

Studies in Computational Intelligence

Volume 700

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/7092>

Tomasz Traczyk · Włodzimierz Ogryczak
Piotr Pałka · Tomasz Śliwiński
Editors

Digital Preservation: Putting It to Work

 Springer

Editors

Tomasz Traczyk
Institute of Control and Computation
Engineering
Warsaw University of Technology
Warsaw
Poland

Piotr Pałka
Institute of Control and Computation
Engineering
Warsaw University of Technology
Warsaw
Poland

Włodzimierz Ogryczak
Institute of Control and Computation
Engineering
Warsaw University of Technology
Warsaw
Poland

Tomasz Śliwiński
Institute of Control and Computation
Engineering
Warsaw University of Technology
Warsaw
Poland

ISSN 1860-949X

ISSN 1860-9503 (electronic)

Studies in Computational Intelligence

ISBN 978-3-319-51800-8

ISBN 978-3-319-51801-5 (eBook)

DOI 10.1007/978-3-319-51801-5

Library of Congress Control Number: 2016961660

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Long-term digital preservation, the process of maintaining digital objects through time to ensure continued access, has become a crucial issue in recent years. The amount and the areas of digitized information are constantly increasing resulting in obsolescence of the software and hardware required to preserve digital information. Despite recognized need for preservation action, still more work is required to effectively address the issue in theory and practice.

The book is divided into two parts. Part I starts with discussion of the basic problems of long-term digital preservation. There are widely discussed and analyzed concepts and requirements for long-term digital preservation. Further, since metadata play important role in long-term digital preservation, processing of metadata in long-term digital archives is discussed. Long-term digital archives usually preserve all metadata transferred to them with corresponding digital objects and often use a subset of obtained metadata to manage archive assets. They also create preservation metadata, which describe processes and preservation actions applied to digital objects in the archives.

In Part II of the book, a framework based on the Digital Document Repository project CREDO is presented. Within the CREDO project, a demonstrative version of a digital repository enabling short- and long-term archiving of large volumes of digital resources has been designed and launched. The repository acts both as a secure file storage and as a digital archive providing metadata management and including the resources in archival packages. Reliability of information readouts is ensured by the repository through the data recording replication and monitoring mechanisms in the repository's file system, as well as through the distributed nature of the system that enables storing copies of the resources in more than one location. Advanced management system supports scheduling of operations on the archival storage while respecting the low energy consumption requirements.

One of the system primary functions is the support for various currently available data carriers such as hard drives, solid-state drives, and tapes. However, the repository architecture is multi-tiered and it enables (together with the emergence of new technologies) replacement and continuous upgrades of the individual components.

This solution has been designed for institutions that store large digital resources for long periods of time, e.g., cultural institutions, mass media, state administration offices, and healthcare institutions. An evaluation of our framework is presented, which illustrates the viability of our approach in retaining accessibility, authenticity, and usability.

Warsaw, Poland
November 2016

Tomasz Traczyk
Włodzimierz Ogryczak
Piotr Pałka
Tomasz Śliwiński

Acknowledgements

The project entitled *Digital Document Repository CREDO* was granted as a part of the pilot undertaking of the Polish National Centre for Research and Development, entitled '*DEMONSTRATOR+* Supporting scientific research and development works for demonstration scale' (Grant No. WND-DEM-1-385/00). The project was co-financed by the European Union through the European Regional Development Fund under the Operational Programme 'Innovative Economy' for the years 2007–2013, Priority Axis 1—Research and development of modern technologies.

Contents

Part I Problems of Long-Term Digital Preservation	
Requirements for Digital Preservation	3
Tomasz Traczyk	
Metadata in Long-Term Digital Preservation	15
Grzegorz Płoszajski	
Part II Solutions Proposed by the CREDO Project	
The CREDO Project	65
Tomasz Traczyk and Włodzimierz Ogryczak	
CREDO Repository Architecture	77
Tomasz Traczyk	
Information Processing in CREDO Long-Term Archive	93
Tomasz Traczyk	
Metadata in CREDO Long-Term Archive	109
Tomasz Traczyk and Grzegorz Płoszajski	
Persistence Management in Long-Term Digital Archive	123
Piotr Pałka	
Power Efficiency and Scheduling Access to the Archive	133
Tomasz Śliwiński	
Information Management in Federated Digital Archives	143
Piotr Pałka and Tomasz Traczyk	
Index	157

Contributors

Włodzimierz Ogryczak Institute of Control and Computation Engineering,
Warsaw University of Technology, Warsaw, Poland

Piotr Palka Institute of Control and Computation Engineering, Warsaw
University of Technology, Warsaw, Poland

Grzegorz Płoszajski Institute of Control and Computation Engineering,
Warsaw University of Technology, Warsaw, Poland

Tomasz Traczyk Institute of Control and Computation Engineering, Warsaw
University of Technology, Warsaw, Poland

Tomasz Śliwiński Institute of Control and Computation Engineering,
Warsaw University of Technology, Warsaw, Poland