Designing the Repository of Documentary Cultural Heritage

Nataliia Kunanets[1](✉), Viktoriya Dobrovolska[2], Nataliia Filippova[3], Kazimi Parviz[4], Halyna Lypak[5], Oleksii Duda[5], Nataliia Veretennikova[1], Marta Karp[6], Chrystyna Shunevych[1], and Luybov Dubrovina[3]

[1] Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine
nek.lviv@gmail.com, nataver19@gmail.com, krishirak@gmail.com
[2] Department of Cultural Studies and Informational Communications, National Academy of Managerial Staff of Culture and Arts, Kyiv, Ukraine
vika_dobrovolska@ukr.net
[3] Vernadsky National Library of Ukraine, Kyiv, Ukraine
filippova.natalya@hotmail.com, lyubov.dubrovina@gmail.com
[4] Department of Bibliography, Library and Information Faculty, Baku State University, Baku, Azerbaijan
pkazimi@mail.ru
[5] Department of Computer Science, Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine
halyna.lypak@gmail.com, oleksij.duda@gmail.com
[6] Applied Linguistics Department, Lviv Polytechnic National University, Lviv, Ukraine
martha24@ukr.net

Abstract. Digitization of communication processes generates the need to design repositories of documentary cultural heritage, providing an open access to historical artifacts. The process of forming repositories is to combine information resources, which is carried out to meet qualitatively the information users' needs. To preserve such large arrays of information resources, the authors have proposed an information technology based on the creation of data warehouses where the resources of social memory institutions are consolidated. The problem of certification of electronic data warehouses, which is currently carried out by CoreTrustSeal, is analyzed and the requirements for the design of such data warehouses are determined. A prototype of an information technology platform for a repository of documentary cultural heritage is proposed.

Keywords: Repository · Cultural heritage · Information technology · Information consolidation

1 Introduction

The design of archival repositories of documentary cultural heritage is a rather complex scientific problem, and its solution is one of the most important tasks today. Awareness of

N. Shakhovska and M. O. Medykovskyy (Eds.): CSIT 2020, AISC 1293, pp. 1034–1044, 2021.
https://doi.org/10.1007/978-3-030-63270-0_70

the undeniable significance and uniqueness of cultural heritage sites, due to the specific historical, cultural, social and political preconditions of their formation, actualizes the study of certain aspects of preservation and functioning of information resources in society. Therefore, understanding the importance of providing an open access to full texts of sources of scientific information contributes to the emergence of scientific research on the organization and functioning of repositories as one of the most representative forms of presenting scientific results and guaranteeing their promotion through openness to the general public.

1.1 Analysis of Recent Researches and Publications

Publications of the specified problems cover scientific achievements of researchers of both the specified subject field, and the connected scientific directions. Especially, the role of libraries, archives, museums as repositories of national and cultural heritage, carriers of social memory, the ideological basis for the formation of social and cultural environment, as well as forms and benefits of their interaction are covered in the works of authoritative Ukrainian scientists such as Dubrovina et al. [1]. The publications by Kovalchuk and Lobuzina [2] are devoted to the problems of creating an integrated web resource of books in Ukraine and information technologies of creating national bibliographic resources. The organization of long-term preservation of library electronic resources, including the digitized ones, in the context of implementation of digital library projects is considered in the monograph by Lobuzin [3]. The issue of preservation of digital cultural heritage, which can be divided into two main blocks, namely digital copies of cultural objects ('digital surrogates') and digital objects by origin ('born digital') is explored by Dobrovolska [4]. The scientific achievements by Kunanets include the study of modern information technologies for the preservation of historical and cultural heritage [5].

In turn, the experience of forming printed and electronic resources of reference biographical information in Ukraine and its importance for the deployment of integrated, thematic and regional information resources are reflected in the scientific works by Popyk [6]. The results of studying the methods and technologies of archiving electronic library resources are presented in the publication by Strilets [7]. The work by Polovynchak [8] and Karpyuk [9] is devoted to the issue of preserving information of the interactive space (in particular, network documents of social media).

Promising directions of repository activity in Ukraine on the example of trade repositories are covered in the paper by Semchenkova [10], while in the publication by Vodorezova, devoted to the economic and legal aspect of legal regulation of repositories, the research results of features and the current state of development of "open science' in Ukraine through the prism of legal regulation, funding and access to scientific repositories is presented. Among other things, the researcher identified the main functions of scientific repositories as a key element of open science [11].

Despite some achievements in scientific understanding, the project activities for the formation of repositories of documentary cultural heritage remain poorly studied and need further research.

2 Consolidation of Information Resources as a Type of Repository Activity

Repositories are the result of combining information resources, which is carried out to qualitatively meet the information needs of users. The request to the repository must be satisfied by providing such an information product, which would contain selected from various sources and complete information that fully reveals the essence of this request. These criteria are met by consolidated information, which, according to experts [12] are obtained from several sources and systematically integrated various types of information resources that together are endowed with features of completeness, integrity, consistency and constitute an adequate information model of the problem area for its analysis, processing and effective use in decision support processes.

In a broad sense, consolidation means strengthening, joining, integrating, uniting something. Regarding information consolidation, it is distinguished a physical consolidation, i.e. information actually gathered in one place, and logical (or federal) consolidation, i.e. distributed information, which from the user's point of view is in a single repository with a common directory and equal access to it. Most often, the concept of consolidated information is used in the context of physical consolidation, when it refers to analytical and synthetic activities to select complete and relevant information to support decision-making in a problem area. In the process of creating consolidated information, data is first transformed into information, then – into knowledge and, finally – into an information product.

The final product of information consolidation should contain processed information based on the found, filtered according to the given criteria, facts presented in a user-friendly form and at a specified time without reference to the physical location of a user. This should be considered when forming a documentary information retrieval system that serves as a kind of guide to the repository and should contribute to achieving a specific goal. Therefore, the process of information support of user needs involves the use of various types of information resources.

When creating information products by means of information consolidation, their quality would correspond to modern trends in the information society, take into account the requirements for completeness and reliability of information, the possibility of timely delivery in the required form and at appropriate time for a user. The basis for the creation of information products of consolidated information is the latest, complete, reliable, and relevant information.

The consolidated information product of the repository may be electronic resources of cultural and memorial institutions, selected to facilitate the full satisfaction of the user's request, and which can be presented by electronic data (information in any form), electronic programs (sets of operators or subprograms that ensure implementation of certain tasks, in particular data processing) or a combination of these types in one resource. As a rule, they are remote access resources, so in this case, the concept of 'electronic resource' means the generalized concept of an electronic document and other types of electronic information, including local and global information networks and technical means to provide access to them [12].

Consolidated information product can be considered as "not just an ordered set of individual fragments of the studied industry, but a holistic picture that reflects the object of interest of the user in a convenient perspective, contains options for alternative, but relevant, information from different sources, as well as makes it possible to perceive the object in its dynamics and from different points of view" [12].

The main components of the repository as a consolidated information resource of any industry are the information database of a given subject area and the automated information retrieval system as a means of interaction with it.

The formation of repositories as information systems of consolidated information pursues the following goals:

- creation of information resources of a certain information industry;
- development of practical measures to optimize the use of information resources to perform their inherent functions;
- thorough processing of information flows, identification of trends and patterns of development of the industry;
- establishing a system of relationships of basic concepts, ways of integration of theoretical research with the introduction and use of information systems and technologies for information consolidation.

3 Certification Requirements for the Repository of Documentary Cultural Heritage

In the modern scientific community, there is a well-established idea that the best option for exchanging scientific data is the use of professional repositories, characterized by a combination of three main factors of their operation such as specialization, ease of use and reliability. Despite the existence of several main types of repositories, all of them combine the main advantages over other possible ways of data exchange, including long-term preservation, guaranteed attention to metadata (i.e. data about data) that allow search relevance, support for different data formats and citations [13].

The issue of certification of electronic data warehouses, which is now carried out by an organization called CoreTrustSeal [14], also deserves special attention. It offers any interested data repository a basic level certification based on the DSA-WDS Core Trustworthy Data Repositories Requirements catalog and procedures [15].

This universal catalog of requirements reflects the main characteristics of existing data warehouses and is the result of cooperation between the World Data System of the International Science Council (WDS) and the Data Seal of Approval (DSA) under the aegis of the Alliance for Research Data, which combined the relevant certificates of data repositories – WDS and DSA.

Thus, CoreTrustSeal certification has effectively replaced DSA certification and WDS permanent member certification by offering the appropriate CoreTrustSeal Data Repository Application Management Tool. It is assumed that repositories that have previously been DSA certified will receive CoreTrustSeal when they renew their certificate, as well as all new certification applications. Similarly, WDS requires its permanent members to apply for or renew their membership in order to obtain a CoreTrustSeal data repository certificate.

As CoreTrustSeal displays the characteristics of reliable data repositories, all catalog-regulated requirements are defined as mandatory and equally important, but separate characteristics of electronic data warehouses. These requirements are regulated by three main documents:

- Introduction to the Requirements for reliable CoreTrustSeal data repositories 2020–2022;
- Actual Requirements for reliable CoreTrustSeal data repositories 2020–2022;
- Glossary of Requirements for Reliable CoreTrustSeal Data Repositories 2020–2022 [16].

Introduction to the Requirements for Reliable CoreTrustSeal Data Repositories 2020–2022 identifies the benefits of certification, which in the modern world is available at different levels, from basic to advanced and formal ones. It is noted that certification at the basic level already provides significant benefits for the repository and its manager. It certifies the reliability and stability of electronic data storage, helping to ensure the transparency and proper quality of its operation, which is achieved through compliance with existing standards. Among the external advantages of basic certification is the increase in the level of trust to the repository, the growth of its authority and representativeness in the social and communication environment. In turn, the internal advantages include the possibility of comparative analysis in order to determine the pros and cons of data warehouses, including preparation for accreditation processes. This provides an outline of prospects for further development of the repository to increase its reliability and quality of operation in the information space. Ultimately, the basic level certification serves as a kind of step for the procedure of passing the higher level certification in the future.

Directly, Requirements for reliable data repositories CoreTrustSeal 2020–2022 inform about preconditions, general recommendations and actually recommendations for reliable electronic data warehouses. In particular, when forming an application for certification, it is necessary to choose one of the types of repositories regulated by Requirements:

- subject repository;
- institutional repository;
- national repository system, including governmental one;
- repository of publications;
- library;
- museum;
- archive;
- repository of research projects;
- etc.

Thus, in the context of the formation of the documentary cultural heritage repository, it is advisable to define its type as a library repository, considered its long-term functioning in order to represent the cultural heritage.

Requirements for the organizational infrastructure of the repository provide a clear outline of the basic characteristics. It is distinguished the purpose and scope of the repository; rules of use; continuity of access; confidentiality and ethics; organizational infrastructure; expert guidance; data integrity and reliability; documented procedures for managing the archival storage of repository data; conservation plan; data quality; work processes; data detection and identification; data reuse; technical infrastructure; security [16].

Thus, the requirements for repositories outline three groups of characteristics, namely the organizational infrastructure (paragraphs 1–6), the management of digital objects (paragraphs 7–14) and technological characteristics (paragraphs 15–16). When applying for certification, it is necessary to indicate the level of compliance with each of the above requirements on the following scale: 0 – not implemented; 1 – not yet represented in the repository; 2 – the repository has a theoretical concept; 3 – the repository is in the implementation phase; 4 – fully implemented in the repository. Compliance with these characteristics will ensure that the repository receives a certificate of reliability of CoreTrustSeal data warehouses.

Glossary of Requirements for Reliable CoreTrustSeal Data Repositories 2020–2022 provides definitions of concepts used in the CoreTrustSeal 2020–2022 Trusted Data Repository Requirements, such as access rights information; rating; archive; authenticity; user; curation; data; database; data set; community; digital object; digital storage; domain or subject repository; data entry; integrity; institutional repository; knowledge base; longevity; long-term preservation; migration; national repository system, including government one, Open Archival Information System (OAIS); preferred formats; producer; information about the origin; publication repository; reference model; repository; repository of research projects; reuse; management; plan of legal succession [16].

A repository is defined as an organization that stores, manages, and provides access to many types of digital materials in a variety of formats. It is also noted that materials in online repositories are used to ensure their search, openness and reusability. In addition, in accordance with the Requirements, it is necessary to provide with a sufficient level of control over the authenticity, reliability, availability of digital material and possibility of its use on a regular basis [16].

4 Documentary Cultural Heritage Repository Project

When designing the information technology platform of the repository of documentary cultural heritage, a six-level architecture has been chosen, which provides the support of context-dependent services (Fig. 1).

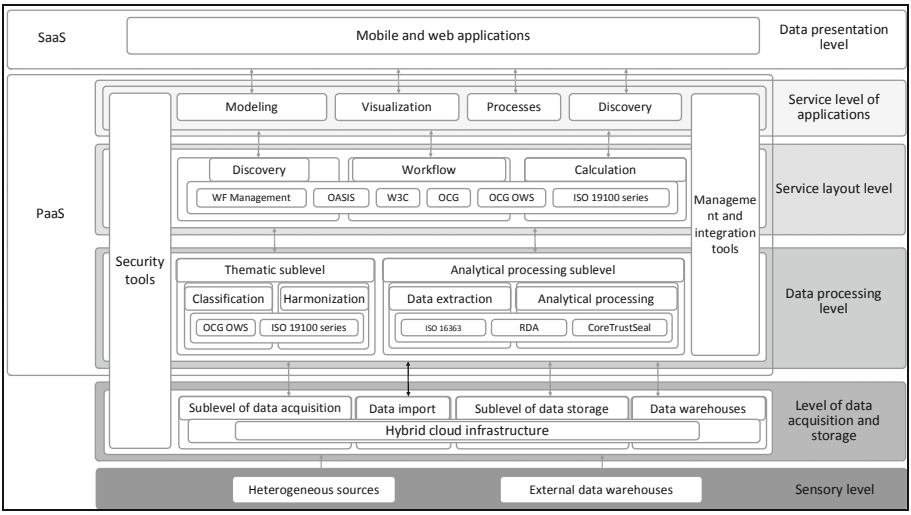

Fig. 1. Information and technology platform of the repository of documentary cultural heritage

This multitier architecture can be used to create a responsive platform based on SaaS (Software as a Service) and PaaS (Platform as a Service). In the case of building a service model, it is transformed into a PaaS solution, on the basis of which a set of other services can be formed. If you use an architecture to create a service that interacts directly with users, the application is transformed into a SaaS solution.

At the sensory level there are various sources of documentary data collections, and the next level contains an integration platform of hybrid cloud infrastructure, which implements the functions of receiving, importing and storing data. The collected data are analyzed and adapted to the requirements of specific tasks by means of the next level.

In the process of data transmission at higher levels, each of the layers of the proposed architecture contributes to the processing of contextual information.

The service layout layer contains service and work processes, identifies data sources, and links to processing components, providing a computing platform for implementing context-dependent services.

The service application layer uses lower-level results and services in the form of specific tools to perform contextual data analysis in order to take correct and comprehensive decisions.

The management and integration layer is used to automate the distribution of filtered data streams between horizontal layers. This, in turn, ensures that the obtained results in one layer will be adequately related to each other and syntactically correctly transferred to another layer. This level also implements change management processes and reduces the time spent on managing a tiered architecture. The security layer implements the functions of authentication, authorization and audit of data use and service delivery processes.

Functionally, such a repository, based on the presented information technology platform, consolidates information resources from various sources (libraries, archives, museums, etc.), and in accordance with consumer search queries produces a relevant query information product (Fig. 2).

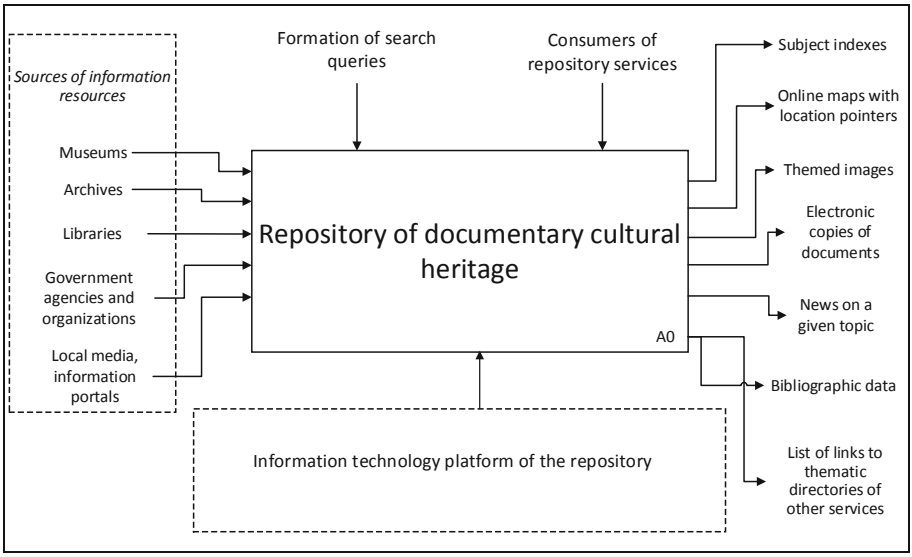

Fig. 2. Functional diagram of the repository of documentary cultural heritage

It should be noted that to ensure the relevance and completeness of information retrieval results, it is necessary to provide a description of the metadata of source information resources in the repository according to a single unified standard. According to experts, such a universal standard is RDA: Resource Description & Access [17].

Developing the prototype of the Web interface of the information technology platform of the repository (Fig. 3), the main attention was paid to providing potential users with convenient tools for finding the necessary materials and quick access to them.

In the process of searching, the user will be provided with such aids as keywords, grouping by various criteria, etc. Depending on the available content, the interface of the Web page can be supplemented with specific sections, menus and other user tools.

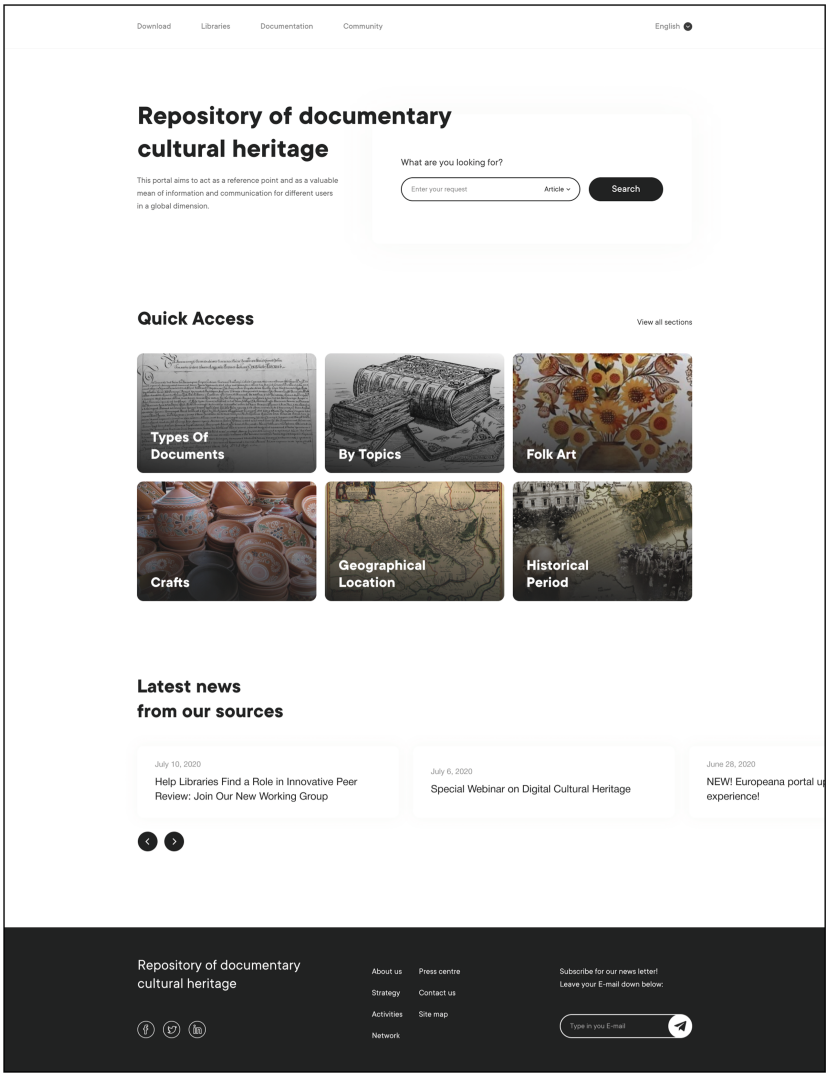

Fig. 3. Prototype of the home page for the repository of documentary cultural heritage

5 Conclusion

A prototype of an information technology platform for a repository of documentary cultural heritage based on a six-level architecture is proposed, which provides consolidation of information resources of libraries, museums and archives. This approach contributes to the comprehensive support of information requests on the basis of consolidated resources.

The relevance and completeness of meeting the information needs is ensured by the correct presentation of metadata for documents included in the repository, according to a single unified RDA standard.

The web interface of the information technology platform of the repository is designed as a convenient tool for finding the necessary documents and quick access to them. Convenient algorithms based on various search criteria are used for search.

References

1. Dubrovina, L., Kyrydon, A., Matyash, I.: Archives, libraries, museums - the source of national memory, cultural heritage of Ukraine. Libr. Bull. **1**, 3–10 (2017). (in Ukrainian)
2. Kovalchuk, H., Lobuzina, K.: Problems of creating an integrated web resource of books in Ukraine. In: Materials of the III International Bibliological Readings "Old Prints and Rare Editions in the University Library", pp. 8–17 (2016). (in Ukrainian)
3. Lobuzin, I.: Digital library projects: technological solutions and management of the life cycle of collections, Kyiv, 216 p. (2016). (in Ukrainian)
4. Dobrovolska, V.: Preservation of the digital cultural heritage: basic problems, tasks, principles of selection. Soc. Hum. Sci. Polish-Ukrainian Sci. J. **03**(23) (2019). https://sp-sciences.io.ua/s2663927/dobrovolska_viktoriya_2019_._preservation_of_the_digital_cultural_heritage_basic_problems_tasks_principles_of_selection._social_and_human_sciences._polish-ukrainian_scientific_journal_03_23
5. Lypak, H., Kunanets, N., Pasichnyk, V., Veretennikova, N.: Modern information technologies for preservation and presentation of historical and cultural heritage at small museums. In: 9th International Conference on Advanced Computer Information Technologies (ACIT), Ceske Budejovice, Czech Republic, 5–7 June 2019, pp. 483–486 (2019)
6. Popyk, V.: New developments in the formation of domestic resources of biographical information. In: Place and Role of Libraries in the Formation of the National Information Space: Materials of the International Scientific Conference, pp. 352–355 (2014). (in Ukrainian)
7. Strilets, N.: Methods and technologies of archiving electronic library resources. Libr. Sci. Document. Informol. **3–4**, 40–46 (2012). http://nbuv.gov.ua/UJRN/bdi_2012_3-4_7. (in Ukrainian)
8. Polovynchak, Yu.: Activities of libraries for the preservation of digital heritage of interactive space. Vernadsky Natl. Libr. Ukraine **49**, 123–132 (2018). http://nbuv.gov.ua/UJRN/npnbuimviv_2018_49_11. (in Ukrainian)
9. Karpyuk, D.: Network documents of social media as an object of deposit. Vernadsky Natl. Libr. Ukraine **42**, 42–51 (2015). (in Ukrainian)
10. Semchenkov, O.: Perspective directions of repository activity in Ukraine. Finance Ukraine **7**, 116–127 (2016). [in Ukrainian]
11. Vodorezova, S.: Legal regulation of repositories: economic and legal aspect. Law Innov. **4**(20), 55–61 (2017). (in Ukrainian)
12. Kunanets, N., Pasichnyk, V.: Introduction to the Specialty "Consolidated Information": Textbook, p. 196. Lviv Polytechnic National University Publishing House, Lviv (2010). (in Ukrainian)
13. Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., Wutich, A.: Qualitative data sharing: data repositories and academic libraries as key partners in addressing challenges. Am. Behav. Sci. **63**(5), 643–664 (2019). https://journals.sagepub.com/doi/10.1177/0002764218784991
14. CoreTrustSeal. https://www.coretrustseal.org/
15. CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022, 20 November 2019 (2019). https://doi.org/10.5281/zenodo.3638211

16. CoreTrustSeal Trustworthy Data Repositories Requirements. https://www.coretrustseal.org/why-certification/requirements/. Accessed 02 July 2020
17. Tosaka, Y., Park, J.: RDA: resource description & access – a survey of the current state of the art. J. Am. Soc. Inf. Sci. Technol. **64**(4), 651–662 (2013)