



Center of Excellence for Traceability

Problem Statements and Grand Challenges

Center of Excellence of Traceability Technical Report COET-GCT-06-01-0.9 September 10, 2006

Editors: Jane Cleland Huang (De Paul University)
Alex Dekhtyar (University of Kentucky)
Jane Huffman Hayes (University of Kentucky)

Authors: Giuliano Antoniol (Ecole Polytechnique de Montreal)
Brian Berenbach (Siemens)
Alex Eyged (Teknowledge; University of Southern California)
Stephanie Ferguson (NASA)
Jonathan Maletic (Kent State University)
Andrea Zisman (City University of London)

with Participation from: Elizabeth Ashlee Holbrook (University of Kentucky), Senthil Sundaram (University of Kentucky), Cindy Zou (DePaul University), Orlena Gotel (Pace University).

Introduction

This document describes problems and challenges related to various aspects of traceability in software systems. Members of the traceability community from academia, industry, and government, participated in the First Workshop on Grand Challenges for Traceability (GCW'06) held on August 4-5, 2006 at NASA's Independent Verification and Validation Facility in Fairmont, West Virginia, USA. The two-day workshop included presentations, brainstorming sessions, and plenary discussions of traceability related topics, and among other things resulted in the delivery of this strawman version of the Grand Challenges.

Glossary

Term	Definition
Traceability Community	Practitioners and academics who are actively working to improve traceability practices and techniques.
Stakeholders	Researchers, developers, analysts, and business people with an interest in traceability. (For the purposes of this document)
Semantic traceability information	Information that indicates something about the nature of the relationship between elements, such as the type of link, the longevity of the link, the stability of the link.

A. Traceability Knowledge

- A-P1** Traceability is critical to the success of software and systems projects, but there is little consensus on best traceability techniques and methods, few recorded best practices, and a general lack of resources providing a body of knowledge in the field.
- A-P2** Tracing requires communication between stakeholders, but semantic mismatches and disparate use of terminology across various stakeholder groups create communication barriers.

Challenges

- A-GC1** Create a body of knowledge that reflects best practices of traceability experts and practitioners, standard terminology, and additional information such as case studies on traceability.

B. Training and Certification

- B-P1** Traceability is a key success factor for any software or systems project, but very few people are proficient at tracing and there are few educational programs available to impart such proficiency.
- B-P2** Organizations need a way to identify individuals skilled in traceability methods and practices, however few requirements certification programs exist, and few of those include traceability components.
- B-P3** Traceability training programs must teach skills needed by traceability practitioners, however there are no standard traceability skill sets defined.

Challenge(s):

- B-GC1** Identify core knowledge areas and associated skills for traceability.
- B-GC2** Develop effective educational components on the practice of traceability that can be integrated into university, industrial, or certification curriculum.
- B-GC3** Develop effective pedagogical materials to educate managers in the importance and cost-benefits of traceability.

C. Supporting Evolution

- C-P1** Accurate, consistent, complete, up-to-date traceability information is vital to diverse groups of stakeholders working in various domains and applications, however current techniques for link recovery are still human intensive and error prone (e.g., due to documentation quality, level of detail, etc.)
- C-P2** For traceability links to be useful, they must reflect current dependencies between artifacts, however, the cost and effort to maintain links during system evolution is burdensome, and (as a result) the links often erode into an inaccurate state.
- C-P3** Current requirements management tools contain features such as ‘suspect links’ to help analysts manage the evolution of links, but in most non-trivial projects the number of suspect links quickly become excessive, drastically minimizing the usefulness of the suspect link feature.
- C-P4** Traceability links need to synchronously evolve with their related artifacts, however, current change management systems and link semantics are not sufficiently sophisticated to support effective evolution of traceability links.
- C-P5** Product lines and other types of reuse are essential for meeting time-to-market deadlines while maintaining quality goals, however, methods for transforming and reusing trace links synchronously as the product develops are immature.

Challenges

- C-GC1** Develop link recovery techniques for textual artifacts that are at least as accurate as manual processes and are much more time and cost effective.

- C-GC2** Develop incremental, almost real-time, traceability recovery approaches to be integrated into Integrated Development Environments.
- C-GC3** Develop change management systems that effectively support the evolution of traceability links across multiple artifact types.
- C-GC4** Develop techniques for reusing traceability work products.
 - C-GC4.1** Develop techniques that support traceability across products within a product line through maximizing reuse and providing traceability between various versions of the product line.
 - C-GC4.2** Develop techniques for maximizing reuse of traceability links when existing code is reused in a new product.

D. Link Semantics

- D-P1** In order to effectively utilize links and understand the underlying traceability relationships, it is necessary to define the semantics (e.g., type) of a link, however defining a formalism to represent the semantics is a non-trivial task and may be domain-specific.
- D-P2** Knowing and establishing the granularity of the elements being linked is important to the consistency of traceability and strongly influences the cost-benefits of the traceability effort, however granularity is influenced by the organization, domain, project, application, and mood of the stakeholder, etc., and there is no clear cost-benefits model for consistently determining the correct trace granularity.

Challenges

- D-GC1** Define a meta-model to represent semantic information of traceability links and provide examples of instantiation to specific domains.
- D-GC2** Develop techniques and processes for determining the correct granularity of links within a project.
 - D-GC2.1** Formally specify the notion of granularity and develop company-specific, domain-specific, etc., models of granularity.
 - D-GC2.2** Develop an infrastructure that supports experimentation with pairs of artifacts by varying the levels of granularity.
 - D-GC2.3** Perform experiments using the infrastructure on various domains, projects, organizations, applications, etc.
 - D-GC2.4** Based on experimental results, develop guidelines on granularity that may be specific to domain, project, or organization characteristics, or upon individual requirements characteristics such as volatility and criticality.
- D-GC3** Based on experimental evidence, develop modeling techniques to relate system longevity to traceability (e.g. granularity, accuracy, etc.).

E. Scalability

- E-P1** Robust and scalable traceability techniques are necessary for building reliable and maintainable industrial-scale software systems; however, it is suspected that current state-of-the-art techniques will not adequately scale for large projects.
- E-P2** Visualization tools are essential to support the comprehension and usage of large amounts of trace information, however current visualization techniques do not scale well, and are not effective at presenting complex information because they lack sophistication in filtering, navigation, querying, etc.
- E-P3** Current traceability methods have been developed to trace well structured data, however, many industrial datasets are composed of large and unstructured documents that are hard to trace.

Challenges

- E-GC1** Obtain industrial scale datasets from various domains and use these datasets to investigate the scalability of available techniques and, if necessary, to create new approaches that scale more effectively.
- E-GC2** Develop effective visual mechanisms to support navigation and querying of large numbers of traceability links and associated artifacts.
- E-GC3** Develop scalable techniques for tracing.
 - E-GC3.1** Develop techniques to trace across heterogeneous data sets.
 - E-GC3.2** Develop techniques to trace large and weakly structured data sets.

F. Human Factors

- F-P1** Automated traceability methods produce candidate trace links; however, the process is useless if the analyst is not able to correctly evaluate the links to differentiate between good and bad ones, or is unable to trust in the completeness and accuracy of the results.
- F-P2** Understanding how humans use traces is essential for building useful traceability methods and tools; however, we do not have an understanding of the entire tracing life cycle and therefore lack accurate stakeholder usage models.
- F-P3** Ideally, tracing during project development should be invisible (because nobody wants to pay attention to tracing), unfortunately, trace generation and usage is interruptive to humans because in current software development environments it is not possible to automate all of the processes.
- F-P4** Traces bridge semantically different artifacts, but these artifacts are created by different people, often written in different documents (lingo, dialect). As a result, users of traces often do not fully understand artifacts on the “other side” of the links.

Challenges

- F-GC1** Based on studies of stakeholders' use of traceability tools, create traceability tools that meet their practical needs.
- F-GC2** Understand the impact and vulnerabilities of human fallibility on the traceability process and develop techniques to help analysts prevent making mistakes and to minimize the impact of mistakes when they do occur.
- F-GC3** Establish user trust in the traceability tools.
- F-GC4** Develop techniques to help humans overcome the semantic barrier in tracing to artifacts of other stakeholders.

G. Cost Benefit Analysis

- G-P1** Complete traceability, in which links are created between all directly related artifacts at a low level of abstraction, may be desirable for comprehension purposes and is often required for legal or process compliance purposes; however, this level of traceability is often impractical and not cost-effective.
- G-P2** Determining the right traceability fit for a given project can result in traceability support at acceptable cost and effort, however, there is a lack of cost-benefit models to differentiate between different traceability needs across various projects and for different potential links within a project.

Challenges

- G-GC1** Define and develop cost-effective techniques for generating and maintaining traceability information.
- G-GC2** Define a practical cost model for generating and maintaining traces that takes into consideration factors such as project size, time and effort, and system quality.
- G-GC3** Define a benefits model for using traces that takes into consideration factors such as criticality and volatility, and incorporates the value achieved from use of traces.

H. Methods and Tools

- H-P1** Establishing traceability to artifacts stored in multimedia formats is useful for tracing stakeholders' contributions and other such information; however, multimedia retrieval methods are not sufficiently sophisticated to effectively support multi-media trace retrieval and little work has been done to incorporate multi-media techniques into traceability tools.

- H-P2** Trace automation is essential; however, it is made difficult by lack of consistency among artifacts, and imprecision of the models.
- HP-3** Traceability involves the activities of link construction or generation, link assessment, link maintenance, and link usage; however, there is no single tool that can cover all of these tasks.
- H-P4** Non-functional requirements often play a critical role in the success or failure of a project and must therefore be carefully managed; however, their tendency to exhibit global impact upon the system and the complexity introduced through their interdependencies and trade-offs make them difficult to effectively trace.

Challenges

- H-GC1** Develop effective methods for tracing multimedia artifacts.
- H-GC2** Build methods and tools with maximized levels of automation to support the entire trace life cycle including link construction or generation, link assessment, link maintenance, and link use.
- H-GC3** Develop methods for tracing non-functional requirements.

I. Tracing across organizational boundaries

- I-P1** End-to-end tracing is critical to the success of a project, but organizational boundaries (e.g. those between marketing and manufacturing or development, or those due to outsourcing) make it difficult to achieve due to differences in skill sets, process, terminology, and tools.

Challenges

- I-GC1** Define effective end-to-end processes and supporting tools and techniques for cross-boundary traceability such as tracing between outsourcing companies or tracing between different business units within a corporation.
- I-GC2** Understand the issues of tracing across systems-of-systems and develop effective, supporting techniques.
- I-GC3** Build infrastructure to allow tracing across distributed artifacts.

J. Process

- J-P1** In order to generate and maintain quality and sound traceability information, an organizational process is required; however, traceability is often not included as an integral part of the development lifecycle.

- J-P2** Automated tracing can provide a cost-effective alternative to manual tracing, but practice has shown that some data sets are difficult to trace using automated methods due to inconsistencies in terminology, standards, terseness of requirements, lack of structure, heterogeneous formats, etc.

Challenges

- J-GC1** Build process models that define the tracing life cycle.
- J-GC2** Develop techniques to evaluate the ability of a given data set to support automated trace methods.

K. Compliance

- K-P1** Standards can help to ensure consistent and complete processes, however, there are a plethora of standards and it is not clear that traceability researchers or practitioners are fully aware of these standards.
- K-P2** The traceability community (both academics and practitioners) are knowledgeable about traceability techniques and processes, but have little influence over the traceability related content of software engineering process standards.
- K-P3** Although traceability standards and regulations exist, it is unclear how compliance to such standards and regulations can be demonstrated.

Challenges

- K-GC1** Establish a communication mechanism to make the traceability community aware of standards related to traceability.
- K-GC2** Attain a presence in the standards community to influence and/or develop traceability standards.
- K-GC3** As a community, develop and promote validation scenarios to prove that the traceability tools/techniques/methodologies comply with standards.

L. Measurement and Benchmarks

- L-P1** Empirical studies are needed to demonstrate the effectiveness of traceability methods and facilitate collaborative and evolutionary work among researchers and practitioners; however, there is a lack of common experimental design, methodologies, and benchmarks.
- L-P2** Assessing traceability links requires a well-defined set of measures, however, current and proposed measures, methods, and metrics have not been validated through empirical studies or benchmarks.

- L-P3** Having good traceability measurements is important, but good traceability benchmarks do not exist or are not compatible.
- L-P4** Researchers have claimed successes in new traceability methods and techniques they have developed, but there are no benchmarks enabling standard comparisons.
- L-P5** Detecting errors in traceability links is necessary to determine the efficacy of the product and process, however, current error detection models in traceability are primitive and not validated.

Challenges

- L-GC1** Define standard processes and related procedures for performing empirical studies during traceability research.
- L-GC2** Build benchmarks for evaluating traceability methods and techniques.
- L-GC3** Define measures for assessing the quality of individual and sets of traceability links.
- L-GC4** Develop techniques to assess traceability methods and processes.

M. Technology Transfer

- M-P1** The ultimate goal of traceability research is to transfer effective traceability solutions to industry; however, industry is reticent to try new and unproven techniques.
- M-P2** Academic researchers and traceability practitioners all want improved traceability techniques, but lack of dialog between the two groups limits researchers' accessibility to real datasets for testing new techniques, and inhibits feedback loops from industry to researchers.
- M-P3** The potential traceability user community is vast; however, it is not well-defined and consists of people with various skill levels and very different incentives to use traceability.
- M-P4** Traceability prototypes are generally designed to show proof of concept, however, they are often not sufficiently rigorous for field testing in industry.

Challenges

- M-GC1** Create an infrastructure and related methods for organizing the technology transfer process.
- M-GC2** Identify successful case studies and publicize them in order to effectively demonstrate the cost effectiveness and technical effectiveness of the traceability techniques to the industry.
- M-GC3** Identify traceability users and define their needs in terms of quality, life cycle, different user groups, communication, etc).

M-GC4 Incorporate state-of-the-art traceability tools into standard IDE's (such as Eclipse) and industrial requirements management tools.

DRAFT VERSION