

10291 Abstracts Collection
Automation in Digital Preservation
— Dagstuhl Seminar —

Jean-Pierre Chanod¹, Milena Dobreva², Andreas Rauber³, Seamus Ross⁴ and
Vittore Casarosa⁵

¹ Xerox Research Center Europe - Grenoble, FR

`Jean-Pierre.Chanod@xrce.xerox.com`

² University of Strathclyde - Glasgow, GB

`dobрева@math.bas.bg`

³ TU Wien, AT

`rauber@ifs.tuwien.ac.at`

⁴ University of Toronto, CA

⁵ ISTI-CNR, Pisa, IT

`casarosa@isti.cnr.it`

Abstract. Digital Preservation has evolved into a specialized, interdisciplinary research discipline of its own, seeing significant increases in terms of research capacity, results, but also challenges. However, with this specialization and subsequent formation of a dedicated subgroup of researchers active in this field, limitations of the challenges addressed can be observed. Digital preservation research may seem to react to problems arising, fixing problems that exist now, rather than proactively researching new solutions that may be applicable only after a few years of maturing.

Recognising the benefits of bringing together researchers and practitioners with various professional backgrounds related to digital preservation, a seminar was organized in Schloss Dagstuhl, at the Leibniz Center for Informatics (18-23 July 2010), with the aim of addressing the current digital preservation challenges, with a specific focus on the automation aspects in this field. The main goal of the seminar was to outline some research challenges in digital preservation, providing a number of “research questions” that could be immediately tackled, e.g. in Doctoral Thesis. The seminar intended also to highlight the need for the digital preservation community to reach out to IT research and other research communities outside the immediate digital preservation domain, in order to jointly develop solutions.

Keywords. Interdisciplinary research in digital preservation, research challenges in digital preservation

Background Document

Digital Preservation (DP) has evolved into a specialized, interdisciplinary re-

2 Jean-Pierre Chanod, Milena Dobрева, Andreas Raube, Seamus Ross and
Vittore Casarosa

search discipline of its own, seeing significant increases in terms of research capacity, results, but also challenges. However, with this specialization and subsequent formation of a dedicated sub-group of researchers active in this field, limitations of the challenges addressed can be observed. DP research may seem to react to problems arising, fixing problems that exist now, rather than proactively researching new solutions that may be applicable only after a few years of maturing. Core IT know-how that is needed to tackle the significant problems that we are facing in DP is not sufficiently present within the DP research community as this is seen as a separate sub-discipline. The goal of the Dagstuhl Seminar is to outline some research challenges in DP, highlighting the need for the DP community to reach out to IT research and other research communities outside the immediate DP domain in general to jointly develop solutions.

Keywords: Interdisciplinary research in digital preservation, research challenges

Joint work of: Rauber, Andreas

Introduction - Research Agendas in Digital Preservation

As a starting point for the discussions during the seminar, the major projects in digital preservation, funded by the European Union under the 6th Framework Program (DPE, PLANETS, CASPAR), were presented here.

Keywords: Preservation research agendas, DPE, PLANETS, CASPAR

Joint work of: Casarosa, Vittore

Breakout Session: Preservation Ready Systems – Digital Preservation and Enterprise Architecture

If we consider an Information System, we can see that in general issues of digital preservation have been considered "after the fact", mainly as a way to preserve the output of the system. At the same time, digital preservation research has focused mainly on the development of Digital Preservation Systems (like OAIS), and on scenarios where Digital Preservation Systems are expected to interoperate with "external" Information Systems.

In many scenarios however, it would be desirable to identify digital preservation requirements from the beginning and to embed the implementation of those requirements into the system itself, making it a "Digital Preservation Ready" system. The research challenges related to this topic should become part of the concerns of the Enterprise Architecture, contributing to the building of systems that in addition to being reliable, usable, interoperable, etc., are also "preservable" or, in other words, systems that are resilient against changes in technology.

Keywords: Preservation ready systems, system engineering, enterprise architecture

Joint work of: Borbinha, Jose

Breakout Session: Beyond Metadata? Information Retrieval, Mining, Visualization, Context

In the breakout sessions, we dealt with four main areas of digital preservation, namely what to preserve, the nature of Digital Objects, the conceptual modeling of Digital Objects and Digital Preservation, and finally, techniques applicable to Digital Preservation. We identified three recurring challenges across the discussed areas of Digital Preservation. These challenges relate to the fluidity of preserved information, the preservation of context, and scalability issues in Digital Preservation. In the remainder of this draft document, we provide some initial insight on the identified challenges with respect to the discussed topics.

Keywords: Preservation of context, scalability issues in preservation, fluidity of preserved information

Joint work of: Constantopoulos, Panos; Plachouras, Vassilis

Breakout Session: Storage Technologies and Protocols

Broadening in some way the scope of preservation, several main topics of interest can be identified: (i) self sufficiency; (ii) "imperfect" digital preservation; (iii) information handling; (iv) storage technologies; (v) distributed Systems and protocols; (vi) new frontiers. Some slides were used to introduce those topics, and this paper briefly summarizes the outcome of discussions.

Keywords: Preservation self sufficiency, imperfect digital preservation, deletion of preserved data, storage technologies, quantum computing

Joint work of: Gschwind, Rudolf; Plachouras, Vassilis

Breakout Session: Policy and Rule Management

It is the current state of the art to explicitly formulate policies and rules for digital preservation and express them separately from the underlying system in order to make them more manageable, adaptive and auditable. They describe how digital preservation is ensured by an institution and digital preservation systems must enforce them. Digital preservation policies and rules also provide important contextual information for evaluating the authenticity of preserved objects and for communicating service expectations and requirements between stakeholders. The slides introduce the main ideas for policy and rule management in digital preservation, and this paper briefly summarizes the discussions held in the breakout session.

4 Jean-Pierre Chanod, Milena Dobрева, Andreas Raube, Seamus Ross and
Vittore Casarosa

Keywords: Digital preservation, policies in digital preservation, rules in digital
preservation

Joint work of: Jens, Ludwig; Lee, Cal

Breakout Session: Ethics, Privacy, Security and Trust

The topics of ethics and privacy are closely connected with security and trust, and their interplay provides a number of interesting research challenges and questions. While these will need solutions to be connected to ethical and legal frameworks, the need was felt to understand the technical implications and possibilities in order to provide guidance on the risks as well as potential mitigation strategies from a technical perspective.

Keywords: Ethics in preservation, privacy in preservation, security in preservation, trust in preservation

Joint work of: Lee, Cal; Kim, Yunhyong

Breakout session: Evaluation and Benchmarking in Digital Preservation

Evaluation and benchmarking is a crucial aspect in digital preservation, considering that in practical terms there can be no quality assurance without evaluation, no marketplace without comparability and no competition without a metric. A pragmatically viable approach is to define the main aspects that need to be measured and address them on a prioritisation basis.

Keywords: Evaluation of digital preservation, benchmarking in preservation

Joint work of: Becker, Christoph

Breakout Session: Application domains

Computer games can represent the "exemplar application" overarching all the other application domains. We might say (with some optimism) that if we can preserve computer games, then we can preserve everything digital. They present a number of technical and legal challenges.

Keywords: Preservation of computer games, preservation of context, preservation of user generated data

Joint work of: Lange, Andreas

Summary reports of the Breakout Sessions

This report gathers the final (draft) reports of each breakout session, with a brief summary of the main points discussed and some suggestions of possible (challenging) research topics.

Keywords: Automation in preservation, preservation ready systems, beyond metadata in preservation, storage technologies and protocols for preservation, policy and rules in preservation, ethics and trust in preservation, evaluation and benchmarking in preservation, preservation of computer games

Joint work of: Casarosa, Vittore

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2902>

Long-term digital preservation in e-Science domains

Jose Barateiro (Polytechnic Institute of Lisbon/INESC-ID - Lisboa, PT)

The complexity of digital preservation increases with the fact that each type of digital object has its own particularities and special requirements. The collaborative environment of the scientific community, and associated services and infrastructures, usually known as e-Science (or enhanced Science), involves the requirement of interoperability and the respective data sharing.

In a broad sense, e-Science concerns the set of techniques, services, personnel and organizations involved in collaborative and networked science. It includes technology but also human social structures and new large scale processes of making science. It also means, on the same time, a need and an opportunity for a better integration between science and engineering processes. Thus, long term preservation can be thought as a required property for future science and engineering, to assure communication over time, so that information that is understood today is transmitted to an unknown system in the future.

Keywords: Preservation, e-Science, risk management

Extended Abstract: <http://drops.dagstuhl.de/opus/volltexte/2010/2769>

Challenges in preservation (planning)

Christoph Becker (TU Wien, AT)

This short paper attempts to highlight some challenges to be tackled by DP research in the next years, taking as a starting point the perspective of preservation planning.

These challenges are in short: (1) Scalability (up and down) requiring (2) measurement of relevant decision factors, in turn requiring (3) benchmarking and ground truth. (4) Quality-aware emulation. (5) Move from the current closed-systems approach to open structures that accommodate evolving knowledge. (6) Move from post-obsolescence actions to 'longevity engineering'.

6 Jean-Pierre Chanod, Milena Dobрева, Andreas Raube, Seamus Ross and Vittore Casarosa

Keywords: Preservation planning, software engineering, scalability, measurements, benchmarking, ground truth, longevity

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2768>

Some (provocative) considerations about preservation, systems and perfect worlds (using a lot of quotes)

Jose Borbinha (Polytechnic Institute of Lisbon/INESC-ID - Lisboa, PT)

My position is that maybe we should not continue research in order to build dedicated digital preservation systems, but instead we should focus our efforts on understanding which should be the basic principles under which systems (simply that) could be designed in order to have, naturally, digital preservation properties.

Keywords: Ubiquitous digital preservation, preservation as a property of systems

Extended Abstract: <http://drops.dagstuhl.de/opus/volltexte/2010/2897>

Preserving the Intent behind

Jean-Pierre Chanod (Xerox Research Center Europe - Grenoble, FR)

Considering that in many areas, especially outside cultural heritage, preserving the intent behind activities or projects is at least as important as preserving the associated digital data, we argue that project policies are the natural means to describe an intent in a sustainable implementation-independent way, after the project actors have reached an agreement.

Keywords: Digital preservation, policy, intent preservation, consensus

Joint work of: Jaquin, Thierry; Déjean, Hervé; Chanod, Jean-Pierre

Extended Abstract: <http://drops.dagstuhl.de/opus/volltexte/2010/2770>

Preserving activity context

Panos Constantopoulos (Athens University of Economics and Business, GR)

Preservation of digital resources should include the ability to re-enact current usage practices and the ability to use the resources in new ways.

Prerequisite to the first is an understanding of the original intended use, of the activities within which the resources are being used and of the usage processes, whereas the second can also benefit from the knowledge of usage histories and usage-related features. Capturing, representing and maintaining the above knowledge is part of a broader set of actions related to the context in which the curation and preservation of digital resources take place. These include the modeling of goals and usage, domain models (ontologies and application models), authorities and the management of all these as a distinct set of resources, in addition to the original collection of digital resources. These complementary resources constitute a representation of the context in which the resources are interpreted and used.

We focus here on the modeling of activities, in particular scholarly activities, which process and produce potentially preservable digital resources. The activity model provides a detailed conceptualization of entities involved in scholarly information practice, clarifying the relationship between information activities or processes in which researchers engage in practice, normative procedures, methods, tools etc.

Keywords: Context preservation, modelling of scholarly activities

Beyond metadata ? Information Retrieval/Mining/Visualization/Context

Panos Constantopoulos (Athens University of Economics and Business, GR)

In this presentation we discuss a number of examples and issues related to the curation of digital objects, whose solution in the end should lead us to the preservation of digital resources, rather than digital objects. In the future we should be able not only to access those digital resources, but we should also be able to understand and use them in different ways, and to re-interpret them and generate new knowledge.

Keywords: Preservation of digital resources, Curation, Information retrieval, query patterns, context

Digital Preservation: Evaluation Issues

Milena Dobrevá (The University of Strathclyde - Glasgow, GB)

The aim of this presentation is to start a discussion on topics related to digital libraries evaluation, in order to address the user needs better. A number of questions can be raised here. What methods should be used for the evaluation of digital preservation systems? How can we compare different systems from a user-centric point of view? How can we show to a user that a particular tool really can be trusted on the long term? How can we demonstrate OAIS compliance? The

starting point will be to outline how users are addressed in a number of models of digital libraries (DELOS Digital Library Reference Model, 5S, 5M) and in the international standard ISO 9241-210:2010(E); examples of typical user studies will also be presented.

Keywords: Evaluation, qualitative, quantitative methods, evidence

Core challenges for digital preservation from a records management/archival perspective

Fiorella Foscarini (University of Toronto, CA)

This abstract briefly outlines two of the core challenges that should be included in a digital preservation (DP) research agenda. The first, more general challenge refers to the lack of a consistent theoretical framework that would help coordinate the various DP initiatives in a coherent and solid conceptual infrastructure. The archival discipline is expected to play a major role in the definition of this framework. The second issue here introduced refers to our limited understanding of the context of records creation and use. This abstract makes a case for the use of bottom-up, 'softer' approaches to the study of real-world situations.

Keywords: Theoretical framework, archival discipline, records creation context, soft methodology

Collecting Usage Data for Digital Preservation

Muriel Foulonneau (Henri Tudor Research Center, LU)

While IT environments are moving towards personalized and context-aware adaptive content and services, digital preservation systems should go beyond the current mechanisms to preserve digital objects. Social and personal experiences need to be investigated as part of the context of digital resources, i.e., the way in which a resource was used and perceived, by retaining usage data for instance. Overall, users have to be further involved in the digital preservation processes, in the creation of context metadata, in the storage and migration of resources, in particular for personal archives.

Keywords: Perception, personalization, multimedia, digital preservation

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2766>

PEVIAR (Permanent Visual Archive)

Rudolf Gschwind (Universität Basel, CH)

Even decades after the advent of computer technology, preserving information in digital archives remains a challenging task. Technological progress on all levels calls for migration as an essential component of any digital archive.

Migration, however, is not only expensive, but also makes it difficult to ensure the authenticity of digital documents. The Imaging and Media Laboratory at the University of Basel has developed PEVIAR (Permanent Visual Archive), an archiving solution addressing one of the fundamental challenges of digital archiving, migration, and introducing the notion of digital originals. Essentially, digital documents are inseparably bound to a migration-free medium, thereby overcoming the transient nature of present state-of-the-art storage systems.

Keywords: Digital preservation, long-term archiving, document authenticity, data migration, visual data representation

See also: Müller, F., Fornaro, P., Rosenthaler, L., and Gschwind, R. 2010. PEVIAR: Digital originals. *ACM J. Comput. Cult. Herit.* 3, 1, Article 2 (June 2010), 12 pages. DOI = 10.1145/1805961.1805963 <http://doi.acm.org/10.1145/1805961.1805963>

Research challenges in Digital Preservation

Michael Hartle (TU Darmstadt, DE)

The extended abstract contains a bit about my background and expertise in Data Format Description, lists a number of long-term visions such as the computer-aided construction and reverse engineering of data format knowledge with related questions concerning future research, and closes with my goals and expectations.

Keywords: Data Format Description, Reverse Engineering

Data, Information, and Knowledge: "where is the Life we have lost in living?"

Yunhyong Kim (The Robert Gordon University - Aberdeen, GB)

This abstract attempts to raise the question of whether current practices in digital preservation properly address the issues of findability of digital objects. It is also intended as a starting point for discussing preservation of digital information in contrast to digital data. The abstract is exploratory and informal.

Keywords: Data, information, knowledge, wisdom, preservation, appraisal, selection, findability

Extended Abstract: <http://drops.dagstuhl.de/opus/volltexte/2010/2765>

Factsheet European Research Project KEEP

Andreas Lange (Computerspiele Museum (im fjs e.V.) - Berlin, DE)

KEEP (Keeping Emulation Environments Portable) is a mid-sized research project funded by the European Union, which focuses on emulators as preservation tools. It is running from February 2009 until Jan 2012. It will develop an Emulation Access Platform to enable accurate rendering of both static and dynamic digital objects: text, sound, and image files; multimedia documents, websites, databases, videogames etc.

KEEP will address the problems of transferring digital objects stored on outdated computer media such as floppy discs onto current storage devices. This will involve the specification of file formats and the production of transfer tools exploited within a framework, and taking into account possible legal and technical issues. KEEP will address all aspects ranging from safeguarding the original bits from the carrier to offering online services to end-users via a highly portable emulation framework running on any possible device.

The project also aims to deliver an understanding about how to integrate emulation solutions with an operational electronic deposit system. Existing metadata models will be researched and guidelines will be developed for mapping digital objects to emulated manifestations.

Keywords: Research Emulators Automation Preservation

Joint work of: Lange, Andreas and others

Computer-Supported Elicitation of Curatorial Intent

Cal Lee (University of North Carolina- Chapel Hill, US)

Individuals across society are generating digital traces of their lives; and these traces can be stored, processed, combined and reused in numerous ways. Many cultural institutions – including libraries, archives and museums (LAMs) – have collecting missions that include personal papers, manuscripts and other non-institutional materials.

Despite a massive increase in the volume and complexity of personal digital collections, research and literature designed specifically to guide this activity has been relatively limited. Information professionals will need to have tools and approaches to better acquire digital collections from individuals. Just as importantly, they must care for collections in ways that reflect the values of the relevant stakeholders, i.e. 'the characteristics or principles by which [they] consider something desirable or worthwhile'.

In this document, I propose a program of research to investigate and address the above issues. I present a case for the development and testing of computer-supported mechanisms to elicit the curatorial intent of individuals in relation to digital objects being transferred to repositories.

Keywords: Personal archives, ethics, values, curatorial intent

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2761>

About the Complexity of a Digital Preservation Theory and Different Types of Complex Digital Objects

Jens Ludwig (Universitätsbibliothek Göttingen, DE)

In this extended abstract I will not present the results or plans of specific projects but some experimental and preliminary thoughts about the theoretical foundation and terminology of digital preservation. For digital preservation a theory in a strong sense may not be possible for task inherent reasons. Nevertheless it is reasonable to refine the underlying models and terminology and different aspects and complexity types of digital objects are distinguished.

Keywords: Digital preservation theory, digital objects, complexity

Aspects of Automatic Content Analysis in the Area of Digital Preservation

Alexander Mehler (Universität Frankfurt am Main, DE)

The content analysis of digital resources is connected with a variety of difficult subtasks. Amongst others, these subtasks have to do with relating a resource and its constituents with other digital artifacts that help to interpret or to further process the original resource. In this sense, automatic content analysis of digital resources goes hand in hand with linking these resources into networks of semiotic artifacts. In this paper we will use this insight to better understand different approaches to automatic content analysis from an abstract point of view.

Keywords: Automatic content analysis, contextualization of digital resources

On the Notion of Genre in Digital Preservation

Alexander Mehler (Universität Frankfurt am Main, DE)

In this paper, we discuss the notion of genre as a basis for addressing the problem of context representation in digital preservation. We outline several reference points for the notion of genre. This includes a review of diplomatic principles that can support and enhance the power of genre as a key to capture information about context relations. Further, we discuss the impact of open genre models and open topic models in information retrieval and finally present a list of research questions concerning future research in automation of digital preservation.

Keywords: Digital preservation, genre analysis, context modeling, diplomatics,
information retrieval

Joint work of: Fiorella Foscarini, Yunhyong Kim, Christopher A. Lee, Alexander
Mehler, Gillian Oliver, and Seamus Ross

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2763>

Characterization and Analysis of Digital Objects

Natasa Milic-Frayling (Microsoft Research UK - Cambridge, GB)

Whether we are concerned with the interoperability of data or long term preservation of digital content, we are often confronted with the issue of migrating digital object formats and assessing the quality of the transformation done on the digital object. This leads to the fundamental question of how to identify and describe the characteristics of the digital objects, which are usually related to the characteristics of the software applications that created them or the "content players" that allow their consumption. The ability and relevance of having a complete characterization of digital objects depends on the context.

We present here some results of experiments with office documents and simple content properties such as layout. For a document file that has been transformed to another format, we can use screen rendering as the action through which the object is consumed. Thus it is of interest to characterize the viewing experience and provide measures for visual characteristics related to the appearance and interaction with the rendered content. We devised a technique that captures the screen images of two files and thus creates a common representation that can be used for comparison. Using the OCR software as the common tool to investigate properties of the image representation, we extract features and compare. The example illustrates the complexity involved in devising practical tools/devices for characterizing and measuring properties of digital objects.

Keywords: Characterization, analysis, digital objects

Digital Object Characterization: Document Conversion and Quality Assurance

Natasa Milic-Frayling (Microsoft Research UK - Cambridge, GB)

Whether we are migrating document formats to achieve interoperability or ensure long term preservation, we are faced with the issue of assessing the quality of the digital object transformation. However, comparing two digital objects is not straightforward. It raises the issue of properties that are inherent to the digital objects and those that are dependent on the environment in which the objects are created, viewed, and compared to one another. That has implications for

devising methods to extract document properties, interpret observed characteristics, and apply similarity metrics. Furthermore, in order to take actions based on collected measurements, we need to define or learn the significance of individual document properties from the perspective of human perception and usage scenarios. We illustrate the complexity of these issues by presenting a method for comparing converted office documents and discussing the challenges from the technical and methodology point of view.

Keywords: Characterization, quality assurance, format migration, file conversion

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2901>

Digital Archives and Documentary Genre

Gillian Oliver (Victoria University of Wellington, NZ)

This abstract briefly outlines two distinct areas for research, which are core challenges for archivists. The first area is related to the appropriateness and relevance of current digital archive models, which are presently largely based on the transfer of custody, where the owner or creator of the digital objects transfer the responsibility of long term preservation to the archive. Anecdotal evidence from archives institutions in Australia and Europe indicate however that these archives, however well designed, are only partially successful as expected transfers or deposits are not materializing.

The second area is related to the exploration of the genre concept, which has the potential to enrich traditional archival functions of appraisal, description and access. The challenge is to explore how the genre concept can be applied without loss of context.

Keywords: Preservation, digital archives models, documente genre

Integrating Information Retrieval models and Digital Preservation

Vassilis Plachouras (Athens University of Economics and Business, GR)

While Information Retrieval provides a mechanism for accessing digital archives, the retrieval models do not necessarily incorporate concepts from the area of Digital Preservation, which potentially result in enhanced access to the archived records, as well as improved support of decisions taken by archivists. Related to the appraisal of archived records, one challenge is to develop methods for the prediction of the record's importance in the future, given the information about its archived versions. Regarding the concept of forgetting, it may be integrated to the information retrieval model in order to manage the complexity of the

archived record representation. For example, the size of the search engine's index may be adjusted by pruning the entries marked to forget.

As a particular example to examine the above issues, we focus on Web archiving, where the collected data corresponds to a set of versioned records

Keywords: Preservation, Information Retrieval, Web archiving

Research questions in Cultural Heritage digital repositories

Peter Stanchev (Kettering University - Flint, US)

This discussion is about innovative solutions for assembling multimedia digital repositories for collaborative use in specific contexts and communities and enhancing scholarly understanding and experiences of digital cultural heritage. Several aspects are stress such as the dynamic aggregation of cross-media resources across existing institutional digital libraries and repositories. Research questions about the scalability, interoperability and distributed architectures, aggregation, and semantic search tools are addressed.

Keywords: Culture Heritage, Digital Repositories

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2762>

See also: www.kettering.edu/~pstanche

(Semi-)Automated digital preservation archives for small institutions and private users

Stephan Strodl (TU Wien, AT)

An increasing quantity of digital collections is held by small institution with limited know-how and awareness of digital preservation. Digital assets are becoming more important for an increasing number of institutions in the long run (e.g. legal obligation, intellectual property or business data). The limited resource in these institutions for archiving drives the need for new approaches of (fully or semi)-automated archiving systems. Research and development in the area of digital preservation is mainly done by memory institutions and large businesses. Consequently, the available tools, services and models are developed to meet the demands of professional environments.

Automated archiving systems are needed for institutions with little professional know how in digital preservation. Important aspects are hiding the complexity of the processes, providing support for decision making and automated error handling. The automation of preservation work flows raises a number of research questions, e.g. meta data management, quality assurance and tolerable limit of loss of preservation actions and automated preservation planning.

Keywords: Digital Preservation, Automation, Small institution

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2764>

Abstract Architecture-independent Workflow Automation

Dirk von Suchodoletz (Universität Freiburg, DE)

The creation of most digital objects occurs solely in interactive graphical user interfaces which were available at the particular time period. Archiving and reservation organizations are posed with large amounts of such objects of various types. At some point they will need to automatically process these to make them available to their users or convert them to a commonly used format. We present methods and a system architecture for emulation services which enables the preservation of interactive environments and their workflows in a reliable manner. This system includes a framework for describing interactions with an interactive environment in an abstract manner, supporting reliable playback in an automated way and finally ensuring the preservation of specific operation knowledge by documenting and storing all components in a dedicated software archive.

Keywords: Digital preservation, workflow automation, emulation, predictable emulator, API, software archive

A Future Emulation and Automation Research Agenda

Dirk von Suchodoletz (Universität Freiburg, DE)

After significant research and proven usefulness especially for complex, dynamic and interactive objects emulation is not widely adapted in digital preservation at all. While some significant building blocks of emulation based strategies are present a number of components are still unsatisfactory or missing. This paper proposes a research agenda for the future integration of emulation into preservation workflows. It discusses prerequisites and requirements for fully automated services operating in large scale environments. Those include the replacement of user interaction by using a standard interfacing protocol like Virtual Network Computing, proper system image and software components archiving and the "preservation aware" emulator. To achieve especially the latter additional goal channels to control the emulator and monitor its states are required. This paper analyses the state of the art in emulation and motivates the need for an advanced research agenda in this field.

Keywords: Digital preservation, workflow automation, emulation, predictable emulator, API, software archive

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2010/2771>