Protocol

1. **Brief Description:** The goal of this experiment is two-fold: a) determine how documents are reused and related to each other in the workflow of modern information workers, and b) determine how best to interrupt users with potentially useful task organizational information. Our hypothesis is that document reuse and other provenance relationships are common, but poorly supported. By exploiting such information, we can radically improve search and recall. We also believe the necessary data can be collected with little burden to the user. This information will help us design better workflow and desktop management tools more closely modeling the workflow and work style of end-users. The data will be used for future system development, publication, and thesis work.

2. **Background and Significance:** Information workers are overwhelmed by the number of tasks and information sources and artifacts they manipulate and work with on a daily basis. There has for some time been a strong need to improve their workflow and information management, as evident by the number of tools and software systems released designed to do just that (ie. Google desktop, TaskTracer, etc.)

While such efforts have made inroads in the problem, we still lack an understanding on a very fundamental level on the challenges that users face, the opportunities we have for improving workflow, and how to interact with users without getting in their way. This study is aimed at helping us answer these questions by providing one of the first longitudinal instrumented observations useful for answering these questions.

3. **Methods and Procedures:**

1. **Recruitment**

   We will recruit subjects via email (see attached). We need to identify a set of likely candidates, up to 50. Our goal is to avoid recruiting too many IT/tech workers and focus on knowledge workers. The populations we will be recruiting from include:

   - Admin assistants at OSU
   - Faculty at OSU
   - Employees from companies in town, Intel, and SmartDesktop
Subjects will be screened to ensure they have experience with applications tracked by our software prototype (TaskTracer). This means Microsoft Windows, internet Explorer, Outlook, and Microsoft Office. Subjects do not need to use all the above, or use them exclusively, but should be knowledgeable in their use, and use at least some of the above.

2. User Study: (3 weeks total)

Selected subjects will be visited by research staff to more clearly explain their participation, their commitment, and their rights. This includes going over and signing the informed consent document, background survey (see attached). The staff member will then install the TaskTracer software which will track the users’ workflow over the course of the experiment, and go over the user manual and set up the system for their own use (approximately 1 hour total).

TaskTracer will track what applications subjects’ use during the course of the experiment, what documents are accessed, and how they are related to each other (aka provenance, how information flows between documents, and how documents are reused). The system will log usage statistics and information, url’s of websites visited, addresses and subjects of emails sent and received, and file and folder names of files accessed. This information is all necessary to accurately model the users’ workflow.

The system will NOT collect any content from the files or emails in order to protect subjects, and clear instructions will be given, and are documented in the user manual to turn off tracking for sensitive activities, or when subjects choose not to be tracked.

Users will be anonymized in the data-sets, and email addresses and names will be removed from the data after the experiment is concluded and before the data is analyzed. This data does need to be collected during the experiment in order to give users access to real data during predictions, and provenance data.

The first 4 days of the experiment, the user needs to train the system to make accurate predictions. This is achieved by tagging documents and activities logged by the system as belonging together in “activities.” The mechanics of this activity will be covered in the initial training session. The activity is expected to take 10-30 minutes at the end of each of the 4 work-days.

On the 5th day, a research member will debrief the subjects, ensuring that they have completed the tasks as assigned, and refreshing their
understanding of the system and its use. An interruptability survey will be administered at this time (see attached).

Subjects will then use the system for 1 week, without the need for end-of-day activities. They will be free to attend to the systems’ warnings and recommendations if they wish. Periodically (every 2 hours ± randomly chosen interval) they will be given a short survey on how approachable they are at the moment (see attached). This is meant to determine what activities are interruptible, and which are not (5 per day, 2 minutes each=10 minutes).

At the end of week 2, another research staff member will disable the interruptability survey, and collect some of their document history. This document history will be used to render some document relationship graphs (see attached) which users will be asked to recall details about at the end of the experiment. We will use this to determine which features of documents are most salient in users’ minds, and how good recall is in general.

At the end of week 3, research staff will perform the interview discussed in the previous paragraph, as well as administer a debriefing interview (see attached questions), and questionnaire (see attached) (total time 1 hour). The interview will be filmed for ease of transcription. TaskTracer will be completely removed from the system, the data removed, and cleaned for further processing.

4. Risks/Benefit Assessment:
The risks of participating in this study are minimal. There is the possibility that participant may experience difficulty first learning to use the system, and/or that the system will experience technical difficulties. We will provide a tutorial and support throughout to help with this problem.

There is no direct benefit to subjects from participating in this study. Their participation will be instrumental to the design of new workflow management systems which will benefit information workers in the future.

Conclusion - Our evaluation is that the benefits outweigh the risks

5. Participant Population:
50 subjects, all professional information workers using Microsoft applications on a daily basis. Subjects will be screened based on their level of experience and use of said applications, and will be recruited via email from within OSU and local companies.
6. **Subject Identification and Recruitment:**
   Email announcement to OSU mailing lists, SAO mailing lists, or company mailing lists (see attached)

7. **Compensation:**
   Participating subjects will receive $40.00 as compensation for their participation. Participants completing only the first 2 weeks of the study will receive $20.00 in compensation.

8. **Informed Consent Process:**
   Informed consent will be collected at the beginning of the study by the research staff (see attached).

9. **Anonymity or Confidentiality:**
   The volunteers’ personal information will be kept private during and after the study. We will generate randomly a 4-digit number to identify every form, tape, and data-set of the participants.

   Every hard copy record: audio, video, forms, etc, will be kept locked in the Principal Investigator or Research Staff member’s cabinet at any time. The logs recorded from the experiment will be saved in a dedicated directory with a limited access to the Primary Investigator and the Research Staff. All personally identifiable and sensitive information will be deleted from the logs before archiving. The Principal Investigator and the Research Staff member will be the only persons who have access to this information.