ABSTRACT
Software Processes consist of a structured set of activities performed during creation and maintenance of software products. The construction and subsequent maintenance of a software is facilitated by several applications and tools. Some of the tools such as Issue Tracking System (ITS) and Version Control System (VCS) can be classified as Process Aware Information System (PAIS) logging data consisting of events, activities, time-stamp, user or actor and context specific information. Such events or trace data generated by information systems used during software construction (as part of the software development process) contains valuable information which can be mined for gaining useful insights and actionable information. Software Process Intelligence (SPI) is an emerging and evolving discipline involving mining and analysis of software processes. It is modeled on the lines of Business Process Intelligence (BPI), but with the focus on software processes and its applicability in software systems. In this paper, we present a generic framework for Software Process Intelligence and some of its applications.

Categories and Subject Descriptors
H.2.8 [Database Applications]: Data Mining

Keywords
Automated Software Engineering, Business Process Intelligence (BPI), Mining Software Repositories, Process Mining, Software Process Intelligence

1. SOFTWARE PROCESS INTELLIGENCE (SPI) FRAMEWORK
Process mining is an area at the intersection of business process intelligence and data mining consisting of mining event logs from process aware information systems for the purpose of process discovery, process performance analysis, conformance verification, process improvement and organizational analysis. The approaches and algorithms within process mining enables information extraction from event logs or traces generated as a result of execution of a business process [8].

Large and complex software projects use defect tracking systems for managing the workflow of bug reporting, archiving, triaging and tracking. Version control or source code control systems are used to manage changes to project files and documents. Peer code review systems are used to manage peer review of source code before committing the source code to identify defects though inspection. Community based Q&A websites for programmers and online forums are widely used by developers for asking questions and sharing knowledge. Bug databases, version archives, source code repository, peer code review system, community based Q&A websites, mailing lists and online forums for programmers are software repositories containing large volumes of valuable structured data and unstructured data (free-form text) entered by developers during the software development process. These repositories have been primarily serving the purpose of archiving information or recording keeping. Mining Software Repositories (MSR) researchers have investigated social network analysis, data mining, machine learning and information retrieval based approaches to analyze software repositories to uncover interesting patterns and knowledge which can be used to support developers in the process of software maintenance. The work on Mining Software Repositories is based on the premise that historical data present in software repositories can be mined to derive actionable information resulting in increased productivity and effectiveness of developers [1].

Software Process Intelligence (SPI) is an emerging and evolving discipline involving mining and analysis of software processes. This is modeled on the lines of application of Business Intelligence techniques to business processes (Business Process Intelligence (BPI)), but with the focus on software processes and its applicability in software engineering and information technology systems. Software Process Mining or falls at the intersection of Software Process & Mining, and Software & Process Mining. It is a three word phrase which can be viewed from two perspectives: Software + Process Mining and Software + Mining. Figure illustrates Software Process Intelligence (SPI) as an intersection of: Software Processes, Business Process Intelligence (BPI) and Software Archives.

2. SOFTWARE PROCESS INTELLIGENCE (SPI) APPLICATIONS
Some of the business applications of process mining soft-
ware repositories or SPI are: uncovering runtime process model, discovering process inefficiencies and inconsistencies, observing project key indicators and computing correlation between product and process metrics, extracting general visual process patterns for effort estimation and analyzing problem resolution activities [2][3][4][6][7].

Mittal et al. present an approach for mining the process data (process mining) from software repositories archiving data generated as a result of constructing software by student teams in an educational setting [5]. They present an application of mining three software repositories: team wiki (used during requirement engineering), version control system (development and maintenance) and issue tracking system (corrective and adaptive maintenance) in the context of an undergraduate Software Engineering course [5]. Gupta et al. present an application of process mining three software repositories (ITS, PCR and VCS) from control flow and organizational perspective for effective process management [4]. They discover runtime process model for bug resolution process spanning three repositories using process mining tool, Disco, and conduct process performance and efficiency analysis. They identify bottlenecks, define and detect basic and composite anti-patterns. In addition to control flow analysis, they mine event log to perform organizational analysis and discover metrics such as handover of work, subcontracting, joint cases and joint activities [4]. Gupta et al. apply business process mining tools and techniques to analyze the event log data (bug report history) generated by an issue tracking system with the objective of discovering runtime process maps, inefficiencies and inconsistencies. They conduct a case-study on data extracted from Bugzilla issue tracking system of the popular open-source Firefox browser project [2]. Gupta et al. present an application of process cube to software defect resolution process to analyze and compare process data from a multi-dimensional perspective. Process cube facilitates process mining from multiple-dimensions and enables comparison of process mining results across various dimensions. They present a framework, a novel perspective to mine software repositories using process cube. They conduct a case-study on Google Chromium project data in which the software defect resolution process spans three software repositories: ITS, Peer Code Review System (PCR) and VCS. [3].

3. REFERENCES