

1-10-2010

# Digital content preservation across domain verticals

Soha Maad

*Royal College of Surgeons in Ireland, sm@sohamaad.net*

Borislav D. Dimitrov

*Royal College of Surgeons in Ireland, borislav.d.dimitrov@gmail.com*

---

## Citation

Maad S, Dimitrov BD. Digital content preservation across domain verticals. *Journal of Communication & Computer*. 2010; 7(10): 64-71.

This Article is brought to you for free and open access by the Department of General Practice at e-publications@RCSI. It has been accepted for inclusion in General Practice Articles by an authorized administrator of e-publications@RCSI. For more information, please contact [epubs@rcsi.ie](mailto:epubs@rcsi.ie).

**Attribution-Non-Commercial-ShareAlike 1.0**

**You are free:**

- to copy, distribute, display, and perform the work.
- to make derivative works.

**Under the following conditions:**

- Attribution — You must give the original author credit.
- Non-Commercial — You may not use this work for commercial purposes.
- Share Alike — If you alter, transform, or build upon this work, you may distribute the resulting work only under a licence identical to this one.

For any reuse or distribution, you must make clear to others the licence terms of this work. Any of these conditions can be waived if you get permission from the author.

Your fair use and other rights are in no way affected by the above.

---

This work is licenced under the Creative Commons Attribution-Non-Commercial-ShareAlike License. To view a copy of this licence, visit:

**URL (human-readable summary):**

- <http://creativecommons.org/licenses/by-nc-sa/1.0/>

**URL (legal code):**

- <http://creativecommons.org/worldwide/uk/translated-license>
-

# Digital Content Preservation across Domain Verticals

Soha Maad and Borislav D. Dimitrov

*HRB Centre for Primary Care Research, Royal College of Surgeons in Ireland, Dublin 2, Ireland*

Received: June 03, 2010 / Accepted: June 29, 2010 / Published: October 25, 2010.

**Abstract:** The authors present a novel approach to develop scalable systems and services for preserving digital content generated from various application domains. The aim is to deliver an integrative scalable approach for digital content preservation across domain verticals. This would involve consolidating approaches for modeling document workflow, preserving the integrity of heterogeneous data, and developing robust and scalable tools for digital preservation ensuring interoperability across domains verticals. The authors consider various application domains including: healthcare, public, business and finance, media and performing art, and education. The authors focus on specific case studies of digital content preservation across the considered domain verticals. The authors describe an integrative framework for digital content preservation across domain verticals. This framework is developed at four levels and attempts to abstract and integrate the digital content workflow across domain verticals. The authors suggest a test bed to validate our integrative approach for digital content preservation. This integrates the digital content preservation activity along the value chain of domain verticals.

**Key words:** Digital preservation, application domains, scalable systems and services.

## 1. Introduction

Digital Preservation encompasses the activities that try to ensure that digital objects remain accessible and usable in an authentic way for long periods of time. Long term can mean anything from a few years to decades and ultimately centuries: anything that is long enough to experience technological change that threatens the availability of a digital object. The last few years witnessed drastic changes in the field of digital preservation. The topic has moved from niche specialist discussions into broad mainstream awareness and more solutions are becoming available and being deployed. This may be attributed to the provision of significant R&D funds for this massive challenge of keeping digital objects accessible and usable [1].

One irony of the information age is that keeping information has become more complex than it was in

the past. The authors not only have to save physical media and electronic files, we also need to make sure that they remain compatible with the hardware and software of the future [2].

With this background, the authors present a novel approach to develop scalable systems and services for preserving digital content generated from various application domains. The aim is to deliver an integrative scalable approach for digital content preservation across domain verticals. The authors intend to deliver progress beyond state of the art in:

- Modeling digital document workflow;
- Preserving the integrity of heterogeneous data;
- Developing robust and scalable tools for digital preservation;
- Ensuring interoperability across domains;
- Academic and industry wide result exploitation;
- Development of testbed and demonstrator.

The authors consider specific case studies of digital content preservation across domain verticals. The case studies consider the existing digital content workflow for a particular application in the considered domain.

---

**Corresponding author:** Soha Maad, Ph.D., associate member, research fields: converged ICT and media. E-mail: sm@sohamaad.net.

Borislav D. Dimitrov, DM/Ph.D., senior research fellow, research fields: primary care, ICT in healthcare, eHealth. E-mail: borislav.dimitrov@rcsi.ie.

The authors propose scalable and interoperable approach to improve the digital content workflow in the various domains. Integrity, robustness, scalability, and interoperability of the digital content preservation are key consideration for novel digital content workflow across domain verticals.

The paper is composed of five sections including this one. In the following second section the authors present a brief overview of state of the art in digital preservation in the European arena. In section three, the authors consider case studies across various domain verticals. In the fourth section the authors present an integrative scalable approach for digital content preservation across domain verticals. The fifth section concludes with agenda for exploitation plan and development of test bed.

## 2. State of the Art

In this section the authors overview the state of the art in digital preservation in the European arena. This overview considers projects funded within European Union Framework Program six and seven FP6 and FP7.

SHAMAN Sustaining Heritage Access through Multivalent Archiving, is an FP7 project intended to develop and test a next generation digital preservation framework including tools for analyzing, ingesting, managing, accessing and reusing information objects and data across libraries and archives. This involves the development of next generation digital preservation framework combining datagrid and semantic technologies, validation on different types of content and organizational needs [3].

PrestoPrime is an EU FP7 project for long-term preservation of digital audio-visual content. PrestoPRIME will research and develop practical solutions for the long-term preservation of digital media objects, programmes and collections, and find ways to increase access by integrating the media archives with European on-line digital libraries in a digital preservation framework. This will result in a

range of tools and services, delivered through a networked Competence Centre. The project will deliver a preservation framework, complete with risk management and content quality and corruption control measures, capable of supporting audiovisual signal migration and multivalent preservation methods using federated services for distributing and storing content. It will create a metadata conversion and deployment toolkit, with a novel and efficient process for metadata vocabulary alignment, annotation and services for user-generated content metadata. A rights management system and audiovisual fingerprint registry will make it possible to track and manage content at all stages of its lifecycle, in all contexts of use. The project will demonstrate and evaluate an integrated prototype of the preservation framework and software in the networked Competence Centre. The Competence Centre and the European Association for Audiovisual Archives will be established to provide business models, registry and best practice services and training [4].

CASPAR-Cultural, Artistic and Scientific Knowledge Preservation is an EU FP7 project for access and retrieval, preservation of digital content of various domains, including high volume and multimedia content. It is based on OAIS-compliant platform. CASPAR set out to build components for a preservation infrastructure which user communities can use to share the cost of preserving their digital resources and which advances the state of the art in digital preservation. This large-scale project addresses the growing challenge of preserving digital information, upon which society is increasingly dependent but which is intrinsically fragile. The CASPAR team will build a framework of tools and infrastructure components to support the end-to-end preservation of all types of digitally encoded information and thus help producers, curators and users of digital resources share the burden of preservation. To validate the research work, CASPAR is tested with different types of digital information, in a wide range of user communities: science, performing arts and cultural heritage. The test

beds produced will be embedded in operational systems within the CASPAR consortium, and will be easy to integrate into many other operational systems [5].

Planets, Preservation and Long-term Access through Networked Services, are a four-year project co-funded by the European Union under the Sixth Framework Program to address core digital preservation challenges. The primary goal for Planets is to build practical services and tools to help ensure long-term access to Europe digital cultural and scientific assets. The project develops preservation planning services that empower organizations to define, evaluate, and execute preservation and tools for transforming and emulating obsolete digital assets [6].

Keeping Emulation Environments Portable KEEP is a medium scale research project that develops portable emulators to enable accurate rendering of static and dynamic digital objects stored on outdated computer media, e.g., early computer games [7].

PROTAGE, PReservation Organizations using Tools in AGent Environments, develops agent-based technology and methodology for preservation-friendly content production, transfer and ingest, archival monitoring and access to information [8].

LiWA Living Web Archive is an FP7 project intended to develop web archiving tools to improve archive fidelity and authenticity and to ensure long term interpretability of web content. LiWA will extend the current state of the art and develop the next generation of Web content capture, preservation, analysis, and enrichment services to improve fidelity, coherence, and interpretability of web archives. By developing methods which improve archive fidelity, the project will contribute to adequate preservation of complete and high-quality content. By developing methods for improved archive coherence and interpretability, the project contributes to ensuring its long-term usability [9].

Digital Culture (DigiCULT) (digicult-web) is a Support Measure to establish a regular technology watch for cultural and scientific heritage. DigiCult

mission is benefiting the Cultural Heritage sector, through monitoring and assessing existing and emerging technologies that provide opportunities to optimize the development, access to, and preservation of Europe's rich cultural and scientific heritage, within the emerging digital cultural economy [10].

### 3. Digital Preservation Case Studies

In this section the authors consider digital preservation case studies across domain verticals. The authors suggest a typical digital content workflow and consider potential for scalability and interoperability. Where possible we highlight problems in existing workflow and suggest improvement using novel scalable systems and services for preserving digital content.

*In healthcare*, the authors address the problem of digital preservation to promote preventive healthcare and active life style. Our approach to develop scalable systems and services for preserving digital content aims at modeling preventive healthcare repositories promoting active life style. All healthcare and scientific authorities worldwide are realizing the importance of developing global healthcare infrastructures and access, at least at information and communication technology levels, for exchange of patient and healthcare data, services and provide wider opportunities for clinical research. In recent times, health policies have emphasized the importance of prevention, healthy living, more active lifestyles and community-based services and noted that these are highly influenced by globalization. Converged IT, communication and media industries can be deployed to develop preventive healthcare services at a global level. This would involve: meta-modeling of personalized patient-centred preventive healthcare services; supporting the collaborative authoring of interactive preventive healthcare digital content; developing novel paradigms of interaction among patients, global healthcare team, global healthcare network and the healthcare industry; developing a global healthcare enabling the delivery of

preventive healthcare services not only acting as a repository of info; developing design model for products and services for global health.

*In the public sector*, the authors address the problem along the workflow chain of policy generation and mediation. Policies are traditionally captured in rigid digital document formats. This poses challenges for scalable policy modeling and mediation where a participative policy modeling approach involving various target groups is to be adopted. Various approaches can be adopted for policy development these include:

(1) Analytical and political approaches taking into account the analytical and political dimension of the policy process and the importance of policy communication at an early stage of policy formulation. According to the analytical vision, a policy process can be seen as a cycle of subsequent phases in which problems are solved (agenda setting, policy formulation, policy adoption, policy implementation, policy evaluation) [11];

(2) Cognitive maps taking the form of a set of connected options-outcomes chains. Assertions about the world imply possible policy options which in turn imply strategies. These are often linked to the overarching goals taken to be purposes of the organization or decision maker [12];

(3) Computer assisted approaches to scenario discovery involving the support of policy-makers and analysts in identifying policy-relevant scenarios by interactively applying statistical and data mining algorithms to large databases of simulation-model results [13].

Our approach to develop scalable systems and services for preserving digital content aims at modular approaches for preserving public policies and enhancing the life cycle of policy generation and mediation. The design and structure of library of policies will build on existing work related to the development of policy-based pervasive service description language. We will also develop an access

model for the library of policy services. This will provide diverse pervasive means to mediate policies to various actors involved in the policy formulation, dissemination and implementation process.

*In the financial sector*, the authors address the problem of risk visibility and mediation. Financial risk captured in various documents is hard to uncover and mitigate using existing approaches for digital preservation of financial data and content. Our approach to develop scalable systems and services for preserving digital content aims at devising techniques for preserving financial content whilst providing the potential to extract knowledge from this content such as financial risk visibility. The urgent economic problem, linked to the financial crisis, challenges current research and technological development. The scale of the fiscal crisis that undermined the credibility of the financial system motivates the consideration of “Global Financial State Visibility” as a key global challenge that validates research and technological development activities to support the engineering dynamics of automatically adaptable software services along the “global financial supply chain”. Risk is an elusive term. Several attempts have been made to define the term risk. In general risk is a probability of a negative occurrence which may be neutralized through pre-mediated actions. Our aim is to mediate state of risk in order to reduce its negative impact, hence, the need to convey the state of risk as a service. Our aim is to align the prevalent thinking in terms of mediating risk using reports (static reports; or XBRL dynamic reports) to mediating the state of risk (financial risk visibility) as a service. Key issues to consider are: who will publish, deploy and discover the service and use it, and how the service will be tuned (i.e. automatic adaptation). Various entities (policy makers, regulators, auditors, accountants, investors, consumers, suppliers, producers, individuals) need to access /govern / adapt Financial Risk Visibility Service depending on Service Level Agreements SLA (see Fig. 2). Financial state could be conveyed in various ways: perception the state

of financial risk; perception of financial events; percent of the financial activity; perception of the financial system and regulatory framework. The Financial state visibility challenge has various dimensions:

(1) Domain knowledge dimension involving the building of the financial state knowledge base;

(2) Converged ICT and media dimension involving the development of an interoperability layer at the service infrastructure and interface levels to guarantee instant access to financial state via various converged ICT and media devices;

(3) Global factors dimension involving the adaptation of the financial state to “global context on demand”;

(4) Governance dimension involving the support of the governance of the financial state with greater perception and manipulation capability tools [14-16].

*In the business sector* registers stores huge amount of information about companies and businesses. With the economic slowdown various business content need to be added to the business register in order to increase the transparency of the performance of the business enterprise. Our approach to develop scalable systems and services for preserving digital content aims at devising techniques developing modular business register. Scalability is key issue and consideration.

Huge amount of knowledge is available in media and performing art content traditionally preserved in rigid format. We intend to develop modular techniques to preserve media and performance art content. This would facilitate the exploitation of the knowledge aspect of this content and its use in the development of product and services for the economy.

The exploitation of the wide variety of existing digital content is not yet fully exploited by academic programs in social sciences and humanities. Our approach to develop scalable systems and services for preserving digital content aims at supporting academic programs in social science and humanities with modular, distributed digital content preservation that can improve teaching programs in areas related to

political, historical, and European studies.

#### 4. Integrative Scalable Model for Digital Preservation

The integrative scalable model for digital preservation is developed at four levels [15]. Fig. 1 depicts the development levels of the integrative scalable model for digital preservation across domain verticals.

- Levels 1: this involves the deployment of state of the art digital preservation technologies delivered by projects reviewed in section two.
- Level 2: is the case studies level where the workflow of digital content across various domain verticals is considered.
- Level 3: is the integration level involving the development of an integrated model for scalable and robust and interoperable digital content preservation across domain verticals.
- Level 4: involves academia and industry wide exploitation plans and the development of prototype tested of the integrative scalable model.

Below is a brief elaboration of the activities undertaken at the four levels.

*At level 1*, various state of the art approaches and technologies are borrowed into the development of the integrative scalable model for digital preservation:

- Semantic technologies and tools for the validation of different types of content and organizational needs developed within the framework of SHAMAN [3];
- The tools and services of the networked Competence Centre developed within PrestoPrime [4];
- The components for the preservation infrastructure developed within the CASPAR [5];
- The preservation planning services developed by the Planet project to empower organizations to define, evaluate, and execute preservation and the tools for transforming and emulating obsolete digital assets [6];
- The tools for preservation-friendly content production developed by PROTAGE project [8];
- Tools for long term interpretability of web content

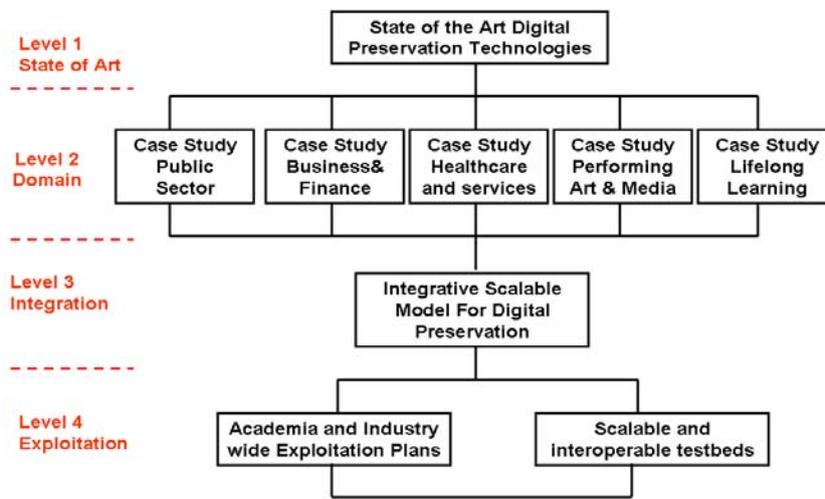


Fig. 1 Development levels of the integrative scalable model for digital preservation across domain verticals.

developed by LiWA project [9].

At level 2, the case studies described in section 3 are considered. For each case study, the authors focus on:

- Domain specific nature of the digital content;
- The digital content workflow;
- The deployment of state of the art infrastructures and interoperability layers identified at level 1;
- The global factor affecting the digital content workflow and preservation;
- The governance and access model of the digital content;
- The target end users and developers.

At level 3, various software services are developed taking into account the focuses of interest across domain verticals as identified at level 2. Fig. 2 illustrates the software services of the integrative scalable model for digital preservation across domain verticals. These services can be grouped into various levels and categories: [14, 16]

(1) Domain Levels Services

- Domain Services: Encapsulate, wrap, and render domain aspect of the digital content;
- Technology Convergence services: Accessibility and Interoperability Services;
- Global Factors Services: encapsulate, wrap, and render global factors affecting the digital content;
- Governance Services: encapsulate, wrap, and render governance & access models of digital content.

(2) Integration Services Exchange

Handle exchange communication between domain level services and integration framework services.

(3) Integration Framework Services

Integrate functionalities of domain level services: These consist of Integrative Framework Services and Integration Framework Service Exchange.

(4) Mediator Services

Mediate Integrative services to various users at various levels according to Service Governance Level Agreement. These consist of Mediator Service Exchange and Adaptation Service Mediator.

The design and engineering of the above described services involves: composition, workflow, adaptation at design time and run time consideration, agency interaction, verification and validation.

Users of the services are conceived at two levels: end users and developers. Users at both levels govern access, discovery, and manipulation of state visibility services according to Governance Service Level Agreements GSLA.

The aim is to align the software service development cycle to the counterpart real world processes and governance infrastructural set ups. Alignment of domain level services involves the alignment of composition, workflow, adaptation at design time and run time consideration, agency interaction, verification and validation to real process and governance development

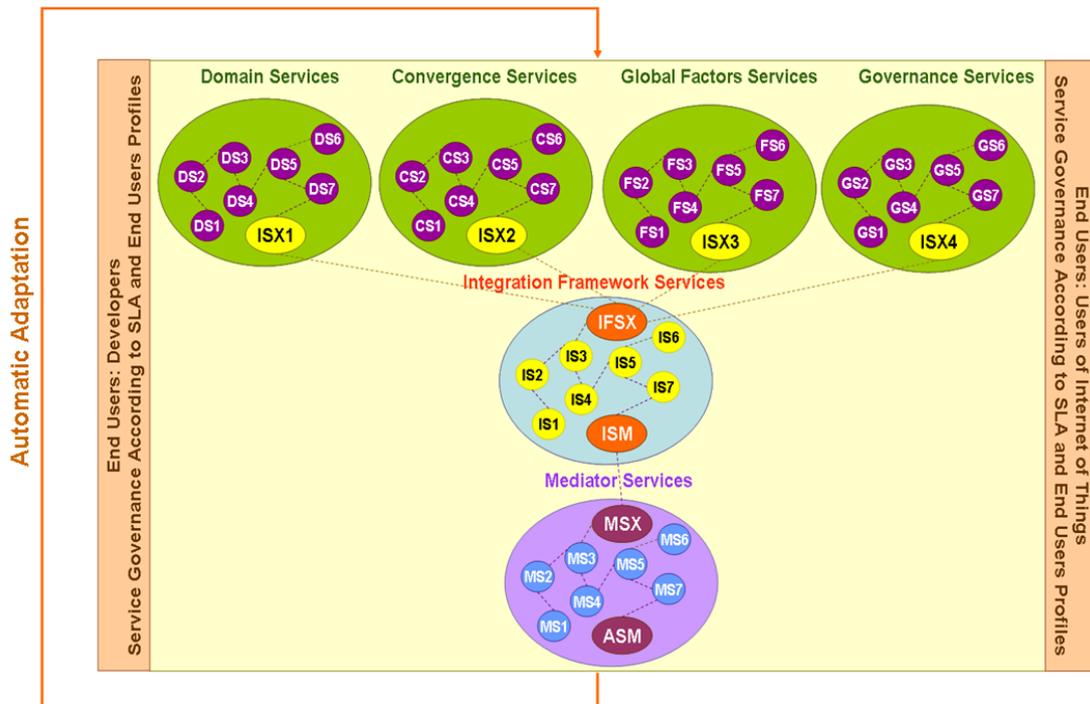


Fig. 2 Software services of the integrative scalable model for digital preservation across domain verticals.



Fig. 3 Value chain across domain verticals.

cycles. Alignment of services depends on the identified level and categories of services.

The alignment of domain level services involves the alignment of workflow and composition of services to the real world activities of knowledge resources: selection; collection; mash-up; extraction; modeling; and abstraction.

The alignment of mediators services involves the alignment of workflow and composition of services to the real world activities of gap measurement, dissemination; policy modeling, policy adjustment, and policy control.

This alignment depends on process management and infrastructural set up at end users points.

At level 4, a test bed demonstrator is thought. This assumes continuity of the activities across domain verticals, i.e., various domain verticals activities forms an integrated value chain. Fig. 3 depicts a typical value chain across domain verticals [17].

### 5. Conclusions

The authors presented an ongoing research work involving the development of novel approach to deliver scalable systems and services for preserving digital content across domain verticals. The digital preservation challenge was examined through various case studies drawn across domain verticals. The considered case studies addressed critical digital preservation challenges. In healthcare we addressed the challenge of modeling preventive healthcare repositories for promoting active life style. In the public domain we addressed the challenge of preserving public policies and enhancing the life cycle of policy generation and mediation. In finance the authors addressed the challenge of raising the visibility

of the financial risk. For businesses, the authors addressed the challenge of developing modular business registers providing greater transparency of the performance of the global business enterprise. In media and performing art, the authors addressed the challenge of knowledge extraction to develop better products and services for the economy. In education, the authors addressed the challenge of providing a greater outreach of the teaching content. For each case study, the authors considered the existing digital content workflow and suggested scalable and interoperable approach to improve the digital content workflow across domain verticals. An integrative framework for digital content preservation across domain verticals was described. This framework is developed at four levels and attempts to abstract and integrate the digital content workflow across domain verticals. Integrity, robustness, scalability, and interoperability of the digital content preservation were key consideration.

## References

- [1] I. Solvberg, A. Rauber, Digital preservation: introduction to the special theme, ERCIM News Special Theme: Digital Preservation, 2010.
- [2] M. Pat, Digital preservation: an evolving landscape, ERCIM News Special Theme: Digital Preservation, 2010.
- [3] SHAMAN-web, available online at: [http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-projects-shaman\\_en.html](http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-projects-shaman_en.html).
- [4] PrestoPrime-web, available online at: [http://cordis.europa.eu/fetch?CALLER=PROJ\\_ICT&ACTION=D&CAT=PROJ&RCN=89482](http://cordis.europa.eu/fetch?CALLER=PROJ_ICT&ACTION=D&CAT=PROJ&RCN=89482).
- [5] PrestoPrime-web, available online at: <http://cordis.europa.eu/ist/digicult/caspar.htm>.
- [6] PLANET-web, available online at: <http://www.planets-project.eu/>.
- [7] KEEP-web, available online at: <http://www.keep-project.eu/ezpub2/index.php>.
- [8] PROTAGE-web, available online at: <http://www.protage.eu/>.
- [9] Liwa-web, available online at: <http://www.liwa-project.eu>.
- [10] Digicult-web, available online at: <http://www.digicult.info/pages/index.php>.
- [11] K.U. Leuven, Minding the gap: applying a service marketing model into government policy communications, *Government Information Quarterly* 27 (2010) 34-40.
- [12] C. Eden, F. Ackermann, Cognitive mapping expert views for policy analysis in the public sector, *European Journal of Operational Research* 152 (2004) 615-630.
- [13] B.P. Bryant, R.J. Lempert, Thinking inside the box: a participatory, computer-assisted approach to scenario discovery, *Technological Forecasting & Social Change* 77 (2010) 34-49.
- [14] S. Maad, D.B. Dimitrov, T. Fahey, A concept for a long-term scalable bioengineering model of primary care, *Int. J. Value Chain Management* (in press).
- [15] S. Maad, *The Horizon of Virtual and Augmented Reality: The Reality of the Global Digital Age*, Intech, January 2010.
- [16] S. Maad, DIGIDO: scalable systems and services for preserving digital content from various application domains, *Proc. Transforming Government*, Brunel University, 2010.
- [17] S. Maad, J.B. McCarthy, S. Garbaya, M. Beynon, R. Nagarajan, Service software engineering for innovative infrastructure for global financial services, *Proc. European and Mediterranean Conference on Information Systems 2010*, Abu Dhabi, United Arab Emirates, 2010.