FORSETI: A Provenance-aware Visual Analysis Environment for the Lifecycle Management of E-autopsy Reports

Baoqing Wang[1], Noboru Adachi[2] and Issei Fujishiro[1]

[1]Keio University, Graduate School of Science and Technology, Yokohama, Kanawaga 223-8522, Japan
[2]University of Yamanashi, Graduate School of Medical Science, Chuo, Yamanashi 409-3898, Japan

Abstract

Autopsy reports are imperative for both medical and legal science. Medical examiners (MEs) and diagnostic radiologists (DRs) cross-reference autopsy findings, while judicial personnel derive legal documents. In a prior study, we proposed a visual analysis system named FORSETI (forensic autopsy system for e-court instruments) with x-LMML (extended legal medicine markup language) for MEs and DRs to author and review e-autopsy reports. In this paper, we outline our extended work in progress to introduce a provenance infrastructure for forensic data accountability to FORSETI, which can be characterized by two technical essences. The first is a provenance management mechanism that combines the forensic autopsy workflow management system (FAWfMS) and `lmmlgit` (a version control system for x-LMML files), allowing a large amount of provenance information about e-autopsy reports and their documented autopsy processes to be individually parsed. The second is authority management, which ensures the confidentiality of e-autopsy reports by deploying strict syntax-guided workflow controls and a custom-tailored tool.

Keywords

Computational forensics, Legal medicine, Accountability, Provenance, Authority

1. Introduction

In forensic science, the generation and utilization of forensic autopsy reports are intrinsically a collaborative data science activity. Usually, forensic autopsy reports are generated by *medical examiners* (MEs) collaboratively working with *diagnostic radiologists* (DRs); the reports then serve as underlying legal documents for MEs and DRs as well as for *judicial personnel* (JP). In these processes, a large amount of forensic data needs to be collected, visualized, analyzed, and annotated. Thus, much work has been done to develop computational tools and techniques for processing forensic data, including forensic autopsy assistance systems [1, 2], virtual autopsy platforms [3], and languages [4, 1]. However, the use of computational environments for forensic data has raised some critical issues—particularly, how autopsy insights and results are obtained from forensic data, how the confidentiality of forensic data is handled, and how to ensure the trustworthiness of the autopsy results. We elaborate on these concerns in the following.

Forensic autopsy reports are commonly generated and used in physical autopsies (PAs) and virtual autopsies (VAs) to record autopsy results and to cross-reference PAs and VAs. Generally, in the refinement of PA (or VA) results, MEs (or DRs) with different experiences perform back-and-forth analyses of forensic data, while they expend substantial efforts recording provenance information. This manual collection of provenance is time-consuming, laborious, and error-prone. In cross-referencing, MEs and DRs may make biased or inaccurate autopsy decisions. This is because the clues of autopsy insights, which serve as interpretative provenance information inspired by MEs and DRs experiences, are not well provided within the autopsy report. Thus, for MEs and DRs to effectively share knowledge and insights, the development of applications supporting the systematic management and analysis of provenance is necessary. In addition, JP finds existing autopsy reports cumbersome because of the deficiencies of non-derivability.

Multiple stakeholders (MEs, DRs, and JP) are involved in complicated pipelines for handling autopsy reports, where ethics and policies are commonly respected to protect postmortem privacy. These ideological and legal constraints are not sufficient for maintaining forensic information security. Clearly, computational tools and system mechanisms ensuring the confidentiality of autopsy reports are needed. The verifiability and confidentiality of data provenance in forensic autopsy workflows are crucial for establishing data accountability, with which e-autopsy can ensure that data contributors are committed to the truthfulness of the data.

In our prior research [1], we introduced a visual analysis system called FORSETI (forensic autopsy system for e-court instruments) with x-LMML (extended version of

Published in the Workshop Proceedings of the EDBT/ICDT 2022 Joint Conference (March 29-April 1, 2022), Edinburgh, UK
wangbaoqing@keio.jp (B. Wang); fuji@ics.keio.ac.jp (I. Fujishiro)
https://fj.ics.keio.ac.jp/en/member/baoqing-wang (B. Wang); https://fj.ics.keio.ac.jp/en/member/issei-fujishiro (I. Fujishiro)
0000-0002-6184-4245 (B. Wang); 0000-0002-8898-730X (I. Fujishiro)

legal medicine mark-up language). The proposed prototype assists MEs and DRs in authoring and browsing e-autopsy reports using x-LMML, but the research was not targeted at the lifecycle management of e-autopsy reports in terms of data provenance.

In this paper, we outline our work in progress that addresses the aforementioned issues for establishing data accountability by extending the prior research to design a provenance-aware FORSETI, which can be characterized by two technical essences. The first is a provenance management mechanism that combines the forensic autopsy workflow management system (FAWfMS) and `lmmlgit` (a version control system for x-LMML files) to allow a large amount of provenance information about e-autopsy reports and their documented autopsy processes to be individually parsed. The second is authority management, which ensures the confidentiality of e-autopsy reports by deploying strict syntax-guided workflow controls and a custom-tailored tool. The paper concludes with directions for future work in pursuit of a provenance-aware FORSETI.

2. Related Work

This section reviews prior work on provenance systems and authority management, both of which are vital components for the core functionalities of the proposed extension to the current FORSETI system.

Provenance, also known as audit trail, lineage, and pedigree, refers to the entire amount of information composing all the elements and their relationships that contribute to the existence of a set of data [5]. Recently, systematic execution of tasks such as collecting, managing, and analyzing provenance information has received significant attention in a wide range of application fields (e.g., bioinformatics, astronomy, ecology, and geology). In this context, two basic types of systems are usually considered. One is *workflow-based system*, generally known as the scientific workflow management system (SWfMS), which involves the linking of components as a task execution plan in the form of workflows whose computation is abstracted by directed acyclic graphs (DAGs) [6]. For defining task workflows, some SWfMSs, such as VisTrails [7], Swift [8], Kepler [9], and Taverna [10] use their own scripting languages, whose syntax is restricted to support the creation of specific types of DAGs. Thus, the SWfMS lacks the flexibility provided by general-purpose scripting languages. The other is a *script-based system*, which refers to the user's interaction with the data processing components through a sequence of commands entered in the shell interface to track provenance data. These kinds of systems, such as PASS [11], ES3 [12], noWorkflow [13], and Lancet [14], provide users with the flexibility to search for, derive, store, and share provenance information via designated commands. However, because these kinds of systems ignore the structure of the script, the user may find it difficult to link the provenance they have collected to the steps in the script.

Unfortunately, the FORSETI prototype [1] does not support provenance functionalities. We therefore extend our original research to introduce provenance awareness to the FORSETI by taking and using the best of the workflow-based and script-based provenance approaches to lifecycle management of x-LMML files and their associated processes.

On the other hand, authority management refers to access control among users for preventing illegal information leaks. Authority management is essential to maintain the confidentiality and objectivity of collaborative data science activities. Our authority management design is mainly inspired by electronic health record systems (EHRS) [15], where authorized information providers can create and manage patients' health information in a digital format (EHR) such that they can be shared with other authorized providers in more than one healthcare organization. Note that the syntax of the derivation relationship is the major difference between our system and EHRS. In contrast to EHRS used in medical organizations, our system needs to satisfy the usage of both medical and legal organizations. In addition, numerous compliance regulations require audit logs for electronic records. The Health Care Portability and Accountability Act (HIPAA) mandates proper logging of access and change histories for EHR [16]. However, this is still a "black box" for e-autopsy reports. Thus, establishing accountability mechanism for forensic data is necessary to ensure the trustworthiness of autopsy reports.

To the best of our knowledge, there are few published works exploring the potential of data provenance with authority management for accountability of forensic data. In addition, the data accountability mechanism can provide some insights for addressing many big data challenges related to data quality and privacy.

3. Problem Statement

In this section, we identify three forensic autopsy goals and associated computational tasks in the processing flow of e-autopsy reports.

In our prior research [1], a general workflow for MEs and DRs to perform collaborative autopsy was identified, in which the use of the e-autopsy report is imperative. As delineated in Figure 1, for performing PA or VA (A1 or A2, respectively), autopsy reports generated from MEs' or DRs' work are integrated into a decision report that contains phased conclusions for the step-by-step refinement of autopsy results. For repetitive and detailed cross-referencing (a structure of A1 with B1 and

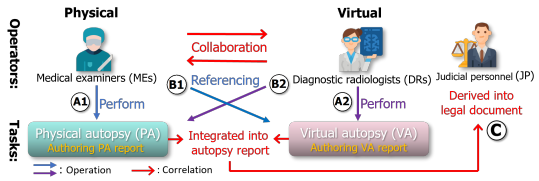

Figure 1: Operators and tasks in forensic activities.

A2 with B2), DRs' (MEs') autopsy reports are viewed as references for MEs (DRs) work. In these processes, back-and-forth reviewing, verifying, and sharing of forensic data are accompanied by the routine work of MEs and DRs. After processing in a forensic hospital, the final autopsy report is transmitted to JP (C), who extracts parts and modifies the form of the information for use in legal document generation and trials. The correctness of the practical workflow relies on the individual correctness of all stakeholders (MEs, DRs, and JP). However, involved stakeholders may act fallaciously in their own interest or make inaccurate decisions according to their oversights.

We first identify three forensic goals (**G**) in terms of authoring and reviewing that describe the target problems. Then, we explicitly state three computational tasks (**T**) of provenance management that our functional design should address.

G1—Accountability and interoperability. The autopsy report is co-authored by multiple doctors (MEs and DRs), so each piece of diagnostic information should have a descriptive and trusted interpretation to allow for shared use.

G2—Reproductivity and traceability. The forensic report must be able to be distributed, reused, and retraced by MEs, DRs, and JP.

G3—Privacy security. There must be a concern for postmortem privacy in authoring autopsy reports using a computational environment.

For addressing these forensic goals, the following three computational tasks can be identified.

T1—Provenance information. The task enables users to reason about, verify and refer to the results; share and reuse the knowledge; and assess data quality and validity.

T2—Lifecycle management. It is essential to facilitate efficient reuse of e-autopsy reports among stakeholders (MEs, DRs, and JP) by intelligently deriving the version, content, format, authoring manner, and viewing manner of autopsy reports based on the stakeholders' duties.

T3—Authority management. The access control system containing tailored workflows and computational tools should be designed for MEs, DRs, and JP to author and reuse the e-autopsy reports.

Note that each of the computational tasks is specified by multiple forensic goals. These forensic goals and computational tasks can provide guidance for the design of a provenance management in FORSETI.

4. Provenance-aware FORSETI

In this section, we give an overview of the provenance management in FORSETI, with a focus on its two core characteristics: the combination of FAWfMS and `lmmlgit` and authority management.

The provenance-aware FORSETI system supports the processing flows of the e-autopsy report in PA, VA, and e-court, enabling the computational tasks outlined in section 3. Figure 2 (a) illustrates the overall picture of the provenance-aware FORSETI, where the input (A), manipulations (B, C, D), output (E), and data model of x-LMML (F1) are existing parts in the current version of FORSETI, while provenance (F2) is the primary component of this work. Fortunately, the original syntax of x-LMML in FORSETI has been well designed, facilitating the incorporation of data provenance functionalities. As shown in F2, a three-dimensional coordinates system is introduced to provide the underlying framework for the lifecycle management of e-autopsy reports in terms of "Time evaluation," "Repository," and "Computational forensic ontology." On the "Time evaluation" axis, each node represents a version of x-LMML files for a different stakeholder, such as DRs, MEs, judges, jury, or the prosecution. These x-LMML files are gradually being refined with stakeholders' processing, achieving the global transition from e-autopsy reports to e-court documents. On the "Repository" axis, each node indicates an x-LMML file storing a forensic autopsy case. Note that the third axis, "Computational forensics ontology" serves as the theoretical basis for support, organization, maintenance, specification, and extension of x-LMML files.

As shown in Figure 2 (b), provenance functionalities in FORSETI consists of three parts: collection (**T1**, **T2**, **T3**), management (**T2**, **T3**), and analysis (**T2**, **T3**). In collection, the navigation interface and the FOSETI system capture mechanisms collect provenance data in x-LMML files at different granularities, such as activity duration, descriptive insights, and expertise explanation. To manage the collected provenance data, a version control system is tailored for the lifecycle management of e-autopsy reports. In analysis, by comparing the related x-LMML files, users can quickly view the differences among the autopsy results, and then utilize the process provenance of these results to make a consensus. In the intersection of the three circles in Figure 2 (b), the core components of three parts are positioned: x-LMML, FAWfMS and `lmmlgit`, and authority management. As shown in the bottom of Figure 2 (b), FAWfMS is defined under a hierarchical structure of workflow management.

4.1. Combination of FAWfMS and lmmlgit

In our design, FAWfMS and `lmmlgit` (**T1**, **T2**, **T3**) inherit the advantages of workflow- and script-based provenance

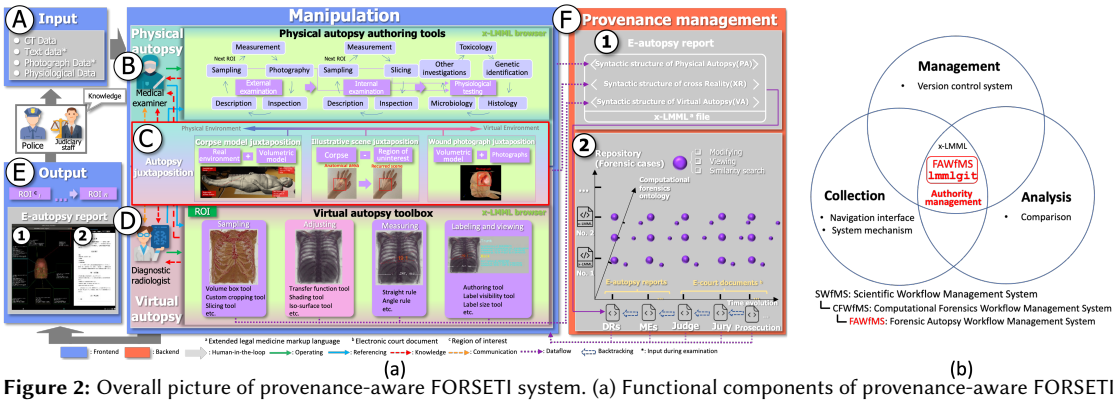

Figure 2: Overall picture of provenance-aware FORSETI system. (a) Functional components of provenance-aware FORSETI system. (b) Abstract structure of provenance functions in FORSETI.

management approaches, respectively. These advantages play an important role in the three components of the FORSETI system. In the following, we explain the roles FAWfMS and `lmmlgit` play in the provenance functionalities and how they are incorporated.

From the lower-left to the top-right corner of Figure 3, the three user interfaces of FAWfMS are shown, navigation interface, node editor interface, and FORSETI interface. The navigation interface is for MEs and DRs to register the person information, clarify the status of the autopsy process, and navigate to their next task. As shown in the left of Figure 3, three branches (PA, Juxtaposition, and VA) with circled numbers are in place. Each circled number represents a set of x-LMML files with their major version number. Users (MEs, DRs, and JP) can click each circle to invoke the node editor interface, where each node graph links an x-LMML file for a particular author or browser, as shown in the middle of Figure 3. The node is not only able to perform some basic operations, such as move, add, delete, and modify, but also has some special features, such as comparison analysis. The node editor can reveal the pedigree of the x-LMML files and their status, such as derived, merged, locked, in-process, and out-process. By double-clicking on the selected node, the user can access the FORSETI interface for authoring and browsing x-LMML files, as shown in the top-right corner of Figure 3.

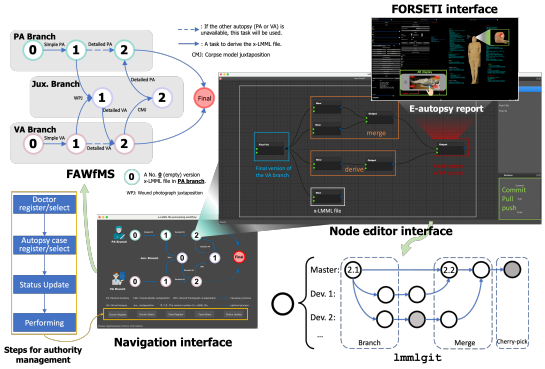

Figure 3: FAWfMS, `lmmlgit`, and authority management in provenance-aware FORSETI system.

The `lmmlgit` is a version control system (VCS) mainly inspired by Git [17], and acts as an expert in processing granularity information, privacy security, and data accountability. All the functionalities in FAWfMS can be carried out by `lmmlgit` commands, but not vice versa. In particular, if a user needs to view a specific target stored in an x-LMML file, it is difficult to use FAWfMS due to a coarser granularity, but `lmmlgit` can be used to access, delete, edit, and check all the targets of an x-LMML file by simply typing designated commands into the shell interface. Thus, `lmmlgit` works as the back-end of FAWfMS for finer handling of process provenance documented in x-LMML files due to its flexibility.

4.2. Authority Management

To build an effective provenance-aware FORSETI system, an authority management (**T1**, **T2**, **T3**) is proposed for e-autopsy confidentiality, in which three works were involved.

The first is strict workflow designs for various stakeholders. As shown in the lower-left corner of Figure 3, four steps for monitoring users' processing are presented in the navigation interface. Through these four steps, the users' personal information, including ID, e-mail, location, affiliation, and position, is stored and verified for assigning access rights. Then, the node editor allows users to author or browse the e-autopsy reports based on the user's level of access. The second is the locking tool installed in the node editor for giving the user control over their own node. Other users can view e-autopsy reports only after obtaining permission from authors or administrators. The third component is particularly important: a well-designed access control syntax supports the first two tasks on the back-end. In the future development plan, the syntactic structure for JP is going to be installed in the authority management to allow the e-autopsy report to be transformed into an e-court document.

5. Concluding Notes

This paper is an initial report on the provenance-aware FORSETI, with an aim to empower MEs, DRs, and JP to accountably author and review e-autopsy reports using a combination of FAWfMS, `lmmlgit`, and authority management.

In the future, incorporation of the provenance functionalities into the autopsy juxtaposition methods, as shown in Figure 2 (a) C, should be a priority. Since the autopsy juxtaposition methods act as a "bridge" for cross-referencing between MEs and DRs, integrating provenance functionalities with these methods will enable a more effective manner of referencing. Indeed, the provenance-supported corpse model juxtaposition allows MEs to understand the insight processes of VA findings in the augmented reality setting, which leads to trustworthy planning of PA. Similar effects can occur in wound photograph juxtaposition and illustrative scene juxtaposition. The next issue is to evaluate the provenance-aware FORSETI with domain experts using reliable and real datasets, which can empirically prove the effectiveness of our system and provide useful feedback for further improvements. The final issue to be confronted is to complete the syntax of access control for authority management for JP.

Acknowledgments

This work has been supported in part by JSPS KAKENHI under the Grants-in-Aid for Scientific Research (A) No. 26240015, 17H00737, and 21H04916.

References

[1] B. Wang, Y. Asayama, M. O. Boussejra, H. Shojo, N. Adachi, I. Fujishiro, FORSETI: A visual analysis environment for authoring autopsy reports in extended legal medicine mark-up language, The Visual Computer 37 (2021) 2951–2963.

[2] Y. Asayama, B. Wang, M. Nakayama, H. Shohjoh, N. Adachi, Y. Kiyoki, I. Fujishiro, THEMIS: Context-sensitive similarity analysis for wound imagery using mathematical model of meaning, in: Proceedings of the 2021 International Conference on Cyberworlds, IEEE, 2021, pp. 129–132.

[3] C. Lundström, T. Rydell, C. Forsell, A. Persson, A. Ynnerman, Multi-touch table system for medical visualization: Application to orthopedic surgery planning, IEEE Transactions on Visualization and Computer Graphics 17 (2011) 1775–1784.

[4] M. O. Boussejra, N. Adachi, H. Shojo, R. Takahashi, I. Fujishiro, LMML: Initial developments of an integrated environment for forensic data visualization, in: Proceedings of the 2016 EuroVis Short Papers, 2016, pp. 31–35.

[5] Y. L. Simmhan, B. Plale, D. Gannon, A survey of data provenance in e-science, SIGMOD Record 34 (2005) 31–36.

[6] J. Cheney, A. Ahmed, U. A. Acar, Provenance as dependency analysis, Mathematical Structures in Computer Science 21 (2011) 1301–1337.

[7] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, H. T. Vo, Vistrails: Visualization meets data management, in: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, 2006, pp. 745–747.

[8] L. M. Gadelha Jr, B. Clifford, M. Mattoso, M. Wilde, I. Foster, Provenance management in Swift, Future Generation Computer Systems 27 (2011) 775–780.

[9] I. Altintas, C. Berkley, E. Jaeger, M. Jones, B. Ludascher, S. Mock, Kepler: An extensible system for design and execution of scientific workflows, in: Proceedings of the 16th International Conference on Scientific and Statistical Database Management, IEEE, 2004, pp. 423–424.

[10] D. Hull, K. Wolstencroft, R. Stevens, C. Goble, M. R. Pocock, P. Li, T. Oinn, Taverna: A tool for building and running workflows of services, Nucleic Acids Research 34 (2006) W729–W732.

[11] K.-K. Muniswamy-Reddy, D. A. Holland, U. Braun, M. Seltzer, Provenance-aware storage systems, in: Proceedings of the Usenix Annual Technical Conference, 2006, pp. 43–56.

[12] J. Frew, P. Slaughter, Es3: A demonstration of transparent provenance for scientific computation, in: Proceedings of the International Provenance and Annotation Workshop, Springer, 2008, pp. 200–207.

[13] L. Murta, V. Braganholo, F. Chirigati, D. Koop, J. Freire, noWorkflow: Capturing and analyzing provenance of scripts, in: Proceedings of the International Provenance and Annotation Workshop, Springer, 2014, pp. 71–83.

[14] J.-L. R. Stevens, M. Elver, J. A. Bednar, An automated and reproducible workflow for running and analyzing neural simulations using Lancet and IPython Notebook, Frontiers in Neuroinformatics 7 (2013) 44.

[15] O. Can, D. Yilmazer, Improving privacy in health care with an ontology-based provenance management system, Expert Systems 37 (2020) e12427.

[16] R. Nosowsky, T. J. Giordano, The health insurance portability and accountability act of 1996 privacy rule: Implications for clinical research, Annual Review of Medicine 57 (2006) 575–590.

[17] J. Loeliger, M. McCullough, Version Control with Git: Powerful tools and techniques for collaborative software development, O'Reilly Media, Inc., 2012.