A workflow learning model to improve geovisual analytics utility

Robert E Roth[1] | Alan M MacEachren[1]| Craig A McCabe[1]
reroth@psu.edu | maceachren@psu.edu | cmccabe@psu.edu

[1]GeoVISTA Center, Department of Geography
The Pennsylvania State University
302 Walker Building, University Park, PA 16802

ABSTRACT

Introduction: This paper describes the design and implementation of the G-EX Portal Learn Module, a web-based, geocollaborative application for organizing and distributing digital learning artifacts. G-EX falls into the broader context of geovisual analytics, a new research area with the goal of supporting visually-mediated reasoning about large, multivariate, spatiotemporal information. Because this information is unprecedented in amount and complexity, GIScientists are tasked with the development of new tools and techniques to make sense of it. Our research addresses the challenge of implementing these geovisual analytics tools and techniques in a useful manner.

Objectives: The objective of this paper is to develop and implement a method for improving the utility of geovisual analytics software. The success of software is measured by its *usability* (i.e., how easy the software is to use?) and *utility* (i.e., how useful the software is). The usability and utility of software can be improved by refining the software, increasing user knowledge about the software, or both. It is difficult to achieve transparent usability (i.e., software that is immediately usable without training) of geovisual analytics software because of the inherent complexity of the included tools and techniques. In these situations, improving user knowledge about the software through the provision of learning artifacts is as important, if not more so, than iterative refinement of the software itself. Therefore, our approach to improving utility is focused on educating the user.

Methodology: The research reported here was completed in two steps. First, we developed a model for learning about geovisual analytics software. Many existing digital learning models assist only with use of the software to complete a specific task and provide limited assistance with its actual application. To move beyond task-oriented learning about software use, we propose a process-oriented approach to learning based on the concept of scientific workflows.
Second, we implemented an interface in the G-EX Portal Learn Module to demonstrate the workflow learning model. The workflow interface allows users to drag learning artifacts uploaded to the G-EX Portal onto a central whiteboard and then annotate the workflow using text and drawing tools. Once completed, users can visit the assembled workflow to get an idea of the kind, number, and scale of analysis steps, view individual learning artifacts associated with each node in the workflow, and ask questions about the overall workflow or individual learning artifacts through the associated forums. An example learning workflow in the domain of epidemiology is provided to demonstrate the effectiveness of the approach.

Results/Conclusions: In the context of geovisual analytics, GIScientists are not only responsible for developing software to facilitate visually-mediated reasoning about large and complex spatiotemporal information, but also for ensuring that this software works. The workflow learning model discussed in this paper and demonstrated in the G-EX Portal Learn Module is one approach to improving the utility of geovisual analytics software. While development of the G-EX Portal Learn Module is ongoing, we expect to release the G-EX Portal Learn Module by Summer 2009.

Keywords: geovisual analytics, workflows, learning, utility, usability, geocollaboration, G-EX Portal, epidemiology

1. Introduction

Here we report on an effort to improve the utility of geovisual analytics tools and techniques using a process-oriented, workflow learning model. In a 2005 manuscript, the National Visualization and Analytics Center (NVAC) introduced a new research area of visual analytics with the goal of supporting sophisticated analytical reasoning through the integration of the human eye-brain system and visual computer interfaces to computational methods (Thomas and Cook, 2005). The GIScience community responded to this call by organizing a new research thrust in *geovisual analytics* to specifically address the spatial component of visually-mediated reasoning (Andrienko et al., 2007). Geovisual analytics separates itself from previous work in cartography and geovisualization in both a focus on supporting the process of analytical reasoning and increasing the amount and complexity of the information that can be analyzed.

As the amount and complexity of information requiring analysis grows, so too does the difficulty of learning and applying geovisual analytics software. Software success is often defined by its *usability* (i.e., a measure of the ease-of-use of the software, called *interface transparency* when it is immediately usable without any training) and *utility* (i.e., a measure of the usefulness of the software when applied to a specific problem or task) (Grinstein et al., 2003, Fuhrmann et al., 2005). Usability and utility often play out as competing forces during software design (e.g., Apple decided to remove features from the second generation iPod to make the interface transparently usable). Both usability and utility can be improved by refining the software, increasing user knowledge about the software, or both. However, in situations requiring a high level of software utility – which is typically the case with geovisual analytics – a greater emphasis must be placed on strategies for increasing user knowledge through the provision of learning artifacts.

In this paper, we describe the **G**eovisual **Ex**plication (G-EX) Portal Learn Module, a web-based application for organizing and distributing cartographic, ESDA, and geovisual analytics learning artifacts. A primary purpose of the G-EX Learn Module is to improve user knowledge about geovisual analytics software (and therefore the usability and utility of the software) when transparent usability cannot be achieved. Our focus in this paper is upon the design and implementation of an interactive workflow interface for the improvement of software utility. In the following section, existing task-oriented learning models for both usability and utility are summarized and the process-oriented workflow model is introduced. In the third section, the G-

EX Portal is introduced and the design and implementation of the workflow-based Learn Module is described. We conclude with a brief outline of future work and summary remarks.

2. Task-oriented versus process-oriented learning & the workflow learning model

2.1. Existing learning models: Task-oriented learning

Software learning artifacts traditionally fall into one of two categories: *software use* (i.e., which buttons to push and in what order, generally relating to software usability) or *software application* (i.e., how to utilize the software in a problem context or knowledge domain, generally relating to software utility). Learning assistance for software use is typically provided through external materials, such as code documentation and instruction manuals, and internal features, such as wizards and popup tool tips. These learning models are designed to walk a user through the required mouse clicks or keyboard inputs needed to use a single feature in the software and have only limited potential to support learning how to apply the software.

The classic model for learning about the application of geovisual analytics software is classroom education, where a set of students would meet in person with a single instructor to discuss how tools and techniques can be applied to solve a specific problem in a given domain. However, the advent and pervasive use of the Internet as a mechanism for distribution and collaboration has led to a growth in distance education or e-Learning (DiBiase, 2000); most higher education courses now offer at least a blended curriculum with classroom and online components. In distance education, learning artifacts about the application of software include textbook–like resources on domain concepts, forums/wikis for group discussion, and multimedia tutorials or demonstrations.

The aforementioned, existing learning models typically assist in completing only one task at a time, an approach to learning described by Lin et al. (2002) as ‘task-oriented’. The task-oriented approach to learning may be insufficient for geovisual analytics, as the emphasis is on logically reasoning through a series of steps to arrive at insights, hypotheses, or decisions. Here, the connections across tasks (i.e., how the user determines to move from one task to the next) are as important as the tasks themselves. To move beyond task-based learning, Lin et al. (2002) recommend a ‘process-oriented’ approach to learning based on the concept of workflows.

2.2 The workflow learning model: Process-oriented learning

A workflow is the formal representation of a work process (van der Aalst, 1998). Rather than restricting the work process to a linear sequence of activities, the workflow model allows for a process conditioned by the outcomes of previous activities and/or the context of the process; the correct pathway through the workflow emerges as each step in the process is completed (Stohr and Zhao, 2001). Because of this conditional characteristic, workflows are commonly represented as flow charts or tree-diagrams. In this paper, we are interested in scientific workflows, or the “analytical pipeline” from data to knowledge (Ludäscher et al., 2006, p1041). Given the context of geovisual analytics, we are also interested in how this knowledge is then used to make a decision or adjust a policy. Nodes in the workflow are the various scientific methods that can be applied to support scientific reasoning throughout the process, with the conditional connections among nodes based upon both the results and their interpretation.

Multiple scholars have argued for the application of the scientific workflow concept to distance education (e.g., Stohr and Zhao, 2001, Lin et al., 2002, Sumner et al., 2005). In a distance education workflow model, a digital learning artifact is prepared for each possible step in the reasoning process; each individual learning artifact makes up a single node in the workflow. When viewed independently of the workflow, these learning artifacts primarily provide instruction on software use; together, they support learning about a process of work using the software.

To move beyond education on software use and towards education on software application, the learning artifacts must support learning related to one or several of the following: (1) background cartography and geovisual analytics theory describing why a tool or technique was implemented in the software as it was, (2) an explanation of *when* or *why* the tool or technique should be applied (in addition to *how* to apply it), (3) specifics about how to adapt the tool or technique to a particular application domain, (4) guidance on how to interpret the results of the tool or technique within a broad process of analytic reasoning, and (5) recommendations on next steps or future directions depending on the results. The workflow learning model, as implemented in the G-EX Portal Learn Module, makes learning focused on these topics possible through three approaches:

(1) Within each workflow node: All five of the above topics designed to improve software application can be incorporated directly into each individual learning artifact. This is the only way that existing learning models support software application. However, including instructions on both software use and software application in a single learning artifact may make the document prohibitively lengthy and overly complex. Because of this, it is rare to cover more than one (if any) of the above educational topics on software application.

(2) Via the workflow connections: A primary advantage of a workflow learning model is that it makes obvious the connections among steps in the analytic process. These graphic connections can serve two purposes: they can connect a step in the analytic process to instructions on how to interpret the results of the step (#4 above or the *if* part of an *if-then* logic statement) and, based on this interpretation, suggest the next course of action in the analytic process (#5 above or the *then* part of an *if-then* logic statement).

(3) Via ongoing communication: An important aspect of the classroom model of learning is the ability for students to ask for clarification about concepts they do not fully understand. The primary way such Q&A is provided in distance education is through a forum or wiki. Once the workflow is implemented in a digital environment, a wiki can be attached to the overall workflow itself, to each of its nodes, and to each of its connections. This allows users to ask extra questions about background theory in cartography and geovisual analytics (#1 above), conditions for application of a specific geovisual analytics tools or techniques (#2 above), and background theory in a potential application domain (#3 above).

3. G-EX Learn Module Implementation

3.1 Introduction to the G-EX Portal

The Geo-EXplication (G-EX) Web Portal is a proof-of-concept, web-based, geocollaboration application currently under development at the Penn State Geographic Visualization Science, Technology, and Applications (GeoVISTA) Center. It is designed to facilitate dissemination of and collaboration about cartographic, ESDA, and geovisual analytics tools, learning artifacts for these tools and associated techniques, and analysis artifacts (datasets, maps, statistics, notes, or any other record of the geovisual analytics process) generated during the application of these tools and techniques (Robinson et al., 2007, Roth et al., 2008). Design of the G-EX Portal was inspired by online, profile-based social network software like del.icio.us, Facebook, and YouTube. Users contribute to the online repository by uploading tools, learning artifacts, or analysis artifacts to their personal profile. They are then able to view and download the artifacts contributed by others and collaborate with these users on larger projects that may involve many uploaded tools, learning artifacts, and analysis artifacts.

The G-EX Portal is organized into four modules, each supporting a different high-level geovisual analytics task: (1) Search, (2) Collaborate, (3) Review, and (4) Learn. The *Search Module* provides the primary portal interface, allowing a user to browse, filter, and retrieve artifacts in the repository that may be relevant to the user's current geovisual analytics project. The *Collaborate Module* provides the primary outlet for asynchronous geocollaboration, allowing users to share, discuss, and annotate artifacts related to their long-term cartographic, ESDA, and geovisual analytic applications. The *Review Module* provides a multimedia alternative to the traditionally text-based review process, allowing reviewers to evaluate all steps and products of the scientific process more thoroughly. Further background theory and scenarios of use for the Search, Collaborate, and Review Modules are available in Robinson et al. (2007). For this research, we are only interested in the design and implementation of the Learn Module, described in greater detail in the following section.

3.2 The G-EX Portal Learn Module

The G-EX Portal *Learn Module* provides software developers with an outlet for public dissemination of learning artifacts about their cartographic, ESDA, and geovisual analytics software. More importantly, the G-EX Portal Learn Module provides a free, online one-stop site for browsing and retrieving learning artifacts about the use and application of software as well as providing a forum for contacting the developers with detailed, follow-up questions. The G-EX Portal Learn Module can handle any file format that is supported by the user's browser, allowing for heterogeneous learning content. The Module currently contains multimedia learning artifacts for several of the GeoVISTA products, including the Exploratory Spatio-Temporal Analysis Toolkit (E-STAT), the GeoViz Toolkit, GeoVISTA *Studio*, the Health GeoJunction, the Pennsylvania Cancer Atlas, and the G-EX Portal itself.

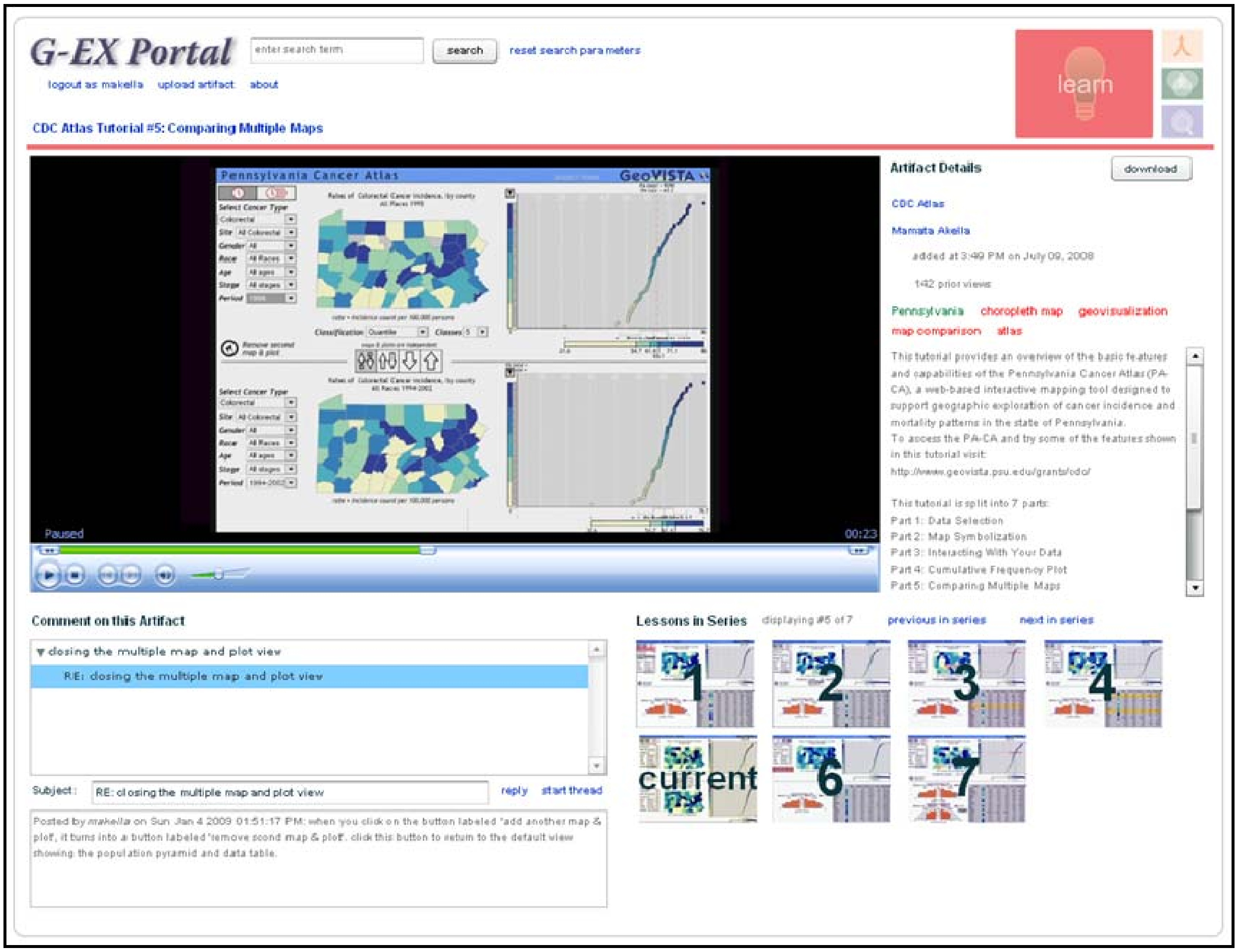

Figure 1. An individual learning artifact page in the G-EX Portal Learn Module.

After a user uploads a learning artifact to the repository, an individual page for the artifact is created. This individual learning artifact page can be reached using the persistent Search functionality at the top of the application or from the contributor's homepage. Figure 1 shows an individual learning artifact page for a movie clip tutorial of the Pennsylvania Cancer Atlas (see Bhowmick et al., 2008, for details). Each individual learning page has four primary interactive components: (1) a viewer for the learning artifact itself at the top-left of the interface, (2) a description and other metadata (the associated project or software, contributor, date added, number of views, attribute and geography tags, etc.) about the learning artifact at the top-right, (3) a wiki allowing users to ask questions about the learning artifact at the bottom-left, and (4) a listing with a graphic preview of other learning artifacts associated with the tool or technique at the bottom-right.

3.3 Extension of the Learn Module to support the workflow learning model

The individual learning artifact pages follow a task-oriented approach to learning about the use of a tool or technique. To support process-oriented learning about the application of software, we are currently extending the G-EX Learn Module to provide a secondary interface for the logical organization of a set of learning artifacts into a workflow. Figure 2 shows a mockup of the work-

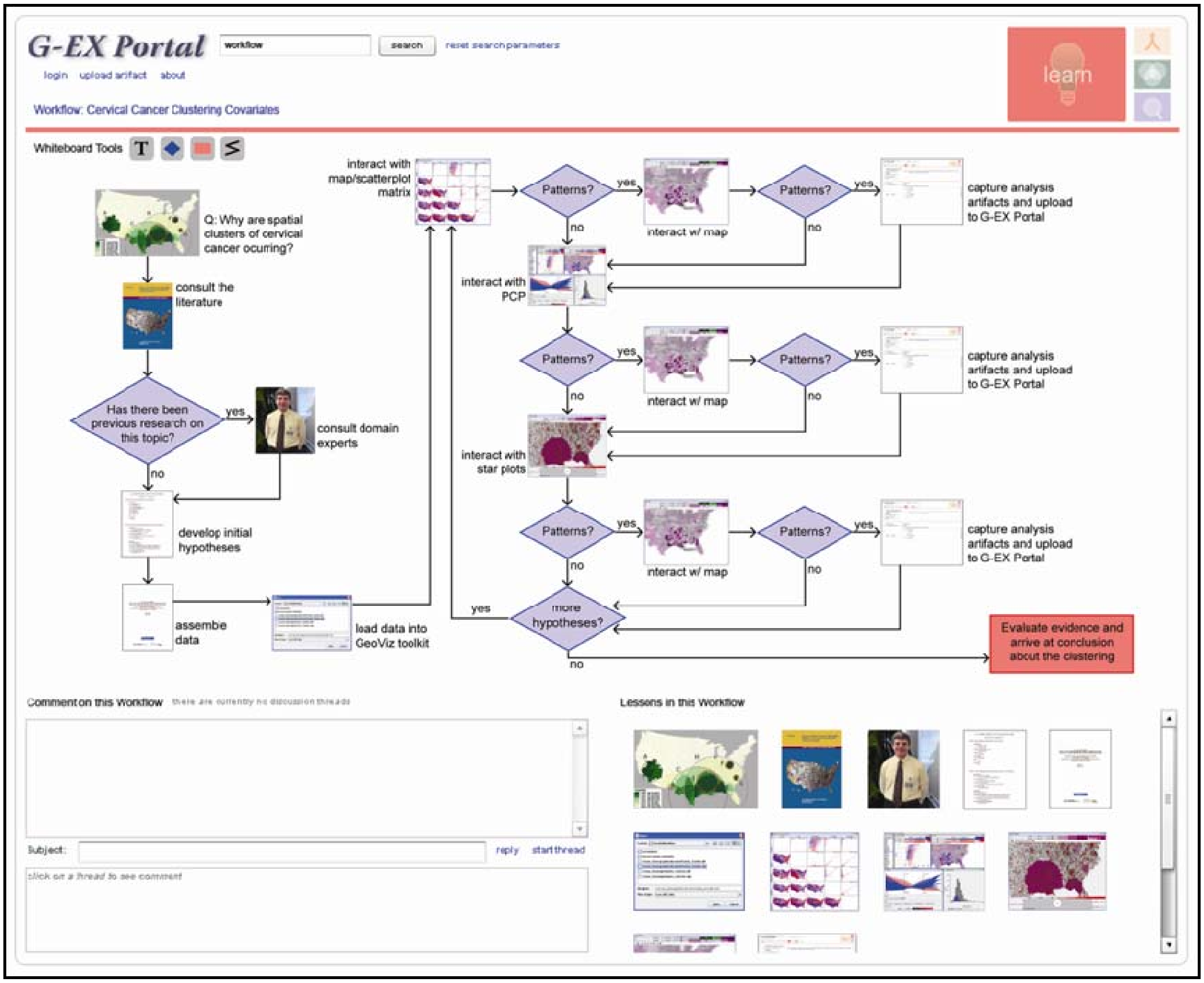

Figure 2. A mockup of the workflow interface with an example in the domain of epidemiology.

flow interface using a workflow learning example in the domain of epidemiology. The workflow interface has three primary interactive components: (1) a central whiteboard for constructing and annotating the workflows, (2) a wiki allowing users to ask questions about the overall workflow and its connections at the bottom-left, and (3) a listing of all individual learning artifacts comprising the workflow at the bottom-right.

Creation of a workflow begins by dragging a learning artifact icon from the bottom-right list interface onto the whiteboard. The user can then reposition these icons according to their logical place in the workflow. The user can annotate the workflow using one of the four drawing tools: (1) a line of text, (2-3) one of two colored shapes (rectangle representing a processing step without learning artifacts and diamond representing a conditional step or decision) and (4) a line tool. The example shown in Figure 2 illustrates a workflow of recent research at GeoVISTA to apply a geovisual analytics approach to enhance established spatial clustering methods applied in cancer surveillance research (see Chen et al., 2008, for details). The specific focus of the workflow is on learning how to investigate factors that may explain why clusters of elevated cervical cancer incidence are occurring in particular regions of the United States. The steps in the learning workflow portray the analytic process that an epidemiologist would follow while exploring this problem, including consultation of literature and domain experts, development of possible explanations or hypotheses, assembly of the requisite spatiotemporal datasets, exploration using geovisual analytics software (here, the GeoViz Toolkit is used – see Hardisty and Robinson, submitted, for details), capture of analysis artifacts using the G-EX Portal, and final interpretation and evaluation of the collected evidence.

Upon first entry to the workflow page, the user receives a general impression about the kind, number, and scale of analysis steps that must be undertaken to solve a similar geovisual analytics problem; it is expected in many cases that the user will be working on his or her own research problem while proceeding through the learning workflow. The user begins interaction with the workflow by clicking on the icons. Each click sends the user to the associated individual learning artifact page for instruction. Where possible, we have incorporated information specific to the domain application of epidemiology into our explanations of the datasets and geovisual analytics tools and techniques (providing learning support for the application of the software through approach #1 from Section 2.2). Once a learning artifact has been reviewed by the user, he or she returns to the workflow interface to determine the next appropriate step (approach #2 from Section 2.2), which may be conditioned by the outcomes of previous learning artifact steps or the context of the workflow application. At any step in the workflow, a user can ask a more detailed question about both an individual learning artifact (on the individual learning artifact page) or about the workflow itself (on the workflow page) (approach #3 from Section 2.3). Progress through the workflow continues until the user has enough expertise to complete a step without guidance or when the user reaches the final step in the workflow.

4. Future Directions and Conclusions

Development of the G-EX Portal, including the Learn Module described above, is an ongoing project at the GeoVISTA Center. At the time of this submission, the G-EX Portal is available for internal or invited use only. In addition to making the interface more robust and opening access to remote users (anticipated completion by Summer 2009), future directions include developing additional controls for the whiteboard interface (e.g., grouping icons or actions, collapsing/expanding these groups, resizing/recoloring the whiteboard drawings), extending the whiteboard interface concept to other modules in the G-EX Portal, enhancing the communication beyond a simple wiki (e.g., RSS feeds for wiki updates, commenting ability on the workflow connections, annotation directly atop the workflow), and improving learning assistance for the application of geovisual analytics software. Further, we are currently revising the design of learning artifacts and their dissemination in the G-EX Portal Learn Module based upon the results of a formal end user survey.

In order to support visually-mediated analytical reasoning about large and complex spatiotemporal datasets, visual interfaces to computation methods must be made available to analysts. To this end, GIScientists are tasked with both the development of this geovisual analytics software and, of equal importance, ensuring that this software is both easy-to-use and useful. In this paper, we contend that the usability and utility can be improved by refining the software, increasing user knowledge about the software, or both. Further, we argue that in situations where the goal of transparent usability must be abandoned to support progress on complex problems requiring sophisticated methods and tools, a greater emphasis should be placed on increasing user knowledge through the provision of learning artifacts. The offered workflow learning model is one possible framework for moving away from task-oriented learning primarily focused on using the software and towards process-oriented learning primarily focused on answering questions and solving problems with the help of the software. We encourage both evaluations of and revisions to this workflow learning model and suggestions for possible alternatives that hold the similar objective of providing process-oriented learning assistance for the application of geovisual analytics software.

References

ANDRIENKO, G., ANDRIENKO, N., JANKOWSKI, P., KEIM, D., KRAAK, M.-J., MACEACHREN, A. & WROBEL, S. (2007) Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science,* 21**,** 839-857.

BHOWMICK, T., ROBINSON, A. C., GRUVER, A., MACEACHREN, A. M. & LENGERICH, E. (2008) Distributed usability evaluation of the Pennsylvania Cancer Atlas. *International Journal of Health Geographics*, 7.

CHEN, J., ROTH, R. E., NAITO, A. T., LENGERICH, E. J. & MACEACHREN, A. M. (2008) Enhancing spatial scan statistic interpretation with geovisual analytics: An analysis of US cervical cancer mortality. *International Journal of Health Geographics,* 7**,** 57.

DIBIASE, D. (2000) Is distance education a Faustian bargain? *Journal of Geography in Higher Education,* 24**,** 130-135.

FUHRMANN, S., AHONEN-RAINIO, P., EDSALL, R. M., FABRIKANT, S. I., KOUA, E. L., TOBON, C., WARE, C. & WILSON, S. (2005) Making useful and useable geovisualization: Design and evaluation Issues. IN DYKES, J., MACEACHREN, A. M. & KRAAK, M. J. (Eds.) *Exploring Geovisualization.* Oxford, United Kingdom, Elsevier Ltd.

GRINSTEIN, G., KOBSA, A., PLAISANT, C., SHNEIDERMAN, B. & STASKO, J. T. (2003) Which comes first, usability or utility? *14th IEEE Visualization (VIZ '03).* Seattle, WA, IEEE Computer Society.

HARDISTY, F. & ROBINSON, A. C. (submitted) The GeoViz Toolkit: Using component-oriented coordination methods to aid geovisualization application construction. *International Journal of Geographical Information Science*.

LIN, J., HO, C., SADIQ, W. & ORLOWSKA, M. E. (2002) Using workflow technology to manage flexible e-learning services. *Education Technology & Society,* 5**,** 116-123.

LUDÄSCHER, B., ALTINAS, I., BERKLEY, C., HIGGINS, D., JAEGER, E., JONES, M., LEE, E. A., TAO, J. & ZHAO, Y. (2006) Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience,* 18**,** 1039-1065.

ROBINSON, A. C., KOUA, E., HARDISTY, F., CHEN, J. & MACEACHREN, A. M. (2007) The G-EX Portal: Web-based dissemination of geovisual analytic results. *ICA Commission on Visualization and Virtual Environments Workshop 'From Geovisualization towards Geovisual Analytics'.* Helsinki, Finland.

ROTH, R. E., ROBINSON, A., STRYKER, M., MACEACHREN, A. M., LENGERICH, E. J. & KOUA, E. (2008) Web-based geovisualization and geocollaboration: Applications to public health. *2008 Joint Statistical Meeting, Invited Session on Web Mapping.* Denver, CO, ASA.

STOHR, E. A. & ZHAO, J. L. (2001) Workflow automation: Overview and research issues. *Information Systems Frontiers,* 3**,** 281-296.

SUMNER, T., AHMAD, F., BHUSHAN, S., GU, Q., MOLINA, F., WILLARD, S., WRIGHT, M., DAVIS, L. & JANEE, G. (2005) Linking learning goals and educational resources through interactive concept map visualizations. *International Journal of Digital Libraries,* 5**,** 18-24.

THOMAS, J. J. & COOK, K. A. (2005) *Illuminating the path: The research and development agenda for visual analytics,* Los Alametos, CA, IEEE CS Press

VAN DER AALST, W. M. P. V. D. (1998) The application of Petri nets to workflow management. *Journal of Circuit Systems and Computers,* 8**,** 21-66.