PROCESS ARCHITECTURE FOR DIGITAL LIBRARIES USING RIVA AND ROLE ACTIVITY DIAGRAMING

Zaheer Abbas Khan, Mohammed Odeh, and Tony Solomonides
Centre for Complex Cooperative Systems, Faculty of CEMS,
University of the West of England, BS16 1QY, Bristol, UK
{Tel: +44 117 3283279, fax: 44 117 3282587}
{Zaheer2.Khan, Mohammed.Odeh, Tony.Solomonides}@uwe.ac.uk

ABSTRACT

A single Digital Library (DL) contains enormous research knowledge which can be used for several application domains e.g. e-learning, e-research etc. Due to enormously increasing digital contents and number of facilitating services, DLs face challenges of seeking wide-scale deployment solutions in Technological Spaces (TS). Such solutions necessitate: (i) understanding the process models of DLs at both macro and micro levels, and (ii) identifying the suitable candidates from TS for their successful deployment. This paper is an effort to cover the first part of the above challenges and captures the socio-technical aspects in DL processes by modelling its processes both in Riva and Role Activity Diagrams (RAD). A first-cut Riva based architecture of DLs provides a macro view of inter-communicating and evolving complex processes. This has been further elaborated to develop a micro view by using RADs applied to the Scientific Publishing Process, as an example. This macro-micro process modelling combination helps to understand, identify and reduce the technical implications that may arise at later stages of DL system development, deployment and evolution. Finally, this paper is a step forward towards identifying DL processes and the role of Riva and RADs towards their enactment.

Keywords: *Digital Library Processes, Scientific Publishing Process, Business Process Architecture, Process Modelling, Riva, RADs.*

1. INTRODUCTION

Digital Libraries (DLs) can be considered as one of the special forms of Computer Supported Collaborative Learning (CSCL), where it can provide backend support for eLearning, eResearch and eEducation. Despite its vast usage and advantages, DLs lack wide-scale deployments. This is because of the increasing data, metadata management and interoperability issues. Such deployment efforts can be significantly more complicated when DLs are involved with multi-disciplinary research domains and use multiple Technological Spaces (TS) [23] as their underlying testbed. In this regard, recent evolutionary trends in DLs domain are in pursuit of adopting distributed technologies based on Services Oriented Architectures (SOA), P2P & Grid computing [5, 7, 8, 10] and Open standards like Open Digital Libraries (ODL) [4]. Such a wide-scale technology adoption may lead to DL system deployments in unstable state which might require additional supporting layers of functionality to cope with the evolving DL requirements. In this regard, the modelling of DL processes helps to reduce the extent of such challenges by highlighting and analysing the underlying development and deployment complications.

Almost, all DLs have some common elements e.g. data, metadata and processes [10]. Furthermore, access to volumes of information spread across multiple nodes is enhanced through DL processes combined with metadata harvesting or federation [4]. In a working DL model, both data and metadata go through different processing activities, where these activities can be grouped into different processes based on some common goals and objectives. These processes interact with each other in a predefined manner to carry out the successful DL execution model. For example, Figure 1 highlights an abstract-level process view of DL showing interaction among different DL processes. This high level view is suitable for a naïve user who aims to understand the relationships among DL processes. This view has been kept highly generic, so that it can be applied to a variety of exiting DLs. In addition to this, a high-level process view of DLs has been developed using the Riva method [13], with a first-cut process architecture of DLs at a macro-level. A further refinement of the architecture using RADs [3] brings up the micro level of individual processes. This micro view shows the activities within individual processes and the way processes interact with each other. RAD process improvement feature may help DLs to refine and find improved deployment solutions within TS.

In this paper, we have limited our discussion to identifying DL processes, their interaction in an organised way by using the Riva method and then modelling of the *Scientific Publishing* (SP) process using RADs as an example of a DL process. In the next section, we briefly address related work in digital libraries followed in section 3 by a discussion on process architecture using the Riva

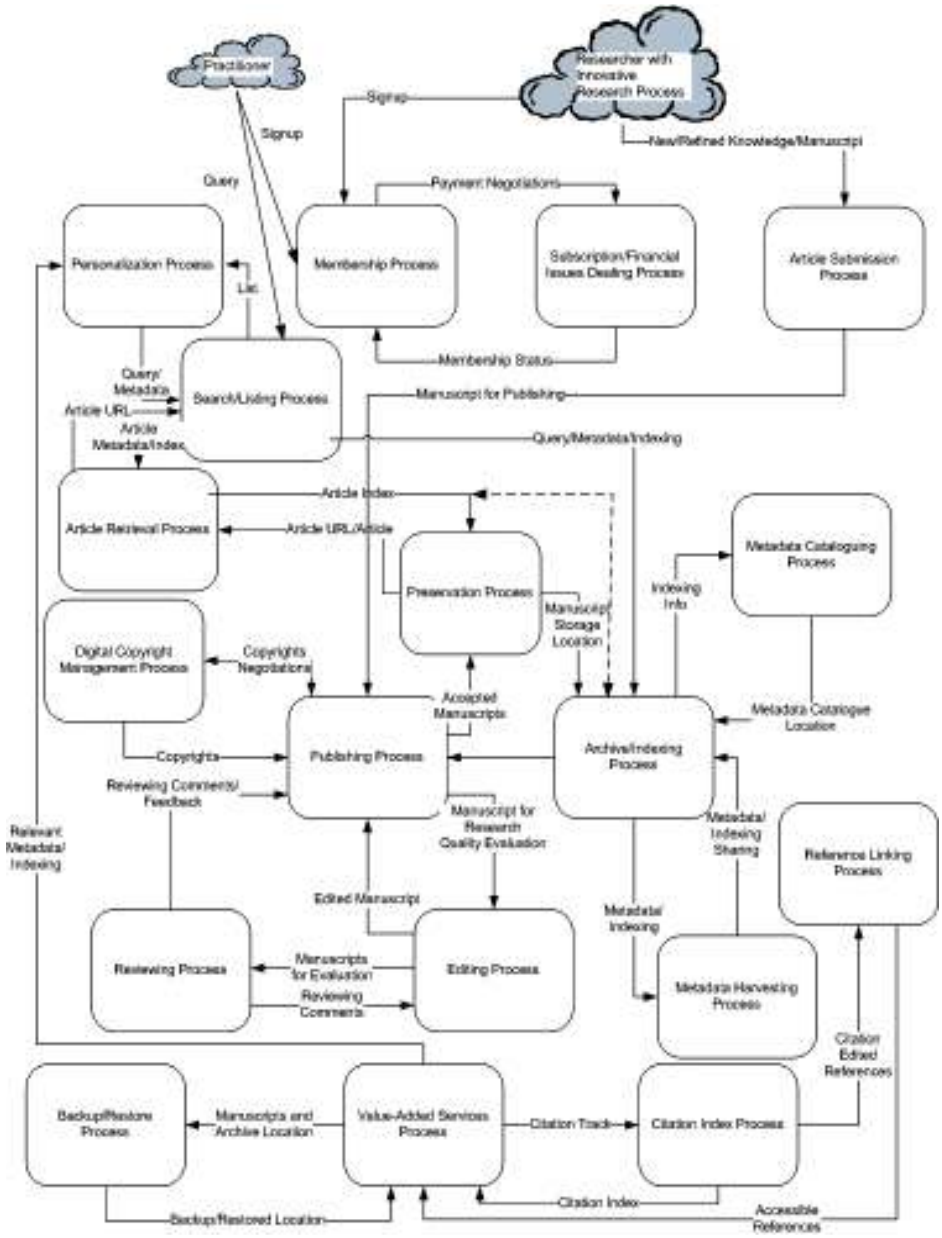

Figure 1:High Level View of DL Processes

method with emphasis on the concept of Unit of Works (UOW) in DLs using UOW diagrams. Section 4 introduces the DL case study along with its processes briefly described and the SP process modelled using RADs. Finally, major observations are discussed in section 5 along with the conclusion and future work.

2. RELATED WORK AND BACKGROUND REVIEW

Recent work in the domain of DLs modelling is well-supported by the 5SL (Stream, Structural, Spatial, Scenarios and Societies Models) language for declarative specification and generation of digital libraries [6] as it provides a comprehensive view of DLs in covering five different modelling aspects. Their algebraic formal approach seems to be the only one existing mathematical formal models for DLs. These models lay down the foundation for covering both socio-technical aspects. On the other hand, these models are comparatively more complex than RAD models. Bjork et. al. modelled the Publishing Process using IDEF notations [9], with emphasis on the functional view of this process as opposed to our RAD modelling of the SP process in covering both behavioural and functional aspects in a step towards a better understanding of the underlying DL processes. Due to its ease of comprehension and simplicity in modelling, RAD has been used in many business applications e.g., Healthcare [19] and Peer Reviews process [18] etc. Some work has been done in transforming RAD models into other modelling languages [17, 20]. This shows that RAD is adaptable to other modelling languages as well as suites the context of TS.

3. DL PROCESS MODELLING USING RIVA ARCHITECTURE AND RAD

RAD modelling captures the systematic coordinated behaviour of business processes [3]. Despite its use in many applications, it appears that RAD modelling emulates more deterministic and static behaviour [21]. This might restrict RAD modelling to represent process dynamism under emerging environments. Furthermore, special considerations are in need to introduce process level active models [1]. These issues are even more critical when organisations tend to adapt process-centric approach, where inappropriate mapping and coordination between individual processes may not work properly to achieve the overall business objectives [12]. These issues have raised the requirement of flexible and adaptable process architecture with continuous support to deal with evolvable process models [12]. In this regard, the Riva method provides quite suitable approach to tackle such a challenge. In this paper, a composition of DL Unit of Works (UoW) is presented in Figure 2, followed by a first-cut process architecture using the Riva method as shown in Figure 3.

Figure 2: UOW Diagram of DLs

Clearly, the Riva method has enabled providing a high level process view of the DL organization. It also covers in depth details of DL processes from management point of view. However, it lacks in explaining what a certain process embodies. This detail is covered by modelling the individual processes using RADs. Each Case Process (CP) and Case Management Process (CMP) in the first-cut architecture can be represented by a complete RAD model. This approach has helped in linking all the individual processes modelled in RAD under a strong decomposable and decentralized process management infrastructure. And, this suites many TS like P2P, Grid and SOA, where a composed complex service is managed independently and interacts with others as per interaction needs.

In short, process architecture aims to identify the available processes in the DLs and highlight the interdependency among them. This wider view of the organizational processes of DLs facilitates incorporating new emerging processes in order to analyse their effect on the existing ones, in addition to enhance the understanding and improvement of DL business prcoesses. But without RAD models, which deal with micro activities and interactions, the enactment of the business process architecture becomes more difficult. In this regard. We modelled the SP process using RADs which made us conclude that it is difficult to identify or limit the scope of individual processes in large organisations such as DLs. One process may encompass other sub-processes or roles and activities which might be represented as separate independent communicating processes. For example, it is obvious from Figure 4 that the model involves certain aspects of "Editing Process" as part of SP process, whereas, the "Editing Process" is shown separately in both Figures 1 and 3. This part-integration and adaptation of the "Editing Process" in the SP process increases its scope and makes its representation more complex. And, such scope restriction may limit the actual representation of the SP process. This can further restrict its users to analyse and seek enactment possibilities.

4. THE CASE STUDY

In this section, the core processes of DLs are identified and modelled using Riva in a first-cut Business Process Architecture with RAD modelling of the Scientific Publication process (covered in section 4.2) of DLs. Concrete library systems may vary from each other due to both organizational policies and architectural differences. But, the general processes among all the systems remain the same.

4.1. DL PROCESSES

DL processes, as shown in Figure 1, are normally initiated by an *InnovativeResearch process*, where researchers generally conduct research and generate knowledge for DL preservation. In simple terms, this process triggers a particular instance of DLs. The very first *Membership process* creates a new membership ID for new members. This ID allows *Practitioners* to access literature within that DL. Membership processes for different DLs may vary due to different subscription policies. Some DLs may not charge fee for membership (e.g. e-prints are accessible at http://arxiv.org/) and some may charge for the services and material accessed through web (like ACM, IEEE DL). Furthermore, membership fee also varies due to membership type, e.g. student, educational institutions and individual

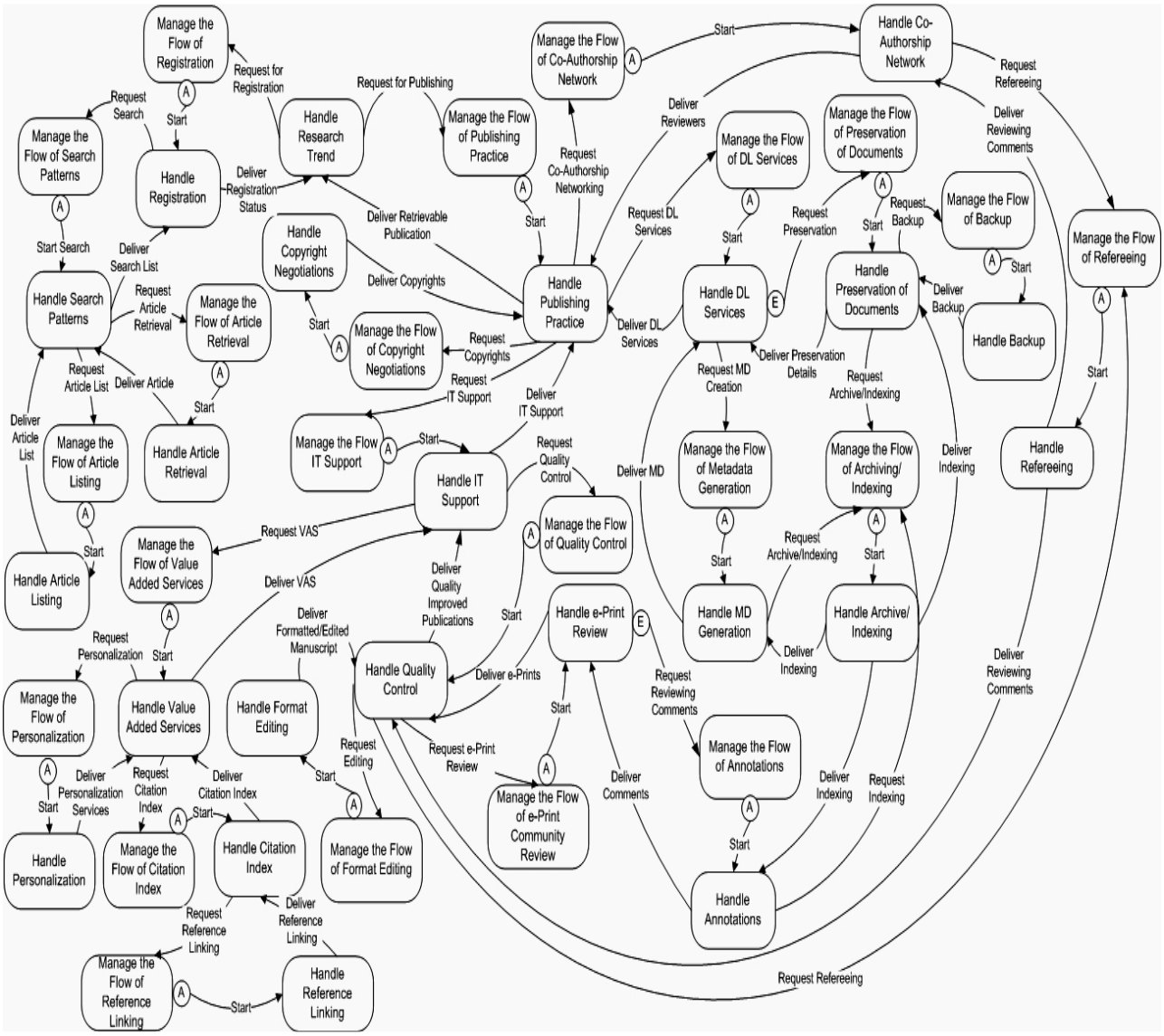

Figure 3: First-cut Process Architecture of DLs

members etc. Additionally, this subscription may be either for short (one time download) or long term (yearly basis). The *Submission process* may be considered a sub-process of the Publishing process which is activated after manuscript submission by the Researcher. In the *Preservation process,* published manuscripts are preserved either at the local or the distributed repository. After preserving a document, it forwards the storage index to the *Archiving/Indexing process* to update the storage records, so that, later retrieval of that document will be highly efficient. This further facilitates *Searching/Listing process* which extracts indexing information and the relevant literature from the widely deployed DLs. It involves specialised and highly performing algorithms to search specific contents from digital repositories. At the end, electronic contents can be accessed through a retrieved entity list from the *Search process.* And, based on user privileges these contents can be either stored locally or directly printed.

In the *Metadata Cataloguing process,* metadata are created and indexed in the metadata catalogue. Also, this process returns the metadata catalogue location to the *Archiving and Indexing process.* This process shares indexing, metadata, and manuscript storage information with major DL processes which are responsible for the successful and efficient functionality of the complete system. This also facilitates *Article Retrieval process* which generates queries and gets the entity listing from the *Searching/Listing process* followed by the article's URL to retrieve it directly from its storage. This search process is further facilitated with the *Metadata Harvesting process* especially when several repositories are part of a DL. By using the metadata harvesting technique, each repository maintains a local metadata catalogue to locally peform users' search. In contrast, metadata federation requires spreading a query on remote repository catalogues [4].

The Editing process ensures both quality and authenticated publishing. Manuscripts are passed to editors for editing. Acquisition editing, copy editing, developmental editing, fact checking, indexing, information design, page design, permissions editing, production editing, project editing, proof-reading and technical editing are different facets of the Editing process [11]. This however emphasises the impact of digital content submission on easing the process of heavy content editing. This process further supports the *Reviewing process, where* manuscripts are reviewed with relevant feedback returned. Also, the *Reviewing* process aims to control the quality of research articles published in conferences, and journals [22]. Some archives do not go through this process as the reviewing process is performed by practitioners after publishing, e.g., e-prints. After editing and acceptance of manuscript for publishing, *Digital Copyright Management process* is initiated to deal with the copyright issues.

A number of Value-Added Service (VAS) that support DL processes are mostly captured in sub-processes like Backup, Citation Indexing, Reference Linking and Personalisation processes etc. For example, the *Personalisation process* normally interacts with practitioners who subscribe to receive specific notification (Push model). The *Backup/Restore process* obtains the references of new manuscripts and creates regular backups. The *Citation Indexing process* is used to keep track of citations made in other papers e.g. CASPER model [16], for instance to calculate the impact factor [15] of specific publication, e.g., CiteBase (http://www.citebase.org/). And, finally the *Reference Linking process* is used to link the active source of the citations in specific publication as it retrieves the edited references through the Citation Indexing process and links them to active and accessible sources.

4.2. SCIENTIFIC PUBLISHING (SP) PROCESS

Publishing is a very complex process. Normally, journal issues and articles, conference/workshop peer reviewed research papers and e-prints go through stringent publishing process before getting into digital libraries. This process involves many other sub-processes like submission, editing, peer-review etc. to cover its wide-scope.

4.2.1. Journal Publishing

Manuscripts are submitted by researchers for Journal publications before the submission deadline of Call for Papers (CFP). Submission status is notified back to Researchers. Successfully submitted manuscripts are forwarded to Editors for both format consistency check and feedback on the peer-reviews. Reviewing may result in acceptance, minor revisions or rejection of submitted papers. Despite the delay and other problems [14, 22] involved in the ***Peer-Reviewing*** process, it authenticates the quality of the research work. The Editorial board uses personal references and/or co-authorship networks to find and agree suitable reviewers/referees [14]. Reviewers, after analysis and evaluation based on reviewing policy, rank the manuscript and forward their feedback to Editors who make final assessment whether to accept or reject the manuscript. Final acceptance is based on the reviewing comments, editorial board's decision and acceptance frequency for specific Journal. Authors of accepted papers are further requested to submit the respective camera-ready papers along with negotiated copyrights. After receiving the camera ready, it is forwarded to Multimedia Content Management (MMCM). MMCM is an abstract term that is referred to with the responsibility of controlling the internal workflow of a DL. It also deals with both storage and retrieval of publications from repositories. Furthermore MMCM initiates the Metadata Cataloguer (MC) and forwards the camera ready for metadata creation. After metadata creation, the camera ready paper is preserved in a repository. Both metadata and storage index are provided to MC that stores the metadata catalogue in the Catalogue Repository (CR).

Results of the above interactions are also inserted either in the external or the internal Bibliographic Services (BS), so that search services can find related records. These BSs play an important role in searching for a specific article and annotating the e-print/pre-print comments by the community. It also supports notification of specific publication based on keywords in the Personalization process.

4.2.2. Open-Archive Publishing

Open-Archives follow different but very simple approach than Journal publishing practice. Documents which are published under such practice are considered as e-prints [14, 22]. Such documents are edited after submission by Editors for appropriate Open Archive preservation format (some archives also use user edited documents) and then forwarded to the MMCM for metadata creation and preservation in the Repository. There is no peer-review process involved in this case and no copyright negotiations are followed as usual. However, readers of such published material can provide their feedback on the quality and value of the submitted work which can be very useful for potential new readers.

4.2.3. Conference/Workshop Publishing

Conference and workshop publishing normally follow the same process described in Journal publishing with the particular assessment of the value and quality of submitted work being a journal submission.

4.3. RAD MODEL OF THE SP PROCESS

RAD model for SP process is shown in Figure 4, where the workflow in this process is presented with the participant roles and their interactions. The roles involved in this process are:

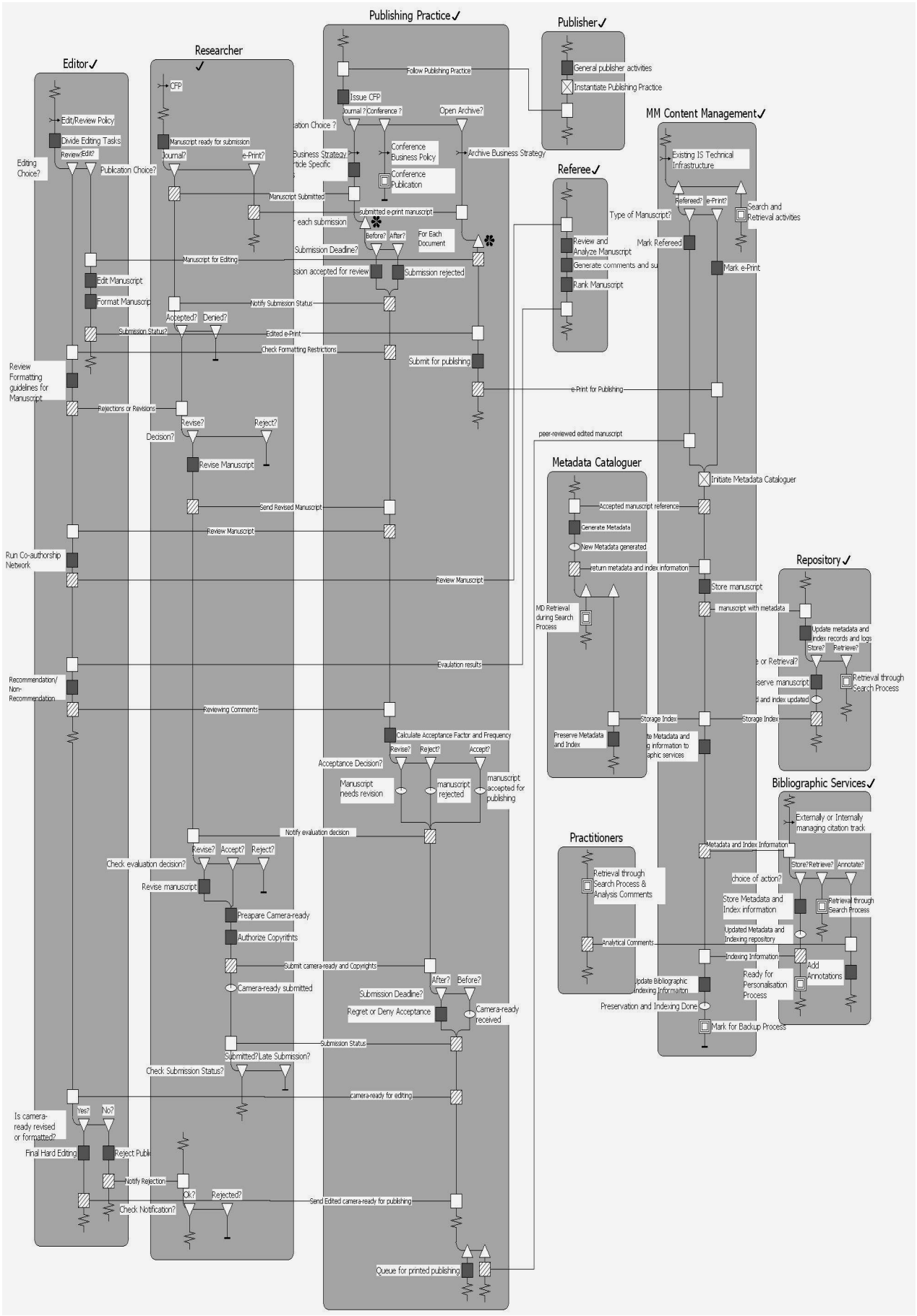

Figure 4: RAD Model of Scientific Publishing Process

- **Researcher:** The research findings of a researcher are transferred or specified in manuscripts which may get feedback from colleagues to improve on such findings and manuscript itself. The researcher is responsible for selecting the place and timely submission. Also, the researcher is responsible for preparing camera-ready of accepted manuscript by incorporating any reviewing comments. Camera-ready papers along with copyrights are to be submitted before the submission deadline. After acceptance of camera-ready manuscript for publication and making it available in digital archives, its indexing information is returned to the researcher for future reference.
- **Publisher:** This role is responsible for instantiating the whole Publication process. In electronic publishing, publication practice imitates the role of Publishers.
- **Practitioner:** Practitioners can be either general audience or researchers. They can extract documents from DLs by using available IT services (e.g. Internet). Normally e-Prints receive reviewing comments from practitioners and are annotated using their metadata/bibliographic services.
- **Referee:** These are also known as Reviewers, play an important role in the peer-reviewed publications. They critically analyse and evaluate the manuscripts against a well defined reviewing policy [22]. They rank the manuscript for acceptance, revision or rejection. Referees also provide comments about manuscript which are then forwarded to Researchers through Editors.
- **Bibliographic/Metadata Indexing Service (BS):** This (role) service is responsible for maintaining the bibliographic metadata of publications and serves as the index or the registry service for future retrieval. It also initiates the personalisation process which supports the push model of information storage and distribution.
- **Editor:** Editors are members of the Editorial Board of a certain type of published work (e.g. journal). The editor ensures that the submitted manuscripts comply with formatting guidelines, and can accept reject submitted work in addition to ensuring adherence to formatting guidelines. Also, the Editor forwards the manuscript to already allocated referees for quality reviews.
- **Publishing Practice:** This role provides a generic approach for the scientific publications of journals, conference/workshop and open access and archive publications. Being in electronic publishing era, this role imitates the activities of the publisher and performs most of the activities for which the publisher is responsible for. After being initialized, it triggers a specific Call for Paper (CFP). Normally, open archives do not need any CFP. It ensures that timely submission is being made by the researcher. All successful submissions are forwarded to editors for formatting review and the peer review process. After receiving reviewing comments from editor, the publishing practice calculates the publication acceptance frequency/factor based on the specific Journal's policy. It also notifies the decision to the researcher. In the case of revision, the camera-ready version of the paper is requested with a specific deadline in addition to negotiating copyright issues with the Researcher. Timely camera-ready submissions are forwarded for preservation to the MMCM.

  The role of the publishing practice varies for e-print open archives. Such archives follow specific business policies. For example, freely accessible articles are mostly published in such archives with the motivation to provide access to articles from specific or heterogeneous scientific domains. Normally, submitted documents are forwarded to editors for necessary editing and formatting tasks, but they are mostly based on the researcher's credibility. Such articles are published as they are without any editing.
- **Multimedia Content Management (MMCM):** The MMCM stores and makes accepted manuscripts accessible using an information system. To keep the track of journal and open archive publications, it initiates the Metadata Cataloguer for metadata management and forwards the accepted manuscripts to repositories which can be both centralised or distributed for storage. Also, MMCM mediates index and metadata Bibliographic Services for external public access, and initiates the backup process for preserved documents.
- **Metadata Cataloguer (MC) & Catalogue Repository (CR):** MC is responsible for creating the metadata for new manuscripts and returns unique references of metadata back to the MMCM. The CR provides storage mechanism for metadata catalogues.
- **Repository:** Accepted manuscripts are stored in specific archives. These archives can be centralised or distributed so as to provide transparent storage and retrieval of accepted manuscripts. In the case of storage, it returns the indexing information added along with metadata to the MMCM for BS and MC registration.

5. DISCUSSUIN AND CONCLUSION

Both Riva based first-cut architecture and the RAD model of the SP process contribute towards a simplified and highly intuitive approach to understanding of digital library processes.. This has resulted in identifying both areas of concern and scope for DL process improvement. For example, since DL has been shown (as per the above models) as a web of complex processes, it introduces certain process scope conflicts. This, however, can be seen in the SP process as it is

inappropriate to separate the scope of editing from publishing process, as they complement each other. In addition, the above proposed process architecture of DLs shows the dependency between processes which poses some further investigation into the cohesiveness of DL processes and the degree of coupling between them. However, this architecture possesses some flexibility in that it allows emerging processes to integrate with existing DL architectures using CMP and CP model in the first-cut Riva process architecture.

Furthermore, the process modelling of individual DL processes at both the macro and the micro levels provides an overall working model of DLs that could facilitate incorporating emerging process requirements. For example, a new DL process like 'Data Replication process' can be integrated in the first-cut Riva model with all its micro details specified in its respective RAD model with less significant impact on the rest of DL processes. Similarly, the SP process can be improved by using different and/or new role composition.

In conclusion, the combined Riva and RAD business process modeling approaches applied to the domain of DLs has demonstrated their significance to complex problem domains such as DLs. Also, the results of these models pave the way for further research investigations into the business process improvement of digital libraries being a relatively new emerging field with the high potential of emerging requirements. In addition, these models of DLs are a step forward for the transparent enactment of organizational business processes, associated with complex domains such as digital libraries, into distributed environments such as Grid-based infra-structures. In particular, our future work concentrates [24] on the deployment of the RAD based DL process models over Grid-based service-oriented architectures using formal approaches, and in particular Pi-ADL being a member of the Pi-Calculus family [2].

REFERENCES

[1] Curtis B. et. al., "Process Modelling", *Communications of the ACM*, Vol 35, No. 9, 1992.

[2] Oquendo F., "pi-ADL: An Architecture Description Language based on the Higher-Order Typed pi-Calculus for specifying Dynamic and Mobile Software Architectures", *ACM Software Engineering Notes*, Volume 29, Number 4, 2004.

[3] Ould M., *Business Processes: Modelling and Analysis for Re-engineering and Improvement*, ISBN 0-471-95352-0. 1995.

[4] Suleman H., *Open Digital Libraries*, PhD thesis submitted to Blacksburg, Virginia Tech, 2002.

[5] Frommholz I. et. al., "Supporting Information Access in Next Generation Digital Library Architectures", *Digital Library Architectures: Proc. of the Sixth Thematic Workshop of the EU Network of Excellence DELOS*, p.p. 49 – 60, 2004

[6] Marcos A. G. et. al., "Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries", *ACM Transactions on Information Systems*, Vol. 22, No. 2, p.p. 270 – 312, 2004.

[7] Candela L. et. al., "Moving Digital Library Service Systems to the Grid", *Proceedings of the Sixth Thematic Workshop of the EU Network of Excellence DELOS*, p.p. 13-24, 2004.

[8] Marchese M. et. al., "Towards a Distributed Architecture for Value Added Services to Digital Libraries", *The Second International Workshop on Peer-to-Peer Knowledge Management*, USA, 2005.

[9] Bjork B., et. al., "A formalised model of the scientific publication process", *Emerald, Online Information Review*, Volume 28, Number 1, p.p. 8 – 21, 2004.

[10] Nurnberg J. P. et. al., "Digital Libraries: Issues and Architectures", *DL '95 Proceedings*. pp. 147 – 153, 1995.

[11] Editor's Forum, Stages for Publication, *Editorial Service Guide, Site URL: [http://www.editorsforum.org/what_do_sub_pages/stages_publishing.html* Last accessed: August 10 2005].

[12] Green S. et. al., "The Primacy of Process Architecture", *5th Int. workshop on BPMDS*, Riga, Latvia, 2004.

[13] Ould M., Business Process Management: A Rigorous Approach, ISBN # 1-902505-60-3, 2005

[14] Marko A. R., "The Convergence of the Digital-Libraries and the Peer-Review Process", *ECCO Working Paper,* 2005.

[15] Eugene G., The use of *JCR* and *JPI* in Measuring Short and Long Term Journal Impact, Last accessed URL: http://www.garfield.library.upenn.edu/papers/cseimpactfactor05092000.html

[16] Paul W. Hamelman, Edward M. Mazze, "Cross-Referencing between AMA Journals and Other Publications", *Journal of Marketing Research*, Vol. 10, No. 2, pp. 215-218, 1973.

[17] Crouch S., Process Modelling for Requirements Capture, *PhD Thesis*, University of Southampton, UK, 2000.

[18] Lei Yu, Agent Oriented and Role Based Business Process Management for Computational Media, *PhD Thesis*, University of Zurich, 2000.

[19] Patel N. V, "Healthcare Modelling through Role Activity Diagrams for Process-Based Information Systems Development", *Requirement Eng 5: p.p. 83-92, Springer – Verlag*, London Limited, 2000.

[20] Odeh M., et. al., "Bridging the Gap between Business Models and System Models", *Information and Software Technology, Special Edition on Modelling Organisational Processes*, Vol. 45, Issue 15, pp 1053-1060, 2003.

[21] Melao N., et. al., "A conceptual framework for understanding business processes and business process modelling", *Information Systems Journal* 10, pp. 105-129, 2000

[22] Fytton Rowland, "The Peer Review Process", *A report to JISC Scholarly Communication Group,* 2002.

[23] Kurtev I. et. al, "Technological Spaces: an Initial Appraisal", *CoopIS, DOA'*2002

[24] Khan Z. A, et. al., "Digital Libraries: From Process Modelling to Grid-based Service Oriented Architecture", ICTTA 2006, Damascus, April 2006.