A Workflow Analysis Perspective to Scholarly Research Tasks

Marijn Koolen*
marijn.koolen@di.huc.knaw.nl
Humanities Cluster
Royal Netherlands Academy of Arts
and Sciences
Amsterdam, Netherlands

Sanna Kumpulainen*
sanna.kumpulainen@tuni.fi
Faculty of Information Technology
and Communication Sciences
Tampere University
Tampere, Finland

Liliana Melgar-Estrada*
lilianamelgar@runbox.com
Utrecht University
Netherlands Institute for Sound and
Vision
Netherlands

ABSTRACT

Since the appearance of digital research infrastructures in the humanities in the last decade, important efforts are being made to understand and model scholarly processes. Different methods are used in those investigations, which often result in abstract representations of research phases, taxonomies of scholarly activities, in conceptual frameworks, or in scholarly ontologies. While the aim of these representations is to inform the design of the digital infrastructures, the complexity and diversity of scholarly work pose the question about the applicability of those models for design and evaluation of research infrastructures and tools. In this paper, we explore a methodology to analyze workflows from a micro-perspective, which aims at capturing the transitions between activities. We use two scholarly projects as case studies, describe their research activities in detail by using existing ontologies and describe the connections between activities, and analyse generic transitions. We discuss what kinds of implications this approach has to evaluation and design of information systems and services to facilitate scholars' complex and varied research processes.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Interactive systems and tools**; • **Information systems → Information systems applications**; **Information retrieval**.

KEYWORDS

scholarly primitives; research workflows; research information interaction

*Authors contributed equally to the paper.

1 INTRODUCTION

Digital scholars use information systems and expect system support not only during searching, but during all the phases of their research process [18, 20]. While in scientific domains research processes are formalized in workflows for which there is explicit support [43], the varied nature of research practices in digital humanities research escapes those formalizations [3]. However, detecting, documenting, and understanding workflows in the humanities is important since there are significant efforts invested in creating digital research infrastructures [1] that are expected to "enhance and support digitally-enabled research and teaching."[2].

In the context of these digital research infrastructures, where data is offered by different providers, and where several tools are built to facilitate working with that data, a process or workflow perspective to information system design and evaluation is of high importance. Particularly in humanities research, where there is little evidence of paradigms operating [27], and there are few established common or shared workflows, more research is needed to tailor research infrastructures to the ways scholarly practice occurs in real work tasks.

Conceptual models of the scholarly process in the field of information behaviour have focused on seeking, but attention to information use behaviour has been limited [34, 37]. Research about the scholarly processes often generates models that propose a macro perspective, originated from either conceptual work or the empirical analysis of overarching research phases or stages (e.g., the models cited in [11]). One limitation of those models is that information and data flows across phases is not reported on sufficient level of detail. Further, since they are abstract representations, they are useful in supporting conceptual overviews, but are limited in capturing the wide variety of scholarly activities. This is crying out for a better methodology to understand and represent scholarly workflows in a way that they can effectively guide information system design and evaluation for specific tasks and research goals.

The main aim of this paper is to contribute to ongoing research about digital scholarly workflows from an information interaction perspective. Instead of aiming to propose a conceptual model, we propose a methodology to increase the level of detail in the analysis of the activities that actually occur within research projects. This analysis has implications for improving the understanding of how to support and evaluate workflows in research work tasks of humanities scholars.

---

[1] such as the US infrastructure HTRC (HathiTrust Research Center), the European infrastructures DARIAH (Digital Research Infrastructure for the Arts and Humanities), and CLARIN (Common Language Resources and Technology Infrastructure), or the the Dutch CLARIAH infrastructure project.
[2] https://www.dariah.eu/about/dariah-in-nutshell/

We seek to address the following research questions:

RQ1. What are the characteristics of research workflows in the selected cases of digital humanities research projects?

RQ2. Is it possible to identify commonalities in the sequences of, or transitions between, research activities across similar digital humanities research projects?

RQ3. What methods are suitable to investigate and compare (humanities) workflows in natural work settings with real research tasks?

We use two research projects in humanities field as cases, and examine their *activities*, the *transitions* between them and provide full *workflow* representations. This method allows us to show critical points where support can be provided, and to identify elements that can improve the understanding of scholarly workflows.

2 BACKGROUND

2.1 Scholarly activities and primitives

Scholarly tasks are typically described using the concept of *scholarly primitive*. This term was coined by Unsworth [46] to refer to some basic functions common to scholarly activity across disciplines. Unsworth [46] defines and separates one primitive from another by its implied and unique functions. The idea of workflows is envisioned based on the primitives, in which the tools that support a primitive are able to produce outputs and receive inputs in standard forms. Unsworth [46] proposes seven scholarly primitives, namely: *discovering*, *annotating*, *comparing*, *referring*, *sampling*, *illustrating*, and *representing*.

Palmer et al. [40] use the concept of scholarly primitives to generate an "activity-based framework" facilitating comparison across domains. They use the concept of *scholarly information activity* to emphasize the explicit role of information in the research processes. They group twenty fine-grained primitives into five broader scholarly information activities. Anderson et al. [2] connected these activities and functions to research methods in a *methodological commons*, with the aim to link them more concretely to research infrastructures offering generic tooling for digital research on a broad range of data sets. They aimed at constructing a framework which describes the methods and functions of scholarly activity in humanities research. They argue that understanding the commons will help to bridge the divide between disciplines and therefore enabling interdisciplinary work and collaboration.

Hoekstra and Koolen [19] added to the discussion about scholarly acitivities with the concept of *data scopes*, which focuses on the central role of modelling in digital humanities research and the importance of understanding data-oriented research as *sequences of transformation activities* that each change the nature of the data and thereby how they are interpreted.

2.2 Research processes and workflows

In the last decade, the creation of digital research infrastructures in the humanities and social sciences has led to the aim of identifying scholarly tasks and matching with specific tool-sets and interface capabilities [7]. In scientific domains, workflows are used as part of daily experimentation and data intensive work. They are conceived as scripts, which can be executed, reused by other scientists, becoming part of a chain of the development of scientific

processes [4, 17]. Scientific workflows were studied empirically in an ethnographic research setting in molecular medicine by Kumpulainen and Järvelin [32],[33]. The series of studies show that the more complex the task, the more varying types of information tools are used [32], and the more difficulties are encountered [33]. Information searching was manifested in differing ways according to task complexity; in complex tasks there were more exploratory searches in sessions and in simple tasks there were more basic single searches and fact checking [30, 32]. According to these studies, research projects include several types of tasks on various levels of difficulty and have various types of information related activities.

In the humanities, research on scholarly workflows became more prominent since the start of the DARIAH research infrastructure,[3] which has been creating a conceptual model for scholarly research activity that is grounded in conceptual and empirical research. This conceptual model aims to serve as the basis to build digital research infrastructures that fit the needs of the actual research life-cycle [6]. This work resulted in a domain independent scholarly ontology (SO), which, according to Pertsas and Constantopoulos [41], represents the domain of scholarly work in the digital age. One extension of this ontology, applied specifically to the humanities is the NeDiMAH [4] Methods Ontology (NeMO). The NeMO ontology contains a hierarchy of 161 research activities and is inspired by Vygotsky's and Leont'ev's cultural-historical activity theory [15], and expressed in terms of the CIDOC Conceptual Reference Model. Also Antonijevic and Cahoy [3] report on a "Digital Scholarly Workflow" project, conducted at the Pennsylvania State University from 2012 to 2016, which analyzed the workflows of faculty members using surveys and field observations to analyse the activities conducted by scholars during one day. Their study focused on the broader aspects of personal collection building and use of digital tools to manage primary and secondary sources, and concluded that there is an "unscripted approach to technology" in scholarly workflows.

2.3 Evaluation frameworks for task processes

Development of information systems, including tools that support information use, requires an understanding of the task process where the systems are used as tools in achieving the task goals [23].

There are few models that support evaluation of information interactions related to information systems in task performance. Stage models provide frameworks that separate different phases during the task processes (see, e.g., Kuhlthau [29], Vakkari [47]). The studies show, that information interactions are different in different stages. In the initiation stage, there are more exploration and seeking for background information, in the focus formulation stage the searching is getting more focused, and in the post-focus stage the information needs are differentiated and searching is more specific [47]. Huurdeman and Kamps [20] studied the stages based on these findings, from initiation to post-focus, on micro level and showed that utilizing both task-level (macro level) and system-level (micro level) support may yield promising results.

Järvelin et al. [23] provided an evaluation framework aiming at analyzing information-intensive activities during the whole task performance process in a learning task. It provides a set of activities

---

[3]Digital Research Infrastructure for the Arts and Humanities, https://www.dariah.eu/
[4]ESF Research Network for Digital Methods in the Arts and Humanities (NeDiMAH.eu)

(task planning, searching and selecting information items, working with information items, and synthesizing and reporting) and related program theories for evaluation purposes.

Task complexity and difficulty have been popular concepts in evaluating task performance, processes and objectives in information seeking and particularly in interactive information retrieval [49]. There are several different task complexity classifications. The Taxonomy of Learning Objectives was developed by Anderson and Krathwohl [1] for educational objects. Its cognitive process dimension included six types of cognitive processes: remember, understand, apply, analyze, evaluate and create. This framework has been used in information retrieval to construct and study search tasks and related searching behavior [22, 26]. Anderson and Krathwohl [1] stated that higher-order thinking skills involve processes of analyzing, evaluating, and creating. The rest of the cognitive process types, remembering, understanding and applying require lower-order thinking skills. Typically, in more complex tasks, such as problem solving or research tasks, people need different types of information sources [12, 32], their search sessions are longer [32] and they may feel uncertain about their tasks [29]. Therefore, higher-order thinking activities require different support than lower-order activities. The latter are more effortlessly transformed into automated workflow scripts, but the understanding of what kinds of supports are needed in more complex tasks is scattered.

There are various frameworks explaining and emphasizing the importance to study digital humanities scholars' practices and work flows. The themes are intertwined with the discussion about the change in the methods and research paradigm. Palmer and Cragin [39] differentiate varying information practices of humanities scholars. Trace and Karadkar [45] call for support that is sensitive to disciplinary based cultures and practices. It is widely accepted, that models supporting scientific workflows do not necessarily fit the scholarly processes in humanities [6].

Evaluation means different things in different contextual levels (see e.g. [21, 24]). Systems may be evaluated on data and system level, information seeking level, task level or across all these levels. In the context of evaluating workflows and research processes that include various types of information interactions and activities with possibly several systems, a need arises for examining also the transitions between the activities that are otherwise difficult to connect in the larger context.

3 RESEARCH DESIGN

3.1 Use cases and sample selection

We approached the study of scholarly work from a micro-perspective and studied research activities and transitions between them, i.e., the workflows. By the term *workflow* we then mean a sequence of activities. An *activity* means a unique type of human mental or physical action that can be separated from other types [46]. These behavioral and cognitive activities contribute to task performance and task outcome [23]. The NeMo [5] ontology provides 161 research activities, and forms the basis for our analysis. The research activities are mapped into two levels of cognitive complexity according to the taxonomy by Anderson and Krathwohl [1] to gain more explanatory power in the analysis.

Our unit of analysis is a *research project*. In this investigation, instead of focusing on long or large research projects, we chose projects with a focused research question, and of which at least one publication came out. In the present research we used the research projects as our *cases*.

We selected two digital humanities projects, with two participants per project, from the area of socio-economic history that have been selected as one of the priority areas in CLARIAH digital infrastructure project [6].

Participants were Dutch researchers but the interviews were conducted in English because of our international research team. We chose projects which: a) had one, maximum two, research questions, and, b) were finished by the time of our data-collection and were already reported in research papers. The first criteria was based on the idea that we wanted to have projects that we were able to handle without too many branches. The second was included, because we wanted to reconstruct the full research life-cycles from the initiation to the end.

The selected projects were:

- Research Project 1 (RP1): The bi-directional influence of politics on science and of science on politics in the management of migration flows in the second half of the 20th century in Europe. The people, organization and topics involved in the international organization of migration are extracted from roughly one hundred political and scientific publications and analyzed per decade to observe changes in the discourse and the network of actors.
- Research Project 2 (RP2): Human capital, immigration and social mobility, about the reconstruction of the careers of migrant and native workers active in the eighteenth century Dutch East India Company (Vereenigde Oostindische Compagnie, VOC). Thousands of payledgers have been transcribed which together contain almost 800,000 contracts between workers and the VOC between 1680 and 1800 to go on 1-5 year journeys to the Dutch East Indies. These contracts state the rank and pay for each individual, and tens of thousands of individual went on multiple journeys, either staying in the same rank or being promoted or demoted.

We note that these two projects are not representative of the broad field of (digital) humanities, but we selected these projects because: 1) both have been carried out in collaboration between domain-experts and a data experts, and we were able to interview collaborators in each project to get a more complete picture, and 2) these projects are different in terms of data size (a hundred publications versus 800,000 contracts) and in method (one focusing on bringing together information from many different sources, the other in extracting new models and insights from a large but homogeneously structured data set), which allows us to approach methodological issues in comparing similarities and differences in workflows.

3.2 Data collection

Studies of real life are considered to be more representative of actors' information behaviors [25]. Therefore, we aimed at studying real work processes by reconstructing two research projects that

represent full, large research tasks from start to end. We wanted to have already finished projects to be able to cover the whole research process, including publication and presentation activities (cf. [23]), and also to have a diversity of sources where we could find descriptions about how each research project was conducted. We thus collected interview data, demonstration data and various documents related to our cases (including, e.g., notes, presentations and publications).

In order to fully cover the viewpoints and the cognitive activities of the humanities scholars, we conducted semi-structured interviews. Interviews and talk-aloud methods are the best means to capture the participants' own insights and cognitive actions [25, 31].

Interviews focused on experienced tasks and task structure. We designed an interview guide (published on Zenodo [28]) inspired by the activities followed by the task-based information interaction evaluation framework [23]. By using the framework we aimed at covering a wide selection of information intensive activities.

In the interviews, we utilized Critical Incident Technique (CIT) [16]. By using CIT we were able to anchor the interviews to a real life research project, and gather more focused answers. During the interviews we asked the interviewees to focus on a pre-selected research project (their choices being RP1 and RP2) and to demonstrate how they worked with their data and various tools during the project. Further, we followed the approach by [6] and focused on activities as characterized by the interviewees; not only asking what they did and how, but also about the aims and goals of the activity (why it was done?), and how they connected to other contextual aspects, such as collaboration and the subject of activity (who did it?).

To increase the accuracy of observed interactions, we also asked the participants to do demonstrations of specific activities (e.g., how did they extract information from their primary sources). During the demonstrations we were able to observe how they worked with the tools and research materials they were working with during the case projects. Therefore we were able to collect in-depth understanding of humanities scholars' work related information activities and how the scholars collect, organize, and use their research material.

Interviews were conducted at the participants' own work desks. Each of the four interviews (two participants per project) lasted between nineteen minutes to two hours. With two participants (one from each project), we conducted a shorter follow-up interview to cross-check their input. The interviews were audio recorded and we used a video camera to collect the demonstrated information interactions and supplemented the videos with hand written field notes. With the notes, we could collect behaviors and interpretations that could not be collected otherwise.

We utilized also all available material about the case studies, including published articles describing the outcomes of their research processes and Jupyter notebooks that contained information about methods and detailed processes. These helped in crosschecking the details and analysis, and provided information that was not necessarily discussed during the interviews.

In our data collection, we used both data triangulation [31] (interview, demonstrations, field-notes and supplementary material on cases) and researcher triangulation [13] (two or three researchers present during every interview and demonstration session) to increase the reliability of our findings. By using researcher triangulation we were able ensure that there were a shared understanding about what happened during the data collection and we were able to collect more accurate accounts of information interactions and activities. The validity was also crosschecked by later showing our analysis (i.e., the workflow charts) to the participants and asked them to comment and validate our findings.

4 ANALYSIS

The data were transcribed (verbatim) and coded in several rounds. First, we cut the data into segments. The segments are relevant snippets of the interview transcriptions that were selected for further coding [cf., 38]. The criteria for selecting a segment was the level of its connection to the research project we were studying. This means that we left out parts of the interview where the researchers reflected, for instance, about how the project was funded, or other more general comments not directly tied to the specific research project's workflow.

During the next stage, we did initial coding of the selected segments. Each researcher coded the same interviews in relation to research activities (which have a goal, method, technique, object). We then discussed the research activity codes we had assigned to each segment. Based on the discussions we refined, deepened and clustered our initial codes to create a codebook. Our codebook creation process was mostly based on the NeDiMAH Methods ontology (NeMO) [7], and also informed by existing taxonomies, such as TaDiRaH [8] and is available on Zenodo [28]. Of the 80 activity codes in our codebook, 69 are in the NeMO ontology. We identified eleven additional activities including *Importing*, *Normalizing*, *Notetaking*, *Reading* and *Sampling*. During the second coding round, we individually reassigned the research activity codes from the codebook to the interviews. We then compared the coding done by each researcher and, after agreeing about the codes, assigned the final activity codes to the selected chunks. In addition, in the different phases above, we also coded for data and tools.

In a second phase, we coded for transitions between activities. We made an attempt to reconstruct the complete workflows, but note that the combination of interviews, research papers, datasets and documentation are not enough to capture the full complexity of the process, with iterations, dead ends, reversals and activities taking place in parallel. Rather, the available evidence leads to a sanitized and idealized version of the process. Still, we argue that these idealized workflows provide useful insights for developing and evaluating humanities research infrastructures.

In generating the activity transition graphs, we use the following assumptions. Each project has one or a few *initiating* activities, i.e. activities that are not preceded by any other activities in the project, and one or more *ending* activities, that have no following activities. Apart from some expected endings like writing and presenting a paper, there are some that indicate unfinished work. All other activities are *intermediate* activities that should have one or more preceding and one or more following activities.

We produced the workflow graphs in a number of steps:

---

[7] See http://nemo.dcu.gr/
[8] http://tadirah.dariah.eu/vocab/)

(1) First pass: coding transitions between activities while going through the interview segments in order. We assigned interview quote IDs to each activity. E.g., hypothesis formulation is described in quote 14, and is followed by problem recognition in quote 15, so the transition is *Hypothesis formulation (14) → Problem recognition (15)*. We add the quote ID to keep separate the different instances of a type of activity.

(2) Second pass: connecting activities with different quote IDs that represent the same instance of an activity. E.g. a specific Hypothesis formulation is referred to in both quotes 14 and 37. We merged both activity/quote ID pairs as Hypothesis formulation (14/37), so the transition code above becomes *Hypothesis formulation (14/37) → Problem recognition (15)*.

(3) Third pass: we identified disconnected parts in the visualization and added missing transitions to connect them.

(4) Fourth pass: we identified the *start* and *end* activities and added missing connections for any *intermediate* activities that had no preceding or following activity.

The resulting workflow visualizations for each project are included in Auxiliary material [28]. We discussed the workflows with the interviewees, who confirmed that these corresponded to their memory of the process and contained no egregious errors.

Lastly, we classified the transitions between activities also according to higher-order and lower-order cognitive complexity, with the aim of understanding when users transition between low and high complexity tasks, as this can inform designing interface features keeping overall cognitive effort in mind. The dichotomous classification is based on the aforementioned taxonomy of information processing skills created by Anderson and Krathwohl [1, pp. 67–68] and applied as cognitive complexity to tasks in IR evaluation by Kelly et al. [26]. The higher-order cognitive complexity represents the three highest levels in the taxonomy: Analyze, Evaluate, and Create. The lower-order complexity includes levels Remember, Understand, and Apply. We mapped activities from our codebook to information processing skills in the taxonomy, whereby some activities combine multiple skills, in which case the cognitive level of the activity is determined by the skill with highest cognitive level. In cases that might belong to both levels, we classified all instances to the same cognitive complexity level best matched the levels in the taxonomy [1]. The coded transitions included transitions between higher to higher, higher to lower, lower to higher and lower to lower. There were some parallel activities that lead to common activity after the transition, and they were coded accordingly. Collaboration was not coded, because it presented both lower and higher complexity.

5 FINDINGS AND DISCUSSION

5.1 Characteristics of research workflows

We describe and zoom into different parts of the workflows by signaling three broad areas within them, which roughly correspond to overarching phases in the each research project's life cycle: (1) initiation, (2) analysis, and (3) synthesis. The selection of activities belonging to each phase was based on the visualizations of the workflows (see [28, pp. 7–8] for the full visualizations). Both research projects included several tasks. Further, tasks included

Table 1: Phases and tasks in RP1

Phase	Tasks
Initiation	(a) conceptualization
Analysis	(b) archival visit; (c) social networks; (d) content analysis
Synthesis	(e) interpretation, contextualization and writing

Table 2: Phases and tasks in RP2

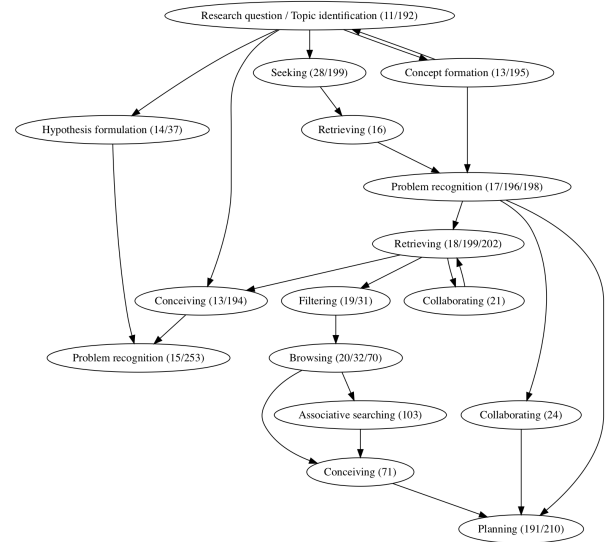| Phase	Tasks
Initiation	(a) conceptualization and gathering data; (b) manual small-scale reconstructing maritime careers
Analysis	(c) small-scale analysis of careers for a presentation; (d) standardizing and linking of place names to geographical gazetteers; (e) spatial analysis of geographic spreading of VOC maritime workers to Twitter and project website; (f) collaborating with a data scientist to cluster and link multiple contracts of individual persons; (g) large-scale reconstruction of maritime careers based on clustered contracts
Synthesis	(h) interpretation, and new hypothesis formulation

Figure 1: Transitions in the initiation phase of RP1.

various activities, that were coded according to our code book. Tables 1 and 2 show the phases and respective tasks in the research projects.

5.1.1 Initiation phase. The initiation phases of RP1 and RP2 are shown in Figures 1 and Figure 2.

- Research activities: both project workflows begin with activities such as 'research question/topic identification', and
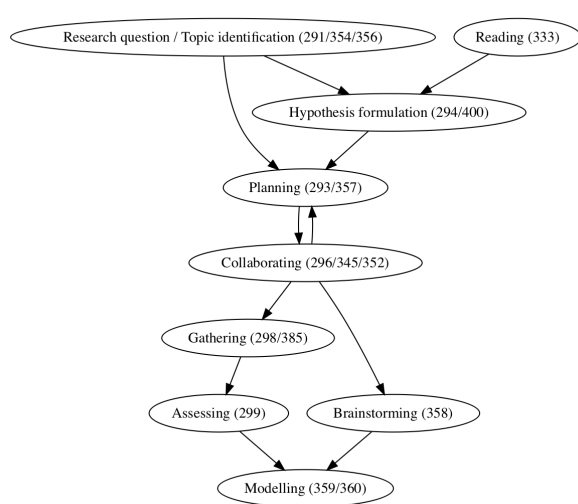
Figure 2: Transitions in the initiation phase of RP2.

'hypothesis formulation'. RP1 includes more information interaction activities (such as 'seeking', 'retrieval', and 'browsing') intertwined with mental/cognitive activities (such as 'concept formation' and 'problem recognition') than RP2. In turn, RP2 includes more data-related activities such as 'gathering,' 'planning,' 'evaluating,' and 'brainstorming.' This can be explained since RP1 departed from a question or intuition that was triggered in the scholar's mind by reading some publications and searching in a catalog, while RP2 departed from the intention to work and answer new research questions with previously well-known datasets by the scholars. In RP2, besides the substantive research question about the reconstruction of careers of migrant and native workers, the problematization of the methodology for linking the entities became a central research question from this initial phase. In both projects, we notice that 'collaboration' as activity begins early in the project after problem recognition.

- Tools: initiation activities were intertwined with seeking and retrieval activities (predominantly in RP1), and with data gathering (predominantly in RP2). In RP1 a library search engine (Worldcat) served as the triggered point for an archival phase (after locating an archive of interest) and for the entire project hypothesis formulation (the scholar formulated the initial hypothesis prompt by a feature offered by the search engine). In RP2, an archive's endpoint was used to obtain the data. In both projects, word processors was used for personal and collaborative note-taking and registration of research work and decisions.

5.1.2 Analysis phase. The analysis phase of RP1 is shown in Figure 3 and of RP2 in Figure 4.

- Research activities: activities related to the analysis phase differ in both projects. In RP1 'modeling' and 'extracting' predominate. The modelling here includes determining which
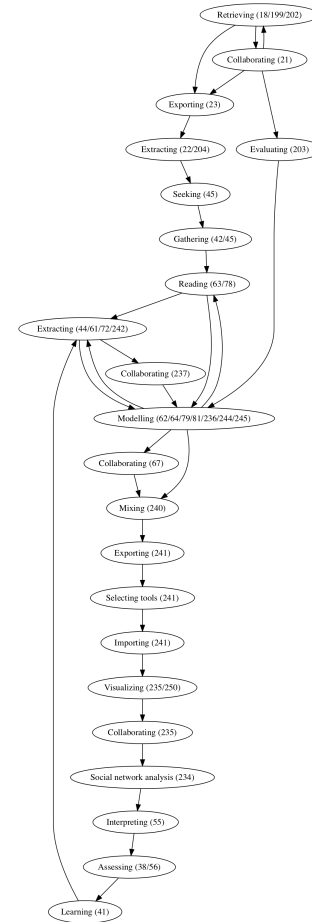

Figure 3: Transitions in the (social network) analysis phase of RP1.

elements to include in the social networks around migration policy, i.e. the names and roles of the publishers of migration journals, the authors of the forewords, the financiers, etc, which was done iteratively as different types of influencers where found in the analysed publications. In RP2 the 'data validation and reconciliation' activity is quite intensive, while the modelling of careers out of the maritime contracts is minimal and mostly left till later in the project and still on going. The activities that occur during the analysis phase are strongly influenced by the selected methods (social network analysis and cloud analysis in RP1, and intensive data reconciliation and linking necessary for spatial and statistical analyses in RP2).

- Tools: Spreadsheet software were intensively used by both projects during this phase, mostly in connection to extracting and modelling activities (in RP1) and to data validation, reconciliation and linking (in RP2), as well as data wrangling tools like Open Refine (RP2). Specific applications strongly connected to the selected method were also used: in RP1,
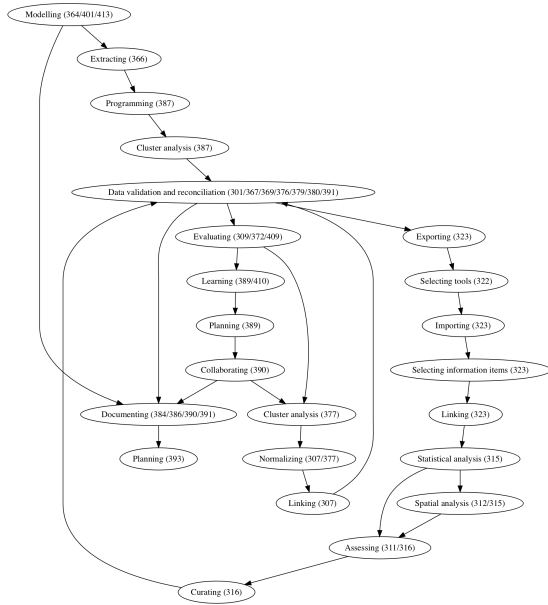
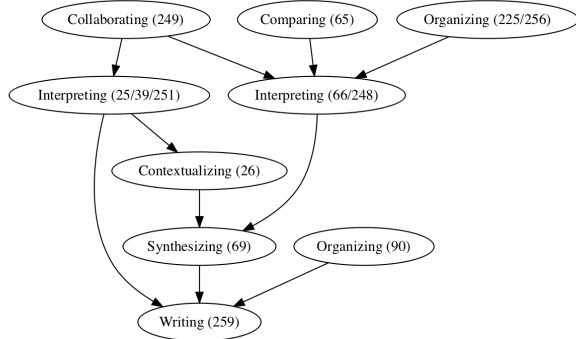**Figure 4: Transitions in the analysis phase of RP2.**

Figure 5: Transitions in the synthesis phase of RP1.

Gephi, an application to process and visualize social networks was used for one of RP1's research questions, and ArcGIS (a geographic information management system) was used in RP2 for their spatial analyzes. Both projects incorporated intensive programming (using for example Python programming libraries), and other algorithms designed by the data experts in the projects.

5.1.3 Synthesis phase. The synthesis phase of RP1 is shown in Figure 5 and of RP2 in Figure 6.

- Research activities: RP1 has a clear ending activity in the writing and presenting of a paper based on the analysis. The participants of RP2 discussed several unfinished threads that led to multiple *end* activities, where the researchers
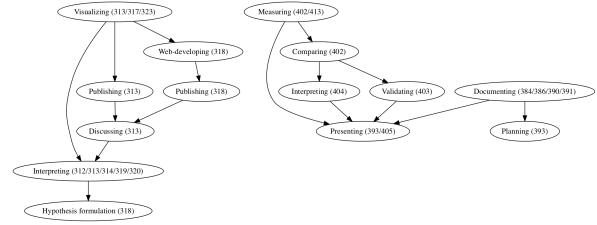

Figure 6: Transitions in the synthesis phases (two moments) of RP2.

have plans for specific next steps and specific hypotheses yet to be tested. Activities such as 'interpreting,' 'contextualizing,'discussing,' 'writing,' or 'presenting' are part of this phase, tied by 'collaborating,' and 'discussing' visualizations and results.
- Tools: besides the use of standard writing and presentation tools, important to highlight is the role that visualizations play in interpretation and synthesis, also in triggering new hypotheses and research questions.

5.2 Workflow transition analysis

We analyzed the transitions, firstly, in the Initiation Phase. Tables 3 and 4 show the most basic transitions types across the cognitive complexity levels. The transition types exemplify also how the analysis proceeded in the subsequent phases. There are more varied types of activity transitions from higher-order to higher-order than in the transitions between other types of cognitive complexity. Some activities include higher-order thinking skills and decision making, and others include computer mediated interactions. Some of the computer mediated interactions may also include manual working, such as copy-pasting or working with printed and paper materials, but some others may be fully computerized and digital, which makes them more suitable for automated handling.

The second phase, Analysis Phase, includes, in RP2, transition types (from lower-order to lower-order) that are more easily to be automatized, even as scripts. However, the situation in RP1 is different; the activity transition types scatter across all the cognitive-complexity types and include collaboration. Also, because the RP1 is divided into two branches (see Figure 3 for social network analysis task) including collaboration, its transitions are complex.

In the third, Synthesis Phase, most of the transition types are higher-order to higher-order, some types from lower-order to higher-order transitions. Combinations of higher-order and lower-order transitions may be difficult to support with systems, and there is also a noticeable level of collaboration involved. Further, the higher-order transitions in all the phases include critical decision making, that affect the subsequent activities and transitions, and they happen in the *strategic level* [cf., 5]. These activities are important to the subsequent activities and the transitions between them.

One general finding is the importance and prevalence of *collaboration*. This is partly due to our selection of collaborative projects, but collaboration is much more common in digital humanities research than in analogue research [44]. Collaboration takes place in

Table 3: Transition Types in Initiation Phase in RP1

Cognitive Complexity	Activity Transitions
Higher → Higher	Topic identification → Hypothesis formulation
	Problem recognition → Reading
	Topic identification → Conceiving
	Topic identification → Concept formation
	Concept formation → Problem recognition
	Concept formation → Topic identification
	Planning → Selecting method
	Hypothesis formulation → Problem recognition
Higher → Lower	Problem recognition → Retrieving
	Topic identification → Seeking
Lower → Higher	Retrieving → Problem recognition
Lower → Lower	Retrieving → Filtering
	Filtering → Browsing
	Browsing → Associative searching
	Seeking → Retrieving

Table 4: Transition Types in the Initiation Phase In RP2

Cognitive Complexity	Activity Transition
Higher → Higher	Topic identification → Planning
	Topic identification → Hypothesis formulation
Higher → Lower	N/A
Lower → Higher	Gathering → Assessing
Lower → Lower	N/A

all research phases, and is part of transitions at different levels of cognitive complexity and connected to different types of activities. However, we have only reported that collaboration exists, but this is a topic for future research.

Another finding is the centrality of *modelling* as activity and its connection to many other activities. It is a complex activity that bridges high level activities such as *brainstorming*, *planning* and *evaluating* to information gathering activities like *reading* and *browsing* and data transformation activities like *extracting*, *linking* and *classifying*.

Transitions happen between activities that may present very different levels of (i) computer/system involvement, (ii) collaboration, (iii) larger/smaller chunks of task processes. According to Bates [5], there are many questions about 1) the degree of user vs. system involvement in the search, and 2) the size, or chunking, of activities, meaning how much and what type of activity the user should be able to let the system to do at once (an intriguing question in designing research support tools is, which tasks that require high cognitive complexity can be performed by the system [48]). We used the cognitive complexity to solve the level of "brain involvement";

the rationale behind this is that in the expert tasks human decision making is crucial to their success. Transitions in the workflows happen fluently between all these cognitive complexity levels and they do not necessarily provide any clear hierarchical structure, as is shown in this research.

From the system design perspective, the value of analyzing transitions between high and low complexity activities is to know which higher-order thinking skills can form the background in which specific lower-order interactions take place, and therefore, to be able to think about ways in which the cognitive load of those higher-order cognitive activities can be supported. For instance, when *annotating* segments of a text or a text with codes from a codebook (low complexity) during *reading* (high complexity) there is a significant cognitive load if the codebook is still in development and the user is at the same time *thinking about what elements should be part of the model* (high complexity) in which the coded information is to be included. In such cases, showing the structure of the developing codebook in the sidebar can reduce the effort of deciding whether and how to modify or extent the codebook.

6 IMPLICATIONS

6.1 Methodological Reflection

This research aimed at reconstructing scholarly workflows of two research projects based on verbal narratives of the participants that were complemented with documents and publications. One challenge of this approach is the temporal order of activities. The order may be described differently in the interviews, in the publications and in documentation, and in places it is left implicit. However, all these traces are simplified accounts of a more non-linear process.

The coding of the activities and transitions brought some issues to the surface. The different background knowledge of the coders resulted in different levels of detail of activities. For instance, from one of the interview quotes, "I did also some data wrangling from the titles (removing stop words, and doing relative frequencies)" and the description in a resulting research paper of how this led to wordcloud visualizations, one of the coders (with an NLP background) included activities that were not directly mentioned but had to have happened, including *parsing*, *tokenizing*, *filtering* and *statistical analysis*, while another coder choose only *analyzing*. We note that this depends on the tool that is used. In this case, these detailed activities were explicitly programmed in the form of a Python script, but other tools wrap all these steps in a single interface feature resulting in a single user interaction. For the final coding, we used the more fine-grained level based on the interviewee showing us the Python script that contained all these activities as sets of coding instructions.

This issue challenges also the usefulness of interviews and other narratives in this kind of research. People describe past happenings as they remember them or how they like to present them, not necessarily as they really happened. However, this is also a question of research economics and not having possibilities to collect data by other means on full scale research projects that may take more than a year to be accomplished. Nevertheless, our work indicates that studying workflows on full research project scale level reveals connections between activities that cannot be studied in laboratory settings. Our study shows that the workflows include transitions

between different cognitive complexity levels, and that lower-order transitions lead to higher levels, and vice versa. This could be further connected to research phases, and would allow us to preserve the non-linear, iterative and detailed level of the workflow and demonstrate its complicated relation to the more idealized research phase level. In the other direction, lower level activities can be more directly connected to tools and data.

More research is needed to validate evaluation methods for process support based on activity transitions. We proposed a method to identify those transitions based on narratives of scholarly projects. Future work includes, on the one hand, to analyze more cases from different domains using the proposed methodology and, on the other hand, to conduct evaluation tests using transitions as the focus in simulated work tasks [9, 10]. However, we believe, that this kind of research still shows important aspects from the human point of view, and covers the research processes at large.

6.2 Implications for system design and evaluation

From a micro-level perspective to analyze scholarly workflows, as proposed in this paper, one of the challenges of research infrastructure projects is to develop support not only for individual activities, but also for transitions between activities. For example, the method proposed in this paper helps to identify potential hurdles for users to switch from one activity, e.g. extracting information, to another, say, clustering or exporting. Within the sciences, a common model for dealing with these transitions is to create pipelines of activities that happen in a fixed sequence [17, 43]. Our findings, together with earlier work on the research processes in the humanities [11, 35, 36], show that there are many complex workflows consisting of activities happening in parallel and activities involved in multiple transitions that loop back in expected and unexpected ways. The highly iterative nature of humanities research challenges the pipeline metaphor.

Our findings suggest other metaphors. Particularly, sets of activities that happen in a single interface bringing together multiple views and interaction options, is already happening in some faceted search interfaces and would also serve our cases. Queries, facets, search results, distribution plots, timelines and item views can be shown in a single interface, with which the user combines activities like seeking, retrieving, browsing, reading, comparing, extracting and annotating, in rapid succession, and combining multiple sources and/or types of information into a single mental model to make sense of complex data [14, 42]. Supporting such combinations of activities requires a *palette* of views and interaction options with which users can liberally mix activities into a complex non-linear workflow. For instance, in RP2, the standardizing and normalizing of person and place names was partially done using the tool Open Refine, which shows data as a spreadsheet and allows typical spreadsheet operations, but in the same interface allows users to e.g. summarize a column by showing the distribution of values with their frequencies in a faceted view on the left-hand side. Within this facet view, they can spot common mistakes in values and correct them, after which both the facet counts and the spreadsheet are updated. For RP2, this combination of different views and functionalities is used in iterations of normalizing, statistically analyzing

and evaluating the data. At any point, the tool offers one or more views on the data in a single screen, and a *palette* of functions to alter the views and modify the data. The *palette* metaphor may also be useful for combinations of activities such as *reading*, *modelling* and *extracting* as in the analysis phase of RP1 (Figure 3). Virtual Research Environments (VREs) like the HathiTrust Research Center[9] and the CLARIAH Media Suite[10] offer tools for searching, accessing and annotating documents in a single interface, which allows users to quickly iterate between such activities. Annotation functionality can bridge the transitions between on the one hand information seeking and retrieval and reading, and on the other hand information extraction and linking [8].

Focusing on transitions between activities instead of on evaluating entire workflows may help system evaluation efforts, which are taking the process perspective. Evaluation could happen in various levels, as in [24] but also taking the process nature into account. In terms of the transition types, evaluation could focus on low-level transitions in the case of data flows, while user evaluations could focus on system and tool flows and supporting transitions between activities. Further, evaluating tools in a workflow perspective should focus on critical transition points. Supporting the transitions from higher-order to higher-order activities could include helping, e.g., in conceptual differentiating and exploration, in making judgments and decisions, and by providing critical cues.

7 CONCLUSIONS

In this paper, we examined workflows from a micro-perspective and captured and analysed the transitions between activities. We used two scholarly projects as case studies, described their research activities and the transitions between the activities in detail. We showed full scale visualizations of the scholarly workflows and shed a light on transitions between activities and that they happen fluently between higher-order and lower-order cognitive complexity, and that collaboration is also critical in humanities research workflows. Our approach has implications to design and evaluate information systems and services that aim to support scholars in their complex and varied research processes. Unless the flows of data, tools, and tasks are taken into consideration, important aspects of information interactions will remain hidden.

REFERENCES

[1] Lorin W. Anderson and David R. Krathwohl. 2001. (abridg. ed.). Longman, NY.
[2] Sheila Anderson, Tobias Blanke, and Stuart Dunn. 2010. Methodological commons: arts and humanities e-Science fundamentals. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368, 1925 (2010), 3779–3796.
[3] Smiljana Antonijevic and Ellysa Stern Cahoy. 2018. Researcher as Bricoleur: Contextualizing humanists' digital workflows. *Digital Humanities Quarterly* 012, 3 (Nov. 2018).
[4] Joan C Bartlett and Elaine G Toms. 2005. Developing a protocol for bioinformatics analysis: An integrated information behavior and task analysis approach. *Journal*

---

[9]See https://analytics.hathitrust.org/
[10]See https://mediasuite.clariah.nl/

of the American Society for Information Science and Technology 56, 5 (2005), 469–482.

[5] Marcia J Bates. 1990. Where should the person stop and the information search interface start? *Information Processing & Management* 26, 5 (1990), 575–591.

[6] Agiatis Benardou, Panos Constantopoulos, and Costis Dallas. 2013. An Approach to Analyzing Working Practices of Research Communities in the Humanities. *International Journal of Humanities and Arts Computing* 7, 1-2 (Oct. 2013), 105–127. https://doi.org/10.3366/ijhac.2013.0084

[7] Agiatis Benardou, Panos Constantopoulos, Costis Dallas, and Dimitris Gavrilis. 2010. Understanding the Information Requirements of Arts and Humanities Scholarship. *International Journal of Digital Curation* 5, 1 (2010), 18–33.

[8] Peter Boot, Ronald Haentjens Dekker, Marijn Koolen, and Liliana Melgar. 2017. Facilitating Fine-Grained Open Annotations of Scholarly Sources. In *DH2017 Digital Humanities 2017 Conference Abstracts*. McGill University and Université de Montréal, Montréal, Canada, 167–169. https://dh2017.adho.org/abstracts/198/198.pdf

[9] Pia Borlund and Peter Ingwersen. 1997. The development of a method for the evaluation of interactive information retrieval systems. *Journal of documentation* 53, 3 (1997), 225–250.

[10] Pia Borlund and Jesper W. Schneider. 2010. Reconsideration of the simulated work task situation: a context instrument for evaluation of information retrieval interaction. In *Proceedings of the third symposium on Information interaction in context*. ACM, New York, NY, USA, 155–164.

[11] Marc Bron, Jasmijn Van Gorp, and Maarten de Rijke. 2016. Media studies research in the data-driven age: How research questions evolve. *Journal of the Association for Information Science and Technology* 67, 7 (2016), 1535–1554.

[12] Katriina Byström and Kalervo Järvelin. 1995. Task Complexity Affects Information Seeking and Use. *Inf. Process. Manage.* 31, 2 (1995), 191–213. https://doi.org/10.1016/0306-4573(95)80035-R

[13] Norman K. Denzin. 2017. *The research act: A theoretical introduction to sociological methods*. Routledge, Somerset.

[14] Brenda Dervin and Michael Nilan. 1986. Information needs and uses. *Annual review of information science and technology* 21 (1986), 3–33.

[15] Yrjö Engeström, Reijo Miettinen, Raija-Leena Punamäki, et al. 1999. *Perspectives on activity theory*. Cambridge university press, Cambridge, UK.

[16] John C. Flanagan. 1954. The critical incident technique. *Psychological bulletin* 51, 4 (1954), 327.

[17] Daniel Garijo, Pinar Alper, Khalid Belhajjame, Oscar Corcho, Yolanda Gil, and Carole Goble. 2014. Common motifs in scientific workflows: An empirical analysis. *Future Generation Computer Systems* 36 (July 2014), 338–351. https://doi.org/10.1016/j.future.2013.09.018

[18] Steffen Hennicke, Stefan Gradmann, Kristin Dill, Gerold Tschumpel, Klaus Thoden, Christian Morbidoni, and Alois Pichler. 2012. *D3.4 – Research Report on DH Scholarly Primitives*. Technical Report. Humboldt-Universität zu Berlin. https://dm2e.eu/outputs/

[19] Rik Hoekstra and Marijn Koolen. 2019. Data scopes for digital history research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, 2 (2019), 79–94.

[20] Hugo C. Huurdeman and Jaap Kamps. 2014. From Multistage Information-seeking Models to Multistage Search Systems. In *Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14)*. ACM, New York, NY, USA, 145–154. https://doi.org/10.1145/2637002.2637020

[21] Peter Ingwersen and Kalervo Järvelin. 2005. *The turn: Integration of information seeking and retrieval in context*. Vol. 18. Springer Science & Business Media, Dordrecht, The Netherlands.

[22] Bernard J. Jansen, Danielle Booth, and Brian Smith. 2009. Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management* 45, 6 (2009), 643–663.

[23] Kalervo Järvelin, Pertti Vakkari, Paavo Arvola, Feza Baskaya, Anni Järvelin, Jaana Kekäläinen, Heikki Keskustalo, Sanna Kumpulainen, Miamaria Saastamoinen, Reijo Savolainen, and Eero Sormunen. 2015. Task-based information interaction evaluation: The viewpoint of program theory. *ACM Transactions on Information Systems (TOIS)* 33, 1 (2015), 3.

[24] Jaana Kekäläinen and Kalervo Järvelin. 2002. Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance. In *Proceedings of the 4th CoLIS conference*. Libraries Unlimited, Greenwood Village, Colorado, 253–270.

[25] Diane Kelly et al. 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval* 3, 1–2 (2009), 1–224.

[26] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*. ACM, New York, NY, USA, 101–110.

[27] Rob Kitchin. 2014. Big Data, new epistemologies and paradigm shifts. *Big data & society* 1, 1 (2014), 1–12.

[28] Marijn Koolen, Sanna Kumpulainen, and Liliana Melgar-Estrada. 2020. A Workflow Analysis Perspective to Scholarly Research Tasks-Auxiliary materials. Zenodo. https://doi.org/10.5281/zenodo.3626814.

[29] Carol C. Kuhlthau. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science* 42, 5 (1991), 361–371.

[30] Sanna Kumpulainen. 2014. Trails across the heterogeneous information environment: manual integration patterns of search systems in molecular medicine. *Journal of Documentation* 70, 5 (2014), 856–877.

[31] Sanna Kumpulainen. 2017. Task-Based Information Searching: Research Methods. In *Encyclopedia of Library and Information Sciences*. CRC Press, Boca Raton, 4526–4536.

[32] Sanna Kumpulainen and Kalervo Järvelin. 2010. Information Interaction in Molecular Medicine: Integrated Use of Multiple Channels. In *Proceedings of the Third Symposium on Information Interaction in Context (IIiX '10)*. ACM, New York, NY, USA, 95–104. https://doi.org/10.1145/1840784.1840800

[33] Sanna Kumpulainen and Kalervo Järvelin. 2012. Barriers to task-based information access in molecular medicine. *Journal of the American Society for Information Science and Technology* 63, 1 (2012), 86–97.

[34] Sanna Kumpulainen, Heikki Keskustalo, Boyang Zhang, and Kostas Stefanidis. 2020. Historical reasoning in authentic research tasks: Mapping cognitive and document spaces. *Journal of the Association for Information Science and Technology* 71, 2 (2020), 230–241. https://doi.org/10.1002/asi.24216

[35] Catherine Marshall and Gretchen B. Rossman. 2014. *Designing qualitative research*. Sage publications, Thousand Oaks, California.

[36] Liliana Melgar, Marijn Koolen, Hugo Huurdeman, and Jaap Blom. 2017. A Process Model of Scholarly Media Annotation. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR '17)*. ACM, New York, NY, USA, 305–308. https://doi.org/10.1145/3020165.3022139

[37] Liliana Melgar Estrada. 2016. *From social tagging to polyrepresentation: A study of expert annotating behavior of moving images*. Doctoral thesis. Universidad Carlos III de Madrid, Madrid, Spain.

[38] Liliana Melgar Estrada and Marijn Koolen. 2018. Audiovisual media annotation using qualitative data analysis software: A comparative analysis. *The Qualitative Report* 23, 13 (2018), 40–60.

[39] Carole C. Palmer and Melissa H. Cragin. 2008. Scholarship and disciplinary practices. *Annual Review of Information Science and Technology* 42, 1 (2008), 163–212. https://doi.org/10.1002/aris.2008.1440420112

[40] Carole L. Palmer, Lauren C. Teffeau, and Carrie M. Pirmann. 2009. *Scholarly information practices in the online environment*. Technical Report. OCLC Research. http://www.oclc.org/programs/publications/reports/2009-02.pdf

[41] Vayianos Pertsas and Panos Constantopoulos. 2017. Scholarly Ontology: modelling scholarly practices. *International Journal on Digital Libraries* 18, 3 (Sept. 2017), 173–190. https://doi.org/10.1007/s00799-016-0169-3

[42] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 269–276.

[43] Ricky J Sethi and Yolanda Gil. 2017. Scientific workflows in data analysis: Bridging expertise across multiple domains. *Future Generation Computer Systems* 75 (2017), 256–270.

[44] Lisa Spiro. 2009. Collaborative Authorship in the Humanities. Digital Scholarship in the Humanities. https://digitalscholarship.wordpress.com/2009/04/21/collaborative-authorship-in-the-humanities/

[45] Ciaran B. Trace and Unmil P. Karadkar. 2016. Information management in the humanities: Scholarly processes, tools, and the construction of personal collections. *Journal of the Association for Information Science and Technology* 68, 2 (2016), 491–507. https://doi.org/10.1002/asi.23678

[46] John Unsworth. 2000. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this. In *Symposium on Humanities Computing: Formal Methods, Experimental Practice*, Vol. 13. King's College, London, 5–00.

[47] Pertti Vakkari. 2001. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation* 57, 1 (2001), 44–60.

[48] Joris J. van Zundert and Ronald Haentjens Dekker. 2017. Code, scholarship, and criticism: When is code scholarship and when is it not? *Digital Scholarship in the Humanities* 32, suppl_1 (2017), i121–i133.

[49] Barbara Wildemuth, Luanne Freund, and Elaine G. Toms. 2014. Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation* 70, 6 (2014), 1118–1140.