Semana de Engenharia 2010
Guimarães, 11 a 15 de Outubro

Relational Databases Digital Preservation
Work Progress Report
Significant Properties in the Preservation of Relational Databases

Ricardo André Pereira Freitas
Advisor: José Carlos Ramalho
Department of Informatics – University of Minho
Braga – Portugal
freitas@fam.ulusiada.pt, jcr@di.uminho.pt

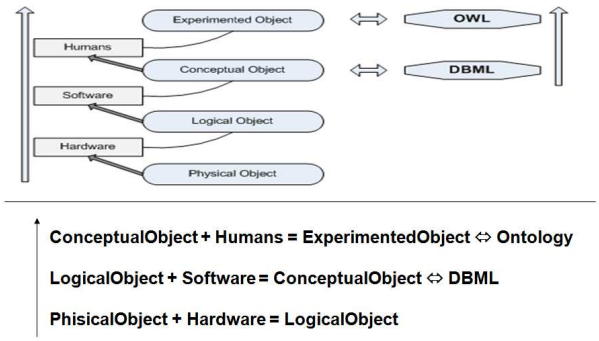
KEYWORDS
Digital Preservation, Significant Properties, Significant Characteristics, Relational Databases, Ontology, OAIS, XML, Digital Objects.

ABSTRACT

Relational Databases are the most frequent type of databases used by organizations worldwide and are the base of several information systems. As in all digital objects, and concerning the digital preservation of them, the significant properties (significant characteristics) must be defined so that adopted strategies are appropriate. In previous work a neutral format (hardware and software independent) --- DBML --- was adopted to achieve a standard format used in the digital preservation of the relational databases data and structure. Currently, in this PhD project we walk further in the definition of the significant properties by considering the database semantics as an important characteristic that should also be preserved. For the representation of this higher level of abstraction we are going to use an ontology based approach. We will extract the entity-relationship model from the DBML representation and we will represent it as an ontology.

PRESERVATION POLICY – CURRENT WORK

In order to walk further on the main topic of our PhD Project – *"Digital Preservation of Relational Databases"*, we intend to also walk further on the determination of the significant properties for this class of digital objects.
After characterizing the relational databases digital objects and establish a division between two levels of abstraction, we need to materialize those ideas into packages of information. These packages are to be used as in the OAIS reference model.
By focusing our strategy/policy on two levels of abstraction we intend to preserve the two correspondent levels of abstraction present on the chain of relationships of digital objects.
The database **Data** and **Structure**, which we identified as significant properties of the database, correspond to conceptual level of this family of digital object. The migration to DBML covers these properties and ensures that its representation becomes neutral.
At the top of the chain of relationships present in digital objects we have the Experimented Object (interpreted by humans). At this level there is an inherent **Knowledge** associated to the database semantics. We intend to captured the experimented object (knowledge) through an ontology based approach.



The ontology approach is adopted to formalize the knowledge present at the experimented object level and also a methodology to create an abstract representation of it.
The research work brought us to a point where we seek to preserve the combination of these levels of abstraction. The main strategy in our approach continues to be Migration which is carried in order to transform the original database into the new format -- DBML + Ontology.

CONCLUSION

At present time, in this PhD project, we address the problem of relational databases digital preservation by pointing at the significant properties of this class of digital objects. A combined strategy is being adopted to integrate as significant properties both conceptual and experimented levels of the digital object. By doing so we intend to provide a neutral (DBML) and abstract (ontology) representation of relational databases.

REFERENCES

1. Angela Dappert and Adam Farquhar, "Signi_cance Is in the Eye of the Stakeholder,
" The British Library, Wetherby, West Yorkshire, 2009
2. Ronald Bourret, "XML and Databases, " Copyright 1999-2005 by Ronald Bourret.
Last updated September, 2005
3. Lee Buck. "Data models as an XML Schema development method", XML 99, Phyladelphia,
Dec. 1999.
4. Consultative Committee for Space Data Systems. "Reference Model for an Open
Archival Information System (OAIS) - Blue Book," National Aeronautics and Space
Administration, Washington, 2002.
5. Edgar Codd, "A Relational Model of Data for Large Shared Data Banks," in Communications
of the ACM, 1970.
6. Michael Day, "The OAIS Reference Model," Digital Curation Centre UKOLN, University
of Bath, 2006
7. Claire Eager, "The State of Preservation Metadata Practices in North Carolina
Repositories," Chapel Hill, North Carolina, 2003
8. Miguel Ferreira, "Introdu_c~ao _a preservacao digital - Conceitos, estrat_egias e actuais
consensos," Escola de Engenharia da Universidade do Minho, Guimar~aes, Portugal,
2006
9. Ricardo Freitas, "Preserva_c~ao Digital de Bases de Dados Relacionais," Escola de
Engenharia, Universidade do Minho, Portugal, 2008
10. R. Freitas, J. Ramalho, "Relational Databases Digital Preservation," Inforum:
Simpsio de Informtica, Lisboa, Portugal, 2009, ISBN: 978-972-9348-18-1; [Online].
Available:
http://repositorium.sdum.uminho.pt/handle/1822/9740
11. M. Jacinto, G. Librelotto, J. Ramalho, P. Henriques, "Bidirectional Conversion
between Documents and Relational Data Bases," 7th International Conference on
CSCW in Design, Rio de Janeiro, Brasil, 2002.
12. B. F. Lavoie, "The Open Archival Information System Reference Model: Introductory
Guide," Digital Preservation Coalition, Dublin, USA, Technology Watch
Report Watch Series Report, 2004.
13. K.-H. Lee, O. Slattery, R. Lu, X. Tang and V. McCrary, "The State of the Art and
Practice in Digital Preservation," Journal of Research of the National Institute of
Standards and Technology, vol. 107, no. 1, pp. 93-106, 2002.
14. "OWL - Web Ontology Language" [Online]. Available:
http://www.w3.org/TR/owl-features/
15. Pat Manson, "Digital Preservation Research: An Evolving Landscape," European
Research Consortium for Informatics and Mathematics - NEWS, 2010.
16. "PLANETS - Preservation and Long-term Access through NETworked Services"
[Online]. Available: http://www.planets-project.eu/
17. J. Ramalho, M. Ferreira, R. Castro, L. Faria, F. Barbedo, L. Corujo, "XML e
Preserva_c~ao Digital," Dep. Inform_atica, Universidade do Minho e Instituto dos Arquivos
Nacionais, Torre do Tombo, 2007
18. J. Ramalho, M. Ferreira, L. Faria, R. Castro "Relational Database Preservation
through XML modelling, " Extreme Markup Languages 2007, Montr_eal, Qu_ebec,
2007
19. J. Ramalho, P. Henriques, "XML and XSL - Da Teoria _a Pr_atica," FCA - Editora
Inform_atica, 2002.
20. "RODA - Repository of Authentic Digital Object" [Online]. Available:
http://roda.dgarq.gov.pt/
21. "SIARD - Format Description," Swiss Federal Archives - SFA, 2008.
22. "SWRL: A Semantic Web Rule Language Combining OWL and RuleML" [Online].
Available: http://www.w3.org/Submission/SWRL/
23. K. Thibodeau, "Overview of Technological Approaches to Digital Preservation and
Challenges in Coming Years," presented at The State of Digital Preservation: An
International Perspective, Washington D.C., 2002.
24. D. Waters, "Good Archives Make Good Scholars: Reections on Recent Steps
Toward the Archiving of Digital Information," 2002
25. C. Webb, "Guidelines for the Preservation of Digital Heritage," United Nations
Educational Scienti_c and Cultural Organization - Information Society Division,
2003.
26. A. Wilson, "Signi_cant Properties Report," InSPECT Work
Package 2.2, Draft/Version 2 (2007), [Online]. Available:
http://www.signi_cantproperties.org.uk/documents/wp22 signi_cant properties.pdf.
27. Wikipedia contributors, "Database models," in Wikipedia, The Free Encyclopedia,
2008. [Online]. Available: http://en.wikipedia.org/wiki/Database models/
28. XML, "Extensible Markup Language", in W3C - TheWorld WideWeb Consortium
[Online]. Available: http://www.w3.org/XML/