ABSTRACT
The goal of this one-day workshop is to bring together researchers and practitioners to consider methods that use data stored in software repositories (such as source control systems, defect tracking systems, and archived project communications) to further understanding of software development practices.

1 WORKSHOP THEME AND GOALS
Software repositories contain a wealth of valuable information for empirical studies in software engineering:
- source control systems store changes to the source code as development progresses,
- defect tracking systems follow the resolution of software defects, and
- archived communications between project personnel record rationale for decisions throughout the life of a project.

Such data is available for most software projects and represents a detailed and rich record of the historical development of a software system. Participants in multiple sites, often on multiple continents, develop software projects without ever meeting in person, as is the case in many large commercial and Open Source projects. This trend makes the use of tools to record all aspects of software project more critical.

Until recently, data from these repositories was used primarily for historical record supporting activities such as retrieving old versions of the source code or examining the status of a defect. Several studies have emerged that use this data to study various aspects of software development such as software design/architecture, development process, software reuse, and developer motivation. These studies have highlighted the value of collecting and analyzing this data. Each of these studies has built its own version of methodologies and tools to address the formidable challenge of utilizing such data to perform their empirical research. Several international efforts have identified the development of approaches to extract, share, and study this data as a research priority.

The goal of this one-day workshop is to bring together researchers, practitioners, and source control systems developers to consider methods that use the data stored in software repositories to further understanding of software development practices. The presentations and discussions in this workshop will facilitate the definition of challenges, ideas and approaches to transform software repositories from static record keeping repositories to active repositories used by researchers to gain empirically based understanding of software development, and by software practitioners to predict and plan various aspects of their project.

2 SCOPE AND TOPICS OF INTEREST
The workshop scope is the general field of software repository mining. Relevant topics include but are not limited to the following:
- New approaches to analyze the data stored in software repositories to:
  - Assist in program understanding and visualization
  - Predict and gauge the reliability and quality of software systems
  - Study the evolution of software systems
  - Discover patterns of change and refactoring
  - Understand the origins of code cloning and design changes
  - Model software processes for development, defect repair, etc.
- Assist in project planning and resource allocation
- Case studies on extracting data from these repositories for large long lived projects
- Proposals for common exchange formats, meta-models, and infrastructure tools to ease the sharing of the extracted data and to enable reuse and repeatability of results throughout the research community
- Suggestions for particular large software repositories to be shared among the community for research evaluation and benchmarking purposes
- Approaches to integrate data between repositories and with other software project data such as static or dynamic analysis data
- Requirements and guidelines for users and developers of source control systems to ease the analysis of the stored historical data

Position papers are no more than 5 pages long and are reviewed by the workshop’s program committee in terms of
their relevance to the aims of the workshop and their technical content. Accepted papers are posted on the workshop web site prior to the workshop at:

http://msr.uwaterloo.ca

3 WORKSHOP FORMAT
The program committee selects for presentation papers that can serve as the basis for fruitful discussions. The papers are chosen so that a broad range of stakeholders from across the software engineering discipline are represented in the workshop.

An emphasis is on making the workshop interactive with many discussion slots assigned throughout the schedule. Presentation are short with strict time limits to ensure ample discussion time. In addition, there is a wrap-up session at the end of the day to set goals for further research in the area of mining software repositories.

4 EXPECTED WORKSHOP OUTCOME
The workshop strives to achieve a consensus among the participants as to the structure of the field, the important research directions the field should take, the research needs of the field and possible future benefits of mining such repositories. In addition, we hope to provide workshop participants with a summary overview of the important emergent issues in the field and to identify potential future research collaborators.