We expect that participants will be found within Treasury “partner agencies” and DATA Act early adopters, like those who participate in the weekly Office Hours calls. As we test other specific groups, recruiting strategies will need to be devised for each user type/persona. Demographic data and test results will be recorded for each participant. Findings from test sessions will inform user stories that address issues and will be populated in the DATA Act Open JIRA environment.

The participants' responsibilities will be to attempt to complete a set of representative task scenarios presented to them in as efficient and timely a manner as possible, and to provide feedback regarding the usability and acceptability of the user interface. The participants will be directed to provide honest opinions regarding the usability of the application, and to participate in post-session subjective questionnaires and debriefing.

### Facilitator

Participants will receive an overview of the usability test procedure, and goals for each session. A premise for user testing will be read aloud as a script or conveyed to participants to inform them of the purpose of testing and their role in the process.

### Procedure

Participants will either take part in Sandbox Testing at the US Mint or testing from their regular office or environment. Testing will be geared towards desk/laptop devices within typical office/home environments. The participant’s interaction with the Web site/Web application will be monitored by a facilitator; either seated in the same office or co-located through a remote connection.

The facilitator will brief the participants on the Web site/Web application and instruct the participant that they are evaluating the application, rather than the facilitator evaluating the participant. The session will be annotated but their privacy of identification will be safeguarded. The facilitator will ask the participant if they have any questions.

The facilitator will instruct the participant to ‘think aloud’ so that a verbal record exists of their interaction with the Web site/Web application. The facilitator will observe and enter user behavior, user comments, and system actions in the data logging spreadsheet.

After each task, the participant will elaborate on the task session with the facilitator.

**Remote Testing**

Participants will take part in the usability test via remote screen-sharing technology. The participant will be seated at their workstation in their work or home environment. Verbal communication will be supported via telephone.

The facilitator will brief the participant and instruct that he or she is evaluating the Web site/Web application, rather than the facilitator evaluating the participant. Participants will respond to pretest demographic and background information questions. Sessions will begin when the facilitator answers all participant questions. The facilitator will inform the participant that exploratory behavior outside the task flow should not occur until after task completion.

The facilitator will observe and enter user behavior and comments, and system interaction in a data logging application.

After each task, the participant will elaborate on the task session with the facilitator.

## Roles

The roles involved in a usability test are as follows. An individual may play multiple roles and tests may not require all roles.

### Facilitator

* Provides overview of study to participants
* Defines usability and purpose of usability testing to participants
* Assists in conduct of participant and observer debriefing sessions
* Responds to participant's requests for assistance

### Subject Matter Experts

* Provides technical expertise about the specifications of DATA Act requirements and reporting specifications.

### Test Observers

* Silent observer
* Assists the data logger in identifying problems, concerns, coding bugs, and procedural errors
* Serve as note takers.

**Test Participants**

* Provides overview of study to participants
* Defines usability and purpose of usability testing to participants
* Assists in conduct of participant and observer debriefing sessions
* Responds to participant's requests for assistance

### Ethics

All persons involved with the usability test are required to adhere to the following ethical guidelines:

* The performance of any test participant must not be individually attributable. Individual participant's name should not be used in reference outside the testing session.
* A description of the participant's performance should not be reported to his or her manager.

## Usability Tasks

Usability tasks are derived from test scenarios developed from use cases and/or with the assistance of a subject-matter expert. Due to the range and extent of functionality provided in USASpending.gov, and the short time with each participant, tasks will cover the most common and relatively complex of available functions. Identical tasks will be tested with all participants of a given user role in the study.

The task descriptions below will be reviewed by the Product Owner to ensure that the content, format, and presentation are representative of real use and substantially evaluate the total application. Their **acceptance is to be documented** prior to usability test.

## Usability Metrics

Usability metrics refers to user performance measured against specific performance goals necessary to satisfy usability requirements. Scenario completion success rates, adherence to dialog scripts, error rates, and subjective evaluations will be used.

### Scenario Completion

Each scenario will require, or request, that the participant obtains or inputs specific data that would be used in course of a typical task. The scenario is completed when the participant indicates the scenario's goal has been obtained (whether successfully or unsuccessfully) or the participant requests and receives sufficient guidance as to warrant scoring the scenario as a critical error.

### Critical Errors

Critical errors are deviations at completion from the targets of the scenario. Obtaining or otherwise reporting of the wrong data value due to participant workflow is a critical error. Participants may or may not be aware that the task goal is incorrect or incomplete.

Independent completion of the scenario is a universal goal; help obtained from the other usability test roles is cause to score the scenario a critical error. Critical errors can also be assigned when the participant initiates (or attempts to initiate) and action that will result in the goal state becoming unobtainable. In general, critical errors are unresolved errors during the process of completing the task or errors that produce an incorrect outcome.

### Non-critical Errors

Non-critical errors are errors that are recovered from by the participant or, if not detected, do not result in processing problems or unexpected results. Although non-critical errors can be undetected by the participant, when they are detected they are generally frustrating to the participant.

These errors may be procedural, in which the participant does not complete a scenario in the most optimal means (e.g., excessive steps and keystrokes). These errors may also be errors of confusion (ex. initially selecting the wrong function, using a user-interface control incorrectly such as attempting to edit an un-editable field).

Noncritical errors can always be recovered from during the process of completing the scenario. Exploratory behavior, such as opening the wrong menu while searching for a function, will be coded as a non-critical error.

### Subjective Evaluations

Subjective evaluations regarding ease of use and satisfaction will be collected via questionnaires, and during debriefing at the conclusion of the session. The questionnaires will utilize free-form responses and rating scales.

## Usability Goals

The next section describes the usability goals for USASpending.gov.

### Completion Rate

Completion rate is the percentage of test participants who successfully complete the task without critical errors. A critical error is defined as an error that results in an incorrect or incomplete outcome. In other words, the completion rate represents the percentage of participants who, when they are finished with the specified task, have an "output" that is correct. Note: If a participant requires assistance in order to achieve a correct output then the task will be scored as a critical error and the overall completion rate for the task will be affected.

**A completion rate of 80% is the goal for each task in this usability test.**

### Error-free rate

Error-free rate is the percentage of test participants who complete the task without any errors (critical **or** non-critical errors). A non-critical error is an error that would not have an impact on the final output of the task but would result in the task being completed less efficiently.

**An error-free rate of 80% is the goal for each task in this usability test.**

### Subjective Measures

Subjective opinions about specific tasks, time to perform each task, features, and functionality will be surveyed. At the end of the test, participants will rate their satisfaction with the overall system. Combined with the interview/debriefing session, these data are used to assess attitudes of the participants.

## Problem Severity

To prioritize recommendations, a method of problem severity classification will be used in the analysis of the data collected during evaluation activities. The approach treats problem severity as a combination of two factors - the impact of the problem and the frequency of users experiencing the problem during the evaluation.



### Impact

Impact is the ranking of the consequences of the problem by defining the level of impact that the problem has on successful task completion. There are three levels of impact:

* High - prevents the user from completing the task (critical error)
* Moderate - causes user difficulty but the task can be completed (non-critical error)
* Low - minor problems that do not significantly affect the task completion (non-critical error)

### Frequency

Frequency is the percentage of participants who experience the problem when working on a task.

* High: 30% or more of the participants experience the problem
* Moderate: 11% - 29% of participants experience the problem
* Low: 10% or fewer of the participants experience the problem

### Problem Severity Classification

The identified severity for each problem implies a general reward for resolving it, and a general risk for not addressing it, in the current release.

**Severity 1** - High impact problems that often prevent a user from correctly completing a task. They occur in varying frequency and are characteristic of calls to the Help Desk. Reward for resolution is typically exhibited in fewer Help Desk calls and reduced redevelopment costs.

**Severity 2** - Moderate to high frequency problems with moderate to low impact are typical of erroneous actions that the participant recognizes needs to be undone. Reward for resolution is typically exhibited in reduced time on task and decreased training costs.

**Severity 3** - Either moderate problems with low frequency or low problems with moderate frequency; these are minor annoyance problems faced by a number of participants. Reward for resolution is typically exhibited in reduced time on task and increased data integrity.

**Severity 4** - Low impact problems faced by few participants; there is low risk to not resolving these problems. Reward for resolution is typically exhibited in increased user satisfaction.

## Reporting Results

The Usability Test Report will be provided at the conclusion of each usability test cycle. It will consist of a report and/or a presentation of the results; evaluate the usability metrics against the pre-approved goals, subjective evaluations, and specific usability problems and recommendations for resolution as reflected in new JIRA stories. The recommendations will be categorically sized by development to aid in implementation strategy. The reports will be delivered at the conclusion of each usability testing cycle to the Project team and stakeholders at each Sprint demo.