

# Professional Solicitor

Linking Canada's Information



## *MANAGEMENT*

## Publisher

Valoree McKay, CAE

## Editor

Rachel Hertz Cobb

## Copy Editor

Rachel Hertz Cobb

## Layout & Design/ Review Coordinator

Beverly Bard

## Member Communications Standing Committee 2014

James Alphonse Bennett (Chair); Lauren Bourdages;  
Gerry Burger-Martindale; Paulina Rousseau;  
Matthew Thomas

Published since 1956 by the Canadian Library  
Association 6 times per year as a membership  
service to CLA members in good standing.

Indexed in the *Canadian Index* and *Library Literature*  
and available online in the Canadian Business &  
Current Affairs Database.

## Deadlines are as follows:

Issue	Editorial	Advertising Space
1	December 12	January 10
2	February 11	March 12
3	April 14	May 11
4	June 16	July 11
5	August 15	September 10
6	October 17	November 8

Acceptance of an advertisement does not imply  
endorsement of the product by the Canadian Library  
Association.

### **Feliciter**

Canadian Library Association  
1150 Morrison Drive, Suite 400  
Ottawa, Ontario K2H 8S9  
613-232-9625  
Fax: 613-563-9895

## Advertising, including career ads:

Beverly Bard  
*Feliciter*, Desktop Publisher /  
Book Review Coordinator  
Tel.: (613) 232-9625, ext. 324  
Fax: (613) 563-9895  
Email: [bbard@cla.ca](mailto:bbard@cla.ca)

©The Canadian Library Association  
ISSN 0014 9802

**Theme:** Data Management  
**Guest Editor:** *Catelynne Sahadath  
& Irena Trebic*  
**Liaison:** *Lauren Bourdages*

## Guest Editorial

**3** Data Management  
*by Catelynne Sahadath & Irena Trebic*

## Theme Features

**8** Copyright Skills in Academic Libraries  
*by Sam Cheng & Christina Winter*

**13** Mischief Managed: A Brief Introduction  
to Data Management  
*by Kim Silk*

**15** Research Data Management at Concordia  
University: A Survey of Current Practices  
*by Alex Guindon*

**18** Is eDiscovery the New Frontier in  
Information Management?  
*by Ariana Ross*

**21** Following the Trail of Breadcrumbs:  
Your Health from Data to Decisions  
*by Lee-Anne Ufholz & Lindsey Sikora*

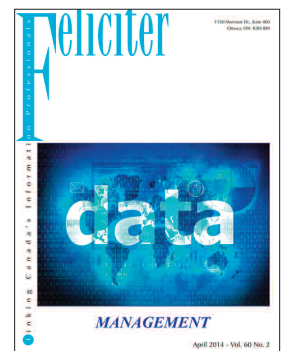
**24** Metadata Management on a Budget  
*by David Cook*

## Front Cover

### This issue's cover image:

Stock Illustration: Data concept  
ID 39055464 © Dwnld777 |  
Dreamstime.com

**Cover Design:** Beverly Bard



---

**27** Beyond the Hype: Data Management and Data Governance  
by *Melanie Sucha*

## Feature Articles

**30** Know Thy Vendor: Getting the Best in Off-Site Records Storage  
by *Guy Robertson*

## Columns

**4** **President's Message**  
Spring: A Time for Celebration and Conferences  
by *Marie DeYoung*

**6** **From the Executive Director's Chair**  
Libraries, So Much More...  
by *Valoree McKay, CAE*

**8** **Copyright Column**  
Copyright Skills in Academic Libraries  
by *Sam Cheng and Christina Winter*

**33** **Book Reviews**

## Departments

**35** *CLA Executive Council & Staff Contacts*

## Guest Editorial

by *Catelynne Sahadath & Irena Trebic*



## Data Management: The Business of Libraries

If your eyes glaze over when someone mentions data management in libraries, or if you think that data management is a concept reserved for technical services or academic institutions, you are not alone. While soliciting articles for the data management-themed issue of *Feliciter*, we were surprised at how many of our library peers were convinced that data management is a concept that does not apply to them. The truth is, data management is not something that became relevant in the information age; it is what we have been doing in libraries all along.

One of the many goals of libraries is the provision of access to information, as well as the delivery of support in interpreting that information. Though these goals have not changed dramatically over time, the nature of information and users' expectations certainly have. So while our goals have remained unaltered, the path we take to achieve them must change to accommodate the evolving nature of information, and of data.

Enter the data management theme issue. We solicited articles for this issue that reflect data management for every library professional, in the hopes of dispelling some of the myths surrounding the topic, such as:

– ***Data management involves nothing more than research data.*** While it is true that research data is a big, important issue in data management, there are many aspects of data management, including metadata, patron data, statistical data, and data security, that are also important for libraries.

*Guest Editorial continued on page 5*





## Spring: A Time for Celebration and Conferences

**Welcome to spring**, or at least to the end of winter and the promise of spring! The countdown to the [Victoria Conference](#) has begun but there are many exciting things in store before we gather at the end of May.

Spring is a time of celebration and I would like to take this opportunity to congratulate all of the new librarians, archivists, information managers and library technicians. Graduation is a very special time, not only do you have an exciting career ahead of you but you also get to recognize the hard work and dedication it took to complete your program.

Our Networks have been very busy, from delivering webinars on *Becoming indispensable: The value proposition* ([Government Libraries Information Professionals Network](#)) to promoting National School Library Day ([Voices for School Libraries Network](#)). I am pleased to report that we have two new networks, the [Prison Libraries Network](#) and the [Next Generation Network](#). All 25 Networks represent grass roots involvement by members across all types of libraries, in urban and rural regions and reflective of the varied interests of our membership.

In March I had the pleasure of spending a week at the Library of Beijing Normal University in Zhuhai, China. It was a wonderful experience, not just to be exposed to the incredibly rich Chinese culture but to hear that colleagues from that country are experiencing the same challenges as many Canadian university and college libraries – declining circulation and demands on staff and electronic resources that are not keeping up with funding. A recent issue of *The Coast*, a weekly alternative newspaper in Halifax had an opinion piece titled *I hang out at libraries, even when I don't need a book* (Sparling, *The Coast* 21:42 2014, pg.6). The author spoke about the special place that libraries have in our communities, as places where users can gain technology skills, find information on government programs and services as well as borrow books. It is the author's "personal mission to fight the attitude that libraries are reserved for the bookish. The most important part of

public libraries isn't the 'library' – it's the public." What affirmation as I see the amazing new Halifax Public Library being completed and slated to open this fall.

In this issue you will have an opportunity to hear from Valoree McKay, our new Executive Director. She has been working very hard on behalf of CLA and I look forward to your meeting her in Victoria.

The conference is just around the corner. I am very excited about the session *Building Bridges to the Future - Colleagues, Collaboration, Consultation (the 3c's)*. At this session you will have an opportunity to hear from industry experts on the need for the 3c's and the impact on our profession. The conversation, moderated by Stephen Abram, will be open to audience questions and comments. Speakers include myself, Barbara Stripling, President, American Library Association (ALA), Shelagh Paterson, Executive Director, Ontario Library Association, Ken Roberts, Past President CLA, Member of the Royal Society of Canada Panel, Status and Future of Canada's Libraries and Archives and Donna Scheeder, Deputy Chief Information Officer, Library of Congress and IFLA President-elect.

I hope you will attend the annual First Timers Breakfast on Thursday. This is a great opportunity to meet new conference attendees and share with them your best advice on getting the most out of the conference. Social events include a pub crawl, where I am sure Victoria will provide ample opportunity to enjoy beverages to suit all tastes and mark your conference program for Friday night where the Greater Victoria Public Library Central Branch will host an evening under the stars. 🌟

See you in Victoria.

[marie.deyoung@smu.ca](mailto:marie.deyoung@smu.ca)



*Guest Editorial continued from page 3*

– **Data management is reserved for technical services.** Like many library functions, managing data in a meaningful way relies on all stakeholders – from the user to the public services team to the technical services team – playing a role in preserving, manipulating, and protecting data.

– **It is OK to put data management on the back burner.** Data management is a crucial element of librarianship and one that will help keep our institutions relevant in the future. Acknowledging its centrality and reacting to the changing data environment is an important step in safeguarding the information profession.

The articles we have selected for this issue cover various aspects of data management and how they apply in disciplines ranging from law to medicine.

Kimberly Silk provides us with an introduction to data management and explains the difference between *data* and *big data*. She discusses the important role data librarians play at universities, and she introduces us to the concept of the codebook, which is used by agencies such as Statistics Canada to help define the content within a data file. Kim also provides a brief introduction to the Research Data Lifecycle developed by the UK Data Archive.


Alex Guindon discusses the findings from his survey of current research data management (RDM) practices at Concordia University. He elaborates on the Concordia researchers' data-sharing preferences and the lack of a means to preserve data in spite of the researchers' desire to do so. He emphasizes the importance of creating new RDM services to help address researchers' needs.

Ariana Ross demonstrates the increasing importance of eDiscovery, which applies records management practices within a legal environment. She compares eDiscovery in Canada to that in the United States and gives us a summary of the twelve Sedona Canada Principles. Ariana maintains that information professionals have the necessary skills to become highly valued members of eDiscovery teams, and it is up to them to market their skills in this non-traditional setting.

Lee-Anne Ufholz and Lindsey Sikora look at the complexities associated with research data management in the health disciplines. They educate us about practices such as de-identification of data, which helps protect the privacy of patients in clinical trials. Lee-Anne and Lindsey have procured a lovely figure that illustrates the complex issues relating to research integrity and the sharing of data from clinical trials.

David Cook emphasizes the importance of metadata management in libraries in order to ensure that the records in our catalogues are accurate and up to date. He encourages us to be brave and try to learn different data management tools in order to save time and money when maintaining records. David shares his expertise with the various software used for manipulating metadata.

Finally, Melanie Sucha discusses the drivers behind data management and data governance as they relate to a corporate setting, which views data as an asset with business value. She describes the differences between master data and transactional data and looks at common requirements for data quality.

It is our hope that after reading these insights on data management, you may also be able to picture yourself in a data management context. 

*Catelynne Sahadath (@MetaCatie) is the Metadata Librarian at the University of Calgary. She genuinely believes that great metadata will save the world, and will take every opportunity to tell you why.*

*Irena Trebic (Irena.Trebic@uottawa.ca) is the replacement Head, Media Library at the University of Ottawa.*



## Libraries, So Much More...

**In my first** two months with CLA I have received nothing but warm welcomes and a desire to assist me in my learning about the Association; its challenges, successes, concerns, and goals; its membership and their needs; and finally about the library and information world. As most of you know, I don't come from this community. I am an association executive with upwards of 15 years of experience guiding associations similar to CLA through change and growth; a skill that CLA is in need of at this point in their organizational growth. I have always had a love of libraries and books and spent many hours within their physical structures throughout the years.

As a child it was to choose a new book to read; I recall the feel of running my fingers over the spines of the books on the shelves as I read the titles looking for my next adventure. As a teenager it was the school library where I did all my studying and some volunteering in between class and practices at the pool. As a young adult it was the stacks at the university where I spent my hours studying (yes, on the floor between the stacks in that quiet corner of the library where no one else visited very often) and researching to complete my degree. Finally, as an adult, it was where I went to research my travels around the world, re-evaluate my career path, and most recently, accompany my niece and nephew on a Saturday afternoon so that they could repeat my childhood experience with their own twist. I will admit, it has been years since I have visited the library for my own purposes: time and the internet have changed my patterns and needs. What I have learned about this community and the people within it thus far is that the library still provides all those services we expect and we used throughout the years and stages of our lives but it is so much more than that. It is an extremely creative, innovative, fun, engaging, and social community that is full of ideas and opportunities.

Libraries, librarians and information specialists are all contributing to the growth and well-being of society and the individuals that are part of the community that

they are serving. Of course, this is something that you are all aware of. However, for those of us who lost sight of the library once we graduated from school, have figured out our career paths, or who don't have children and therefore don't see the libraries they use and how they've changed since our childhood, this is an exciting discovery. I see an immense amount of opportunity for CLA and its members and it is my goal and mandate to continue to learn and to work with CLA members, Executive Council, library and information stakeholders, and CLA staff to share that discovery and rise to the challenges and the opportunities we are presented with and those we seek out.

Since January 27, I have been taking the time to observe and learn about CLA, both as an organization and operationally, and I have been working to piece together all of the information and tie it to the strategic goals. We have been working very hard to finalize the strategic actions that will support the [Strategic Priorities](#) approved by Executive Council in December 2013: we anticipate making these available in the coming weeks. The review of the CLA By-laws in order to ensure compliancy with the new Canada Not-for-Profit Corporation Act is almost complete: these will be shared with membership in advance of the AGM prior to voting on May 30, 2014.

We have secured the services of a communications professional, Ms. Ann Clemensen, on a contract basis until the end of June as we continue to evaluate that role and what is needed to achieve the strategic priorities on a long-term basis. Work is underway on Canadian Library Month and we have released a [Statement regarding Federal Government Library Consolidation and Closure](#), sent letters to all MPs informing them of our position and requesting a meeting to discuss, and we have sent a follow up [letter to the Privy Council Clerk regarding the appointment of a Chief Librarian](#) for Library and Archives Canada. We have several meetings booked with MPs in the month of April.

Last, but certainly not least, preparation for **CLA 2014** in beautiful **Victoria, British Columbia** is well underway: **registration** is rockin' and the exhibit hall is almost sold out. I hope to meet all of you in Victoria as we enjoy spending time with our colleagues and friends, take advantage of what is sure to be spectacular education, learn what is available to us from our exhibitors and enjoy the spring flowers and sunshine already found on the island and sorely lacking here in Ottawa and most of the rest of the country.

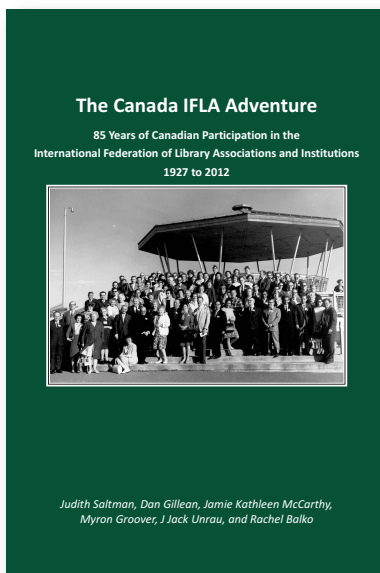
A special thanks to Barb Clubb for all of the time she has spent with me as I learn about CLA, you are an extremely generous individual and I truly appreciate all that you have done to assist me. Thank you to Executive Council for your patience with all my questions, for allowing me the time to learn, and for providing the guidance and support that an ED in a new role requires in such an open and friendly manner. Thank you to those members and stakeholders that have taken the time to get to know me and for sharing your thoughts and goals for CLA with me and finally, thank you to the staff for stepping up and making it happen: yes, there are gaps and yes, some things have been missed or delayed but you all keep coming to work with a smile on your face and the best interests of CLA at the forefront of your daily activities. 🌊

CLA 2014 National Conference & Trade Show

May 28-31, 2014



Join us in beautiful  
Victoria, British Columbia



***The Canada IFLA Adventure:  
85 years of Canadian Participation  
in the International Federation of  
Library Associations and Institutions  
1927 to 2012***

Authors: Judith Saltman, Dan Gillean, Jamie Kathleen McCarthy, Myron Groover, J Jack Unrau and Rachel Balko

Now available from CLA

<http://www.cla.ca/source/Orders/index.cfm?Section=Shop>



2013 • 220 pages • Soft cover • ISBN: 978-0-88802-339-1  
CLA Member rate \$ 35.95  
Non-CLA Member rate \$ 39.95





## Copyright Skills in Academic Libraries

*Margaret Ann Wilkinson ([mawilk@uwo.ca](mailto:mawilk@uwo.ca)) is CLA Copyright Committee editor for the Copyright Column. If you have suggestions for future columns please contact her or John Tooth ([je.tooth@uwinnipeg.ca](mailto:je.tooth@uwinnipeg.ca)). Feliciter's copyright columns are authored by members of CLA's Copyright Committee, and are published after peer-review by the rest of the Committee. The opinions expressed in these columns are those of the authors and do not necessarily represent the position of the Committee or of CLA on any given topic. No column is intended to provide legal advice.*

### Introduction

Over the past several years, many Canadian post-secondary institutions have increased the resources dedicated to the copyright function. Some institutions have created dedicated copyright positions, while others have added responsibility for copyright matters to existing positions. Whether dedicated positions or additional responsibilities in existing positions, in many cases the positions involved are library positions. This article looks at why librarians are well placed to educate our users about copyright, and explores some of the initiatives and tools that they use to do so.

### Copyright and Information Literacy

Copyright issues are prevalent in our everyday activities, and yet many people lack a general understanding of copyright. The Association of College and Research Libraries' *Information Literacy Competency Standards for Higher Education* are widely used as the basis for post-secondary academic librarians' information literacy efforts. Standard Five prompts librarians to educate students on "the economic, legal, and social issues surrounding the use of information [in order that the student] accesses and uses information ethically and legally."<sup>1</sup> One of the performance outcomes is that students demonstrate "understanding of intellectual property, copyright, and fair use of copyrighted material."<sup>2</sup> While other standards have traditionally received more attention during instruction, a comprehensive information literacy program should address this standard.

As others have noted, the aim of grounding copyright education in an information literacy framework is to help individuals gain a deeper understanding of the issues and develop the skills necessary to use information in a way that is compatible with the law – to develop copyright awareness.<sup>3</sup> In addition to being framed as information literacy, copyright issues can also be incorporated into academic integrity sessions to provide context for students.

New technologies are making it easier for people to reproduce and disseminate content, and these activities often involve copyright issues of which users are not aware. For example, there is a general misconception that if it is on the Web, it can be copied and reused without the permission of copyright holders. This lack of user awareness is also seen in consumers signing up for an online service but failing to realize that, in the process, they are relinquishing rights by agreeing to the terms and conditions imposed by the vendor.<sup>4</sup> Library patrons are also sometimes unaware that accessing resources through the library means that they must accept the terms and conditions under which the library has obtained the license to distribute the resource being accessed.<sup>5</sup>

Librarians can engage in meaningful conversations with patrons and develop initiatives to increase understanding of how copyright can impact many of their patrons' day-to-day activities. There are a number of ways librarians can address the need for copyright education.

## New Tools for Copyright Education

There are many different tools that librarians have at their disposal to help individuals navigate copyright. Online tutorials, a dedicated website, and handouts are just some of the ways librarians can communicate information about copyright. In addition to traditional educational methods, librarians are developing new tools that take advantage of technology and collaboration. One such example is the Copyright Literacy in Ontario Colleges Project (the outline of the project can be found on its website).<sup>6</sup> This project involves a series of learning modules from which users can pick and choose. Another example is the University of British Columbia's Digital Tattoo,<sup>7</sup> a joint project between the library and other departments at the university that serves to educate students on issues such as protecting their online identities and exercising their user rights under the *Copyright Act*. These well-designed projects are excellent models of how librarians are engaging with their user base to educate users about copyright in innovative ways.

Several university libraries have developed interactive tools for faculty and students to help them understand and apply the six fair dealing factors to their individual situations. These six factors are guidelines that the Supreme Court has given to determine when a use is fair dealing under the users' right to fair dealing set out in ss.29, 29.1, and 29.2 of the *Copyright Act*.<sup>8</sup> Athabasca University's Fair Dealing Analysis is a simple-to-use application in which users can answer nine multiple-choice questions to receive guidance.<sup>9</sup> Queen's University Library also has a Fair Dealing Evaluator for users to learn how to apply each of the fair dealing factors to the material they want to copy and use.<sup>10</sup> Once users complete the evaluation forms, the Queen's tool provides a date-stamped PDF document for users' records. While these tools claim that they should not be interpreted as legal advice or final judgment, and warn that each evaluation of a use is dependent upon the context of the intended use of the source document, the tools help users to do their own fair dealing analyses and to make educated decisions about the uses they make of materials.

## Serving the Information Needs of Patrons

Librarians, acting in the public interest of promoting access to information, are well-positioned to meet the needs of patrons with respect to copyright. Librarians' skills in assessing reference questions and conducting

reference interviews are useful for helping someone with an information need related to copyright. There may be better resources on the substantive topic about which the user is seeking information than the source the patron initially wants to use but which presents some copyright limitations on the uses the patron can make. Sometimes, the source a patron wants to copy does not permit such an act or, if the user wishes to seek permission for a specific use, locating the copyright holder for permission is not possible. Librarians can recommend alternative resources in these instances. For example, in a situation where a patron wants to convert a copyrighted work into a new format, there may be other library resources on the topic already available in the desired format. Librarians can help patrons make informed decisions on how to meet their information needs while still respecting copyright.

In addition to the discovery of further resources to meet a patron's information need, a copyright question asked at the library can also lead to more effective use of other library services. For an instructor who initially wants to assemble readings into a course pack, the articles may instead be found to be more readily made available to students through direct links to library databases, or the instructor may find it better to use a library service such as electronic reserves to facilitate student access. Librarians can also suggest open access sources, including Creative Commons-licensed works, that have fewer copyright restrictions and are more flexible in how they can be used.<sup>11</sup> Library websites often provide guides to discovering open access content.

## Liaison Activities and Scholarly Communications

Faculty liaison programs are commonly one of the services university and college libraries provide. Liaison activities include collaborating with faculty members in the development of library resources; participating as much as possible in the activities of the departments, schools, and faculties with which they work; promoting library services and collections; and providing library instruction. One of the outcomes of liaison programs is strengthened relationships with faculties, schools, and departments across campus. A 2010 study on copyright communication by Tony Horava found that "[d]ialogue with faculty is essential to raise their awareness of copyright fundamentals for research and teaching, and for their influence on students in developing an

understanding of fair dealing.”<sup>12</sup> Liaison librarians should be important actors in the institution’s overall copyright program because the relationships they have developed allow them to start the conversation about copyright.

Faculty members at universities, and at some colleges, are involved with producing research and scholarly communications. To assist them, librarians can help faculty members to understand publishing options and scholarly impact for the works they create. Copyright ownership would be an important issue in these discussions. The first step for librarians trying to help both students and faculty members decide how to make use of work they themselves have created is to try to help them analyze whether they in fact have ownership of the copyright interests in their work. At universities, where there is a strong tradition of research and scholarly communication, faculty members definitely hold copyright in the works they produce, unless the works are created under a contract that specifies otherwise.<sup>13</sup> In the community college environment, ownership of copyright typically depends either on provisions in a faculty collective agreement or on the employment contracts of faculty members.<sup>14</sup> Students at both universities and colleges always own copyright in their works except in cases when they are contracted to create works outside their studies (for example, as Research Assistants) and there is a written agreement in which copyright ownership is assigned to another (for instance, to the institution or to a faculty member). Therefore, it would be important for librarians to help faculty members and students to understand whether they have the legal rights to pursue the publishing options in which they are interested.

One option for faculty members who are copyright owners of works they have created is to place the works in the institutional repositories that many academic libraries now operate. The Canadian Association of Research Libraries (CARL) encourages the establishment and use of such collections:


An institutional repository (IR) is a digital collection of an institution’s intellectual output. IRs are a key infrastructure component in the digital environment because they provide better access to our digital assets and they ensure that digital objects are managed appropriately.<sup>15</sup>

Such repositories are said to “make accessible the knowledge generated by academic institutions.”<sup>16</sup> However, under s.3(1) of the *Copyright Act*, only the holder of copyright has the right to publish a work or communicate the work to the public by telecommunication ((s.3(1) (f); i.e., to make it publicly available on the Internet). Anyone else, whether an individual or an institution, who publishes or makes it available publicly on the Internet infringes (see s.27(1), primary infringement, and s.27(2), secondary infringement). If a faculty member who does not hold copyright to her or his work places that work in an institutional repository that makes works available on the Internet, not only does that faculty member infringe the rights of the copyright owners, but the institution that manages the repository does as well. For example, if a faculty member has already signed an agreement in which the copyright ownership was assigned to the publisher as part of the publication arrangements, permission from the publisher would be necessary to disseminate the work by other means, including through an institutional repository. Indeed, in a 2005 survey of Canadian repositories, Kathleen Shearer reported that “[c]opyright and intellectual property concerns . . . contribute to low participation rates by faculty.”<sup>17</sup>

Libraries that operate institutional repositories should develop policies and procedures to ensure that, in the case of each deposit, the faculty member or student has the rights necessary to allow his or her work to be made available to the public through the repository or that copyright clearance to do so is obtained from the rights holder(s). Librarians can provide information to help faculty members and students retain the necessary specific rights to permit deposit in their institutional repositories while still having their works published elsewhere. As suggested by the Association of Research Libraries, “outreach to scholarly society leaders is one important activity [that] libraries can do to promote positive change that advances the scholarly communication system through new research and dissemination models.”<sup>18</sup> New publishing models such as institutional repositories and open access journals are options available for those scholars who own copyright to their works. Librarians can offer suggestions to faculty members and students to assist them in their copyright negotiations with commercial publishers in order to allow both commercial publication and institutional repository deposit.



## Conclusion

Copyright law is a balancing act between protecting the rights of creators and protecting the rights of users. A fundamental mandate of libraries is to protect intellectual freedom and advocate for fair and equitable access to information. Copyright issues are always evolving, and librarians can be one of the key resources in this area for universities and colleges. 

**Sam Cheng** ([samccheng@yahoo.ca](mailto:samccheng@yahoo.ca)) was the Copyright Officer at Nova Scotia Community College from 2011 to February of 2014. She has recently relocated to Oakville, Ontario and has become a member of the OLA Copyright User's Group.

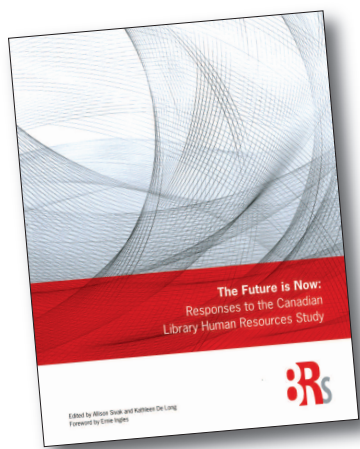
**Christina Winter** ([Christina.Winter@uregina.ca](mailto:Christina.Winter@uregina.ca)) is the Copyright Officer at the University of Regina. Previously she worked in a special library at a law firm in Toronto.

The authors wish to thank Margaret Ann Wilkinson for her insights that contributed to this article.

## Notes

1. Association of College and Research Libraries, *Information Literacy Competency Standards for Higher Education* (Chicago: Association of College and Research Libraries, 2000). Available at <http://www.ala.org/acrl/sites/ala.org/acrl/files/content/standards/standards.pdf>.
2. Ibid.
3. Gail Clement and Stephanie Brenenson, "Theft of the Mind: An Innovative Approach to Plagiarism and Copyright Education," in *Common Ground at the Nexus of Information Literacy and Scholarly Communication*, edited by Stephanie Davis-Kahl and Merinda Kaye Hensley (Chicago: Association of College and Research Libraries, 2013): 45-74, and Tony Horava, "Copyright Communication in Canadian Academic Libraries: A National Survey," *Canadian Journal of Information and Library Science* 34, no. 1 (2010): 1-38, doi: 10.1353/ils.0.0002.
4. Margaret Jane Radin, *Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law* (Princeton: Princeton University Press, 2012): 3-18.
5. See Victoria Owen, Robert Tiessen, Leslie Weir, Davina DesRoches, and Wanda Noel, *E-Books in Research Libraries: Issues of Access and Use*, (Canadian Association of Research Libraries [CARL], April 2008), available at [http://carl-abrc.ca/uploads/pdfs/copyright/carl\\_e-book\\_report-e.pdf](http://carl-abrc.ca/uploads/pdfs/copyright/carl_e-book_report-e.pdf), and the Council of Pacific and Prairie University Libraries (COPPUL), "Checklist for Negotiating License Agreements," April 2013, accessed March 19, 2014, [http://www.coppul.ca/licenses/lic\\_checklist.pdf](http://www.coppul.ca/licenses/lic_checklist.pdf). As identified in the CARL report, user rights permitted under the *Copyright Act* can be modified by a contractual agreement between the library and publisher. COPPUL also stated that "if a license agreement expressly acknowledges fair dealing or is silent as to how license materials may or may not be copied, fair dealing will apply, but if the parties agree to limitations on use that are more restrictive than would otherwise be permitted under copyright law, the licensee cannot later claim broader rights under the law."
6. See <http://copyrightliteracy.wordpress.com/about/>.
7. See <http://digitaltattoo.ubc.ca/>.
8. In *CCH Canadian Ltd. v. Law Society of Upper Canada*, [2004] 1 SCR 339, the Court mentioned that fair dealing in sections 29, 29.1, and 29.2 of the *Copyright Act* is a user's right and provided these factors to help users to decide whether a dealing is fair: (1) the purpose of the dealing, (2) the character of the dealing, (3) the amount of the dealing, (4) the nature of the work, (5) available alternatives to the dealing, and (6) the effect of the dealing on the work.
9. See [http://cldd.athabascau.ca/video/fair-use/FairUse\\_Main.html](http://cldd.athabascau.ca/video/fair-use/FairUse_Main.html).
10. See <http://library.queensu.ca/fairdealing/index.php>. Queen's University Library's tool is adapted from the American Library Association's Fair Use Evaluator: <http://librarycopyright.net/resources/fairuse/>.
11. It is important for librarians and end-users to read the Terms of Use or license information of any open access and Creative Commons-licensed material they want to use. There may be restrictions on what users can do with the material. Creative Commons, for instance, does not replace copyright, since creators can select which rights they want to waive and which rights they want to reserve by choosing a license type. There are six Creative Commons license types, all of which require attribution to the creator, and other restrictions may be combined: share-alike, non-commercial uses only, and/or no derivatives. More information on Creative Commons is available on the organization's website: <http://creativecommons.ca/>.

12. Horava, "Copyright Communication in Canadian Academic Libraries: A National Survey," at 24. See note 3, above.
13. In *Dolmage v Erskine*, 2003 CanLII 8350 (ON SCSM), it is affirmed that university faculty members own copyright to the works they produce unless there is a written agreement between the faculty member and university to the contrary. Under the Copyright Act, the author of a work shall be the copyright owner (s.13(1)) unless the work has been created in the course of employment, in the absence of an agreement to the contrary (see s.13(3)), or unless the author or owner has assigned the rights, either in whole or in part, to someone else (see s.13(4)). In *Dolmage v Erskine*, the court stated that "[t]he academic exception is pervasive in the university community. It has been thoroughly understood and accepted for a very long time, including the 80 years the Act has been in force . . . . Academic exception is an implied 'agreement to the contrary' within the meaning of s.13(3)."
14. In the college environment, there are examples both of college faculty members owning their copyrights and of institutions owning copyright in college faculty members' work.
15. Canadian Association of Research Libraries, "Canadian Institutional Repositories," accessed March 19, 2014, <http://www.carl-abrc.ca/ir.html>. CARL also publishes here a list of institutions that host repositories.
16. University of Saskatchewan Libraries, "Open Access" (LibGuide), accessed March 19, 2014, <http://libguides.usask.ca/content.php?pid=178565&sid=1584233>.
17. Kathleen Shearer, "The CARL Institutional Repositories Project: A Collaborative Approach to Addressing the Challenges of IRs in Canada," *Library Hi Tech* 24, no. 2 (2006): 165, doi: 10.1108/07378830610669547.
18. Association of Research Libraries, "Scholarly Communication," accessed March 19, 2014, <http://www.arl.org/focus-areas/scholarly-communication>.



## ***The Future is Now: Responses to the Canadian Library Human Resources Study (2012)***

**Editors:**

Kathleen De Long and Allison Sivak

**Contributors:**

Alvin Schrader, Deborah Hicks, Donna C. Chan, Marianne Sorensen, Lyn Currie and Carol Shepstone, Islay McGlynn and Joan Cherry

Available from CLA

<http://www.cla.ca/source/Orders/index.cfm?Section=Shop>

ISBN: 978-0-88802-337-7

CLA Member rate \$ 39.95

Non-CLA Member rate \$ 43.95



## Mischief Managed: A Brief Introduction to Data Management

“Data” is a term that has traditionally been used by those working in the social, computer, and applied sciences, rarely crossing over to be a part of day-to-day discussions. But in recent years, the term has been enjoying plenty of time in the spotlight; the use of “data” and especially “big data” is widespread, and has even become common enough in everyday parlance that the folks at the Oxford English Dictionary (OED) added the term “big data” to this year’s edition.

### data

Line breaks: data

Pronunciation: /ˈdɛtə ˌ /

NOUN

[MASS NOUN]

1 Facts and statistics collected together for reference or analysis:

*‘there is very little data available’*

### big data

Line breaks: big data

NOUN

[MASS NOUN]

Computing

Data sets that are too large and complex to manipulate or interrogate with standard methods or tools:

*‘much IT investment is going towards managing and maintaining big data’*

Oxford Dictionaries Online at  
[www.oxforddictionaries.com](http://www.oxforddictionaries.com)

What’s the difference between ordinary data and big data? If you ask data scientists what size a data set has to be to be considered “big,” they are reluctant to give you a specific number. Using the OED’s definition, a data set can be considered big when spreadsheet software such as Excel is no longer able to load it. But given that Excel can only handle data sets smaller than 1,048,576 rows by 16,384 columns, most data users would say that’s hardly big at all. Big data is created by very large information systems, such as those analyzing

health data; another example is the amount of data generated by e-government systems, where a large volume of data is associated with each citizen and is derived from many different databases, each with many variables. The unwieldy volume of data surrounding these systems is what has become so newsworthy, especially since we are generating data faster than we are learning to manage it.

### The Growth of Digital Data

Of course, data has been around as long as any other kind of information, first in print and more recently in digital form. But as Internet technology pushes almost everything we do into a digital format, the amount of data being generated is growing at an incredible rate. As computer technology tracks and records just about everything that happens in the world, these information systems create a lot of data. Like any kind of information, data needs to be managed, but the approach to managing data is different from managing other types of information because data sets have unique characteristics in comparison to text-based information.

### How Data Is Different

As information professionals, we deal with lots of different types of information every day. While traditionally we have worked primarily with information in the physical form of books, serials, photographs, and more, over the past two decades we have become adept at managing these resources in digital form. We’ve embraced metadata as a way to describe digital files so that these resources can be identified and discovered.

When it comes to data, there are some unique challenges. Text-based resources have long had the advantage of being easily described according to title, author, publication date, and so on. When print resources convert to digital, we can use the same descriptors in the metadata. But data sets are more complex for several reasons. First, data files often do not have a stated title or author; the creation date is dependent



upon the software that created the file, and filenames that don't follow a naming convention can be confusing to interpret. Determining the nature of the file's content can be difficult, especially if the variable names are unclear.

Second, data is often not in a "published" state, where it is considered a finished work; rather, various authors update and change it, and the change history can be lost if it is not documented.

To address these challenges, data files are often paired with a data definition file, also known as a codebook. This codebook file is a text document that defines the content within the data file, including its author, a creation date, and an explanation of the variables. Statistics Canada uses data definition files in this way; when StatCan publishes the data files from a survey, such as the census, each data file has a codebook that provides detailed information on the survey such as when the survey was taken, the questions asked, the number of responses, the location, sex, age, and so on of the respondents, and more. As our national statistical agency, StatCan has developed many standards, such as the use of codebooks, to help us understand and use the data sets they provide.

### Defining Data Management

Organizations that acknowledge the value of the data they produce and use have data management plans. A data management plan describes a set of policies, practices, and procedures to manage the data lifecycle. In an academic environment, data management plans pertain to the management of research data in terms of how data created at the university will be processed, analyzed, preserved, accessed, and reused.

Among the leaders in research data management is the UK Data Archive,<sup>1</sup> located at the University of Essex. Responsible for acquiring, curating, and providing access to the UK's largest collection of social and economic data, the UK Data Archive readily shares its best practices and consistently offers data management training to anyone who requests it. The figure and table below are adapted from the UK Data Archive's research data lifecycle, which describes the various stages of data creation, use, preservation, and reuse.

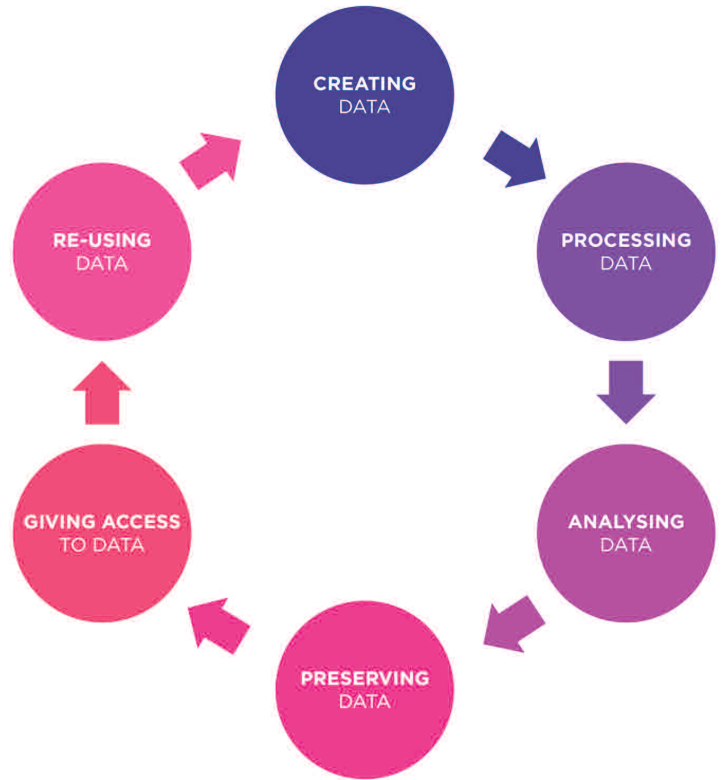


Figure 1: The Research Data Lifecycle

The model developed by the UK Data Archive is highly detailed; the key is for each organization to consider each step and to develop a data management plan that fits accordingly. To ensure success, multiple stakeholders must participate in designing the data management plan, including the data creators, users, and managers as well as the people responsible for policy and governance.

### The Role of the Data Librarian

Academic researchers have been generating, using, and reusing digital data sets since the 1960s: at that time, data was housed on magnetic tape, which would be loaded onto a mainframe computer for researchers to use. By the 1990s, data was available on CD-ROMs, and in the last decade via the Internet and the Web, providing access for a wider variety of people. Today, all research universities in Canada have at least one data librarian, and as the demand in data usage increases, it's expected that more resources will be provided for data management. The majority of data librarians support the social sciences by providing access to a wide range of data from social surveys, censuses, and opinion polls from around the world.

*Mischief Managed continued on page 17*



## Research Data Management at Concordia University: A Survey of Current Practices

### Background

In the summer of 2011, I was a newly minted data librarian looking for an interesting research topic to keep me busy during my upcoming sabbatical year. When I came across the eye-opening awareness kit *Research Data: Unseen Opportunities*, I was only vaguely aware of the issue of research data management.<sup>1</sup> I had heard about the increasing amount of data being produced by researchers and the difficulties associated with curating that data, but I did not imagine it would so soon become a major discussion topic in academia. At that time, it was not yet typical for academic friends to begin conversations by lamenting the *data deluge* or casually mentioning the latest news article on *big data*. How things have changed in just two years! Or have they? We talk about these things a lot, and they have become buzzwords, but what are we actually doing about them? As I was soon to find out, not that much.

Having found a fascinating research topic, I decided that a good way to investigate what was really happening in terms of preserving, documenting, and sharing research data sets would be to devise a survey of current practices of social sciences faculty at my institution, Concordia University. The aim of my survey was twofold: first, to assess what researchers were doing with the data they generated and whether they were interested in sharing it with the academic community; second, to determine what types of research data management services the library could offer. Not wanting to reinvent the wheel, I based my survey and the post-survey interviews on the Digital Curation Centre's *Data Asset Framework*.<sup>2</sup> I decided to survey every full-time faculty member in four departments: Geography; Planning and Environment; Political Science; Psychology; and Sociology and Anthropology. After presenting my project at departmental meetings, I launched my survey. A few reminders later, I was quite happy to have obtained a 37 percent participation rate (forty-one participants out of 111 contacted faculty members).

### Survey results

The vast majority of respondents (92 percent) reported having research data sets. My definition of data was very broad and included qualitative results as well as quantitative data. The data took many forms, including SPSS and Excel spreadsheets, Word documents, recordings of electrical brain activity (EEG), and also, to my surprise, a lot of paper files such as copies of archival documents and handwritten notes from field research. The variety of data holdings underscored the challenges that would be faced when devising research data management (RDM) policies and services. And I was only surveying social sciences departments; had I included science disciplines, the range of data sets would have been even more diverse.

### Data backup and preservation

One of the main goals of my research was to evaluate current practices to ensure data safety and archiving. The findings were quite striking. As many as 43 percent of respondents did not have any plan for backing up their data and only did it on an ad hoc basis (30 percent created a weekly backup, 12 percent backed up daily, and 15 percent only performed a backup on a monthly basis). When asked where they archived their data at the end of their research, an overwhelming majority answered that they used a combination of PC hard drives (85 percent), CDs or DVDs (53 percent), and flash drives (35 percent). Only a minority used safer methods such as departmental servers (15 percent) or the institutional repository (6 percent). Not a single researcher used a dedicated data repository like ICPSR. These results are hardly surprising, as not many data archiving options are available to researchers in most disciplines, and very often the existing repositories are not well known. More importantly, in order to archive data sets in any meaningful way, a thorough curation process is required, involving such activities as data cleaning, creation of appropriate metadata, production of a complete documentation package (data dictionary, description of the methodology, etc.), and creation of data files in an open, interoperable, and sustainable

format. On a positive note, 68 percent of respondents indicated that they were interested in long-term archiving of their datasets in an appropriate repository. What is lacking is not the desire to preserve the data, but rather the means to do it.

## Data-sharing

In the spirit of the open access and open data movements, there is a growing emphasis on making research data, which is generally created with public funding, available to the research community and, where possible, the general public. The Concordia faculty who answered my survey or were interviewed were generally in agreement with this view. Only 15 percent answered that they preferred not to share their data. On the other hand, of the 85 percent of respondents who indicated a willingness to share their data sets, only 31 percent said that their data should be made public. Others preferred to make data available only in a controlled environment (24 percent) or to personally manage access to their data (30 percent). Researchers have several legitimate concerns regarding data-sharing. To cite a few, social science faculty want to preserve the confidentiality of their research subjects, and they are concerned that other researchers might draw inappropriate conclusions based on their data because of a lack of understanding of the methodology or of the context of the original research. There are also concerns regarding licensing and copyright issues. Nevertheless, these results are encouraging: researchers understand the importance of sharing data sets and, given the appropriate parameters, most of them would be happy to do so.

## Research data management services

Data curation is a complicated and time-consuming process and, from the perspective of busy researchers who are primarily concerned with publications and teaching, it may not be a priority. In this context, I was curious to see if there would be some interest in new services aimed at RDM. One of the survey questions listed five possible services that the library or other university offices could offer: 1) the creation of a virtual research environment (i.e., a type of network drive that would allow members of a research team to share and manage their data sets while ensuring data security); 2) help with data archiving; 3) help with metadata creation; 4) help with data management plans; and 5) a data management workshop for graduate students.


A majority of respondents expressed interest in each one of those services, the most popular being the workshop and the creation of services to provide assistance with data management plans and data archiving.

## Conclusion

My survey sample was limited to social sciences faculty from a single institution, and therefore I cannot draw conclusions that could be generalized to all Canadian researchers from all disciplines. Nevertheless, the results and the evidence collected during a series of post-survey interviews were fairly clear: there is a growing need for RDM but not much is being done about it so far. Researchers are aware of the issues and some are actually alarmed about the possibility of losing their data sets due to a lack of archiving resources. There seems to be a significant amount of goodwill towards data-sharing, but the absence of specific funding, the lack of time to devote to RDM activities, and the scarcity of technological infrastructure combine to make this a difficult endeavour.

Canada, and most particularly Quebec, is lagging behind other nations in the development of RDM policies, services, and infrastructure. In the United States and the United Kingdom, research funding agencies are requiring researchers to submit data management plans and commit to data-sharing within a relatively short period after the completion of their research projects. These policies have spurred important developments such as the creation of institutional data repositories, the establishment of expertise centres like the Digital Curation Centre (U.K.), and the Australian National Data Service. In the United States, disciplinary data repositories like ICPSR, institution-based initiatives such as Purdue University Research Repository (PURR), and collective projects like Data-PASS are at the forefront of the RDM movement. There are also interesting initiatives in Canada, like Scholars Portal's Dataverse or the Canadian Polar Data Network, but for most university libraries research data management is just beginning to show up on the radar screen. Academic libraries are only one of several stakeholders in this emerging field, and they cannot be expected to provide solutions to all of the numerous challenges posed by the data explosion. However, if they want to play a role that reflects their well-established strengths in information and knowledge organization, it is time for



them to start planning, in as collaborative a way as possible, for the development of RDM services that will address some of the researcher's needs. 

**Alex Guindon** ([alex.guindon@concordia.ca](mailto:alex.guindon@concordia.ca)) is the GIS and Data Services Librarian at Concordia University. During his sabbatical leave in 2012-13, he conducted a survey of current faculty practices in regard to research data management. Alex is a graduate of Université de Montréal (MLIS) and also holds a Master in Political Science from UQAM.

---

### *Mischief Managed* continued from page 14

In addition to providing access to the data sets and associated documentation, data librarians provide instruction to students and researchers on how to use the data with various statistical software packages, provide data reference services, and manage the licensing for the data collections they procure for their institutions. More and more, data librarians are also playing a leading role in assisting researchers with archiving their research data so that it can be preserved, updated, and in some cases shared with colleagues. Most academic institutions in North America host research repositories, and including the data associated with the published research is becoming more common.

With data librarianship being a recent area of specialization, we are fortunate to have an association of colleagues that work together closely to develop solutions and create best practices. The International Association for Social Science Information Services & Technology (IASSIST) was founded in 1974;<sup>2</sup> the first meeting of the association was held in Toronto, Ontario, in August of that year, and was attended by approximately sixty social science researchers and librarians. Today, IASSIST is an international organization of data scientists, data librarians, and social science researchers; members are highly engaged in multiple areas of data management practice including archiving, preservation and curation, data citation, metadata and


### Notes

1. Canadian Association of Research Libraries, *Research Data: Unseen Opportunities*, 2009, accessed March 11, 2014, [http://www.carl-abrc.ca/uploads/pdfs/data\\_mgt\\_toolkit.pdf](http://www.carl-abrc.ca/uploads/pdfs/data_mgt_toolkit.pdf).
2. Digital Curation Centre, "Data Asset Framework," accessed March 10, 2014, <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/data-asset-framework>.

---

standards, statistical literacy and training, and data access and open data. This summer, IASSIST is celebrating its fortieth anniversary by hosting its annual conference in Toronto, which promises to be a successful gathering of data users from around the world.

### Thinking Forward

While some may feel alarmed at how quickly the mass of data is growing, the good news is that the information profession is well-positioned to tackle data management and its associated challenges. Data, though a unique form of information, can be successfully managed by combining trusted information organization techniques with an understanding of the data lifecycle. 

**Kimberly Silk** ([kimberly.silk@rotman.utoronto.ca](mailto:kimberly.silk@rotman.utoronto.ca)) is the Data Librarian at the Martin Prosperity Institute, part of the Rotman School of Management at the University of Toronto. When she's not playing with data sets, Kim enjoys swimming, biking, and spending time with her family.

### Notes

1. Visit the UK Data Archive's web site at <http://www.data-archive.ac.uk/>.
2. To learn more about IASSIST and the upcoming conference, visit the IASSIST web site at <http://www.iassistdata.org/>.



## Is eDiscovery the New Frontier in Information Management?

### Introduction

You have probably heard the term “eDiscovery” around offices, legal departments, and information agencies throughout North America. But what does this buzzword actually mean? And how will it change the involvement of information professionals in the legal field?

Any introduction to eDiscovery is incomplete without a brief explanation of traditional discovery in the legal context. Discovery refers to the process of compiling and classifying records, correspondence, and other ephemera for production at trial. However, discovery principles can be broadly applied to internal and external audits, investigations, or even general recordkeeping practices. eDiscovery, then, is this process for the digital world. eDiscovery specifically deals with digitized and born-digital content (collectively referred to as “electronically stored information,” or ESI) and the procedures used to preserve, collect, maintain, and describe this data.

Sound familiar? It should. eDiscovery adapts the core principles of records management to a legal framework and utilizes many of the same skills information professionals rely on in their various occupations. The importance of eDiscovery will continue to grow as electronic transactions and interactions become the norm rather than exception, and may even present a new field in which information professionals can market their skills.

The evolution and standardization of eDiscovery regulations have increased the demand for content management systems in agencies that have not typically utilized such software, and have created an increased focus on electronic records and data management in all areas. More environments have instituted records management and eDiscovery procedures than ever, making information professionals who use electronic records management to accurately preserve, restore, and retrieve important data increasingly desirable.

### The Development of eDiscovery in the United States and Canada

The United States utilizes two main court systems: state and federal. While eDiscovery rules and principles can vary widely from state to state, in the federal court system rules governing eDiscovery are more uniform. In 2006, the Federal Rules of Civil Procedure (FRCP) were amended to govern eDiscovery as well as traditional discovery. Some of the most salient changes are the implied inclusion of ESI in the definition of “documents,” unless a discovery request specifically differentiates between traditional paper documents and electronic data. As articulated in the FRCP, the best practice for requesting ESI in the discovery process is to request information stored “in any medium” in order to cover both existing technologies and adapt the request to potential future developments. The 2006 amendment also requires ESI to be produced in its most useful or readable format in order to prevent gamesmanship or even an inadvertent slowing of the discovery process by the production of ESI in a format that is unreadable or unusable by the other side.<sup>1</sup> Given the rapid evolution of technology, the FRCP will be amended to include further updates to eDiscovery rules in 2014.

In Canada, the 2006 amendment to the FRCP only served to underline the need for uniform eDiscovery rules in the Canadian legislative system. In 2008, The Sedona Canada Working Group 7 (commonly referred to as “Sedona Canada”)—a coalition between Canadian legal authorities and a legal think tank based in Sedona, Arizona—released its guidelines for the Canadian eDiscovery process. The Sedona Canada Principles Addressing Electronic Discovery were developed to create a national framework for preservation and production of ESI that would be compatible with existing principles (particularly the FRCP in the United States), but also address needs that were uniquely Canadian.

Most significantly, the Sedona Principles follow Canadian rules for production. In all provinces except Quebec, it is the responsibility of each party to produce

all relevant information during the discovery process, while the general rule in the United States is to produce only information that the opposing party has specifically requested. Although laws governing production in Quebec are more similar to those in the United States, in practice the Quebecois follow the same production principles as other Canadian provinces.

The Sedona Canada Principles elaborate twelve basic rules to govern the eDiscovery process in Canada, and a paraphrased version of the rules is as follows (a full version of the Principles, along with explanations of each, can be found at <https://thesedonaconference.org/download-pub/71>):

1. Electronically stored information is discoverable.
2. Parties should ensure that steps taken in discovery are proportionate and not unduly burdensome, and that the production of ESI does not cause an unreasonable delay in proceedings.
3. As soon as litigation is anticipated, parties should take reasonable and good-faith steps to ensure that potentially relevant ESI is preserved.
4. Parties and their attorneys should meet and confer as soon and as often as possible regarding identification, preservation, collection, review, and production of relevant ESI.
5. Parties should be prepared to produce ESI that is relatively accessible in terms of cost and burden.
6. A party cannot be required to search for or collect residual or deleted ESI without a court order.
7. A party may satisfy its obligation to preserve, collect, review, and produce ESI in good faith by using electronic tools and processes such as data sampling, searching, or using selection criteria to determine potentially relevant ESI.
8. Parties should agree as early as possible on the format in which ESI will be produced.
9. Parties should either agree independently or seek judicial direction on measures to protect privileges, privacy, trade secrets, and other confidential information from being produced with electronically stored information and data.
10. Parties should respect the rules of the forum in which the litigation takes place during the eDiscovery process.
11. A court will consider sanctions against a party that fails to meet its obligation to preserve, collect, review, or produce ESI unless the party can prove that its failure was unintentional.

12. In most cases, the reasonable costs of preserving, collecting, and reviewing ESI will be borne by those producing it.

The Sedona Principles highlight the importance of core records management practices, such as retention schedules, preservation, and classification, to the eDiscovery process. While these developments in eDiscovery may seem obvious or inconsequential to the casual observer, they are indicative of the broader movement toward good records management and eDiscovery practices in North America.

The compatibility between eDiscovery and traditional information and data management may highlight an emerging field for information professionals looking to apply their skills outside the traditional data management setting.

### **eDiscovery and You: An Information Professional's Guide**

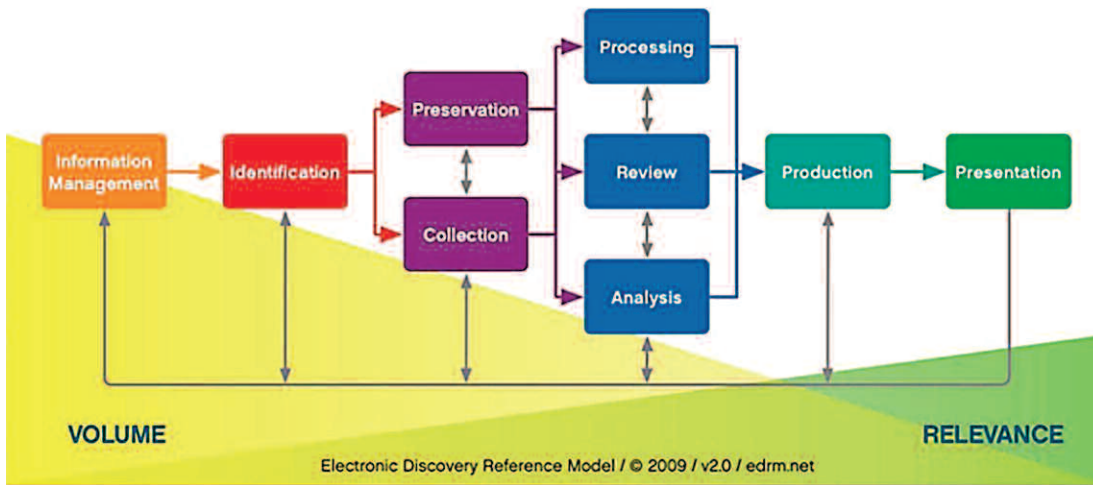
As the role of eDiscovery continues to grow in North American legislative systems, tools that have been utilized by information professionals to maintain intellectual and physical control over records are increasingly adapted to a legal environment.

One group that brings core records management principles to the new world of eDiscovery is the Electronic Discovery Reference Model (EDRM). The EDRM describes itself as “a coalition of consumers and providers working together since 2005 to create practical resources to improve eDiscovery and information governance,” whose members include Microsoft, IBM, and prestigious law firms such as DLA Piper and the Canada-based Osler, Hoskin & Harcourt LLP.<sup>2</sup>

The EDRM is intended to create a framework for both seasoned practitioners and those new to the eDiscovery process to accurately store, manage, and retrieve relevant information. The model itself will seem readily familiar to those used to dealing with retention schedules or other data management schema. While this article provides only a brief introduction to the EDRM, those seeking more information can find the full model at <http://www.edrm.net/resources/guides/edrm-framework-guides>. The steps in the process, such as implementing a preventative information management system, processing and converting data into its most usable format, analyzing patterns in data, and disposition of data where applicable, are all part of



Electronic Discovery Reference Model



Even agencies that may never expect to participate in litigation are beginning to understand the full importance of implementing eDiscovery and records management practices in order to control an ever-growing expanse of digital information. As a result, information professionals will continue to be in demand in this field, even if many employers may not realize that “librarians” have the exact skill set they are looking for. Law firms,

an information manager’s daily routine. In fact, this is why trained information professionals may become the most valued members of eDiscovery teams.

In 2013, EDRM published a framework intended to help companies acquire the best talent for handling eDiscovery. Many of the skills referenced in this model are ones taught in schools of information, particularly in Records Management programs. Information professionals are trained to identify important data, transfer it into a usable format, and make it searchable for later retrieval. These are the exact skills most prized by companies that participate in eDiscovery.

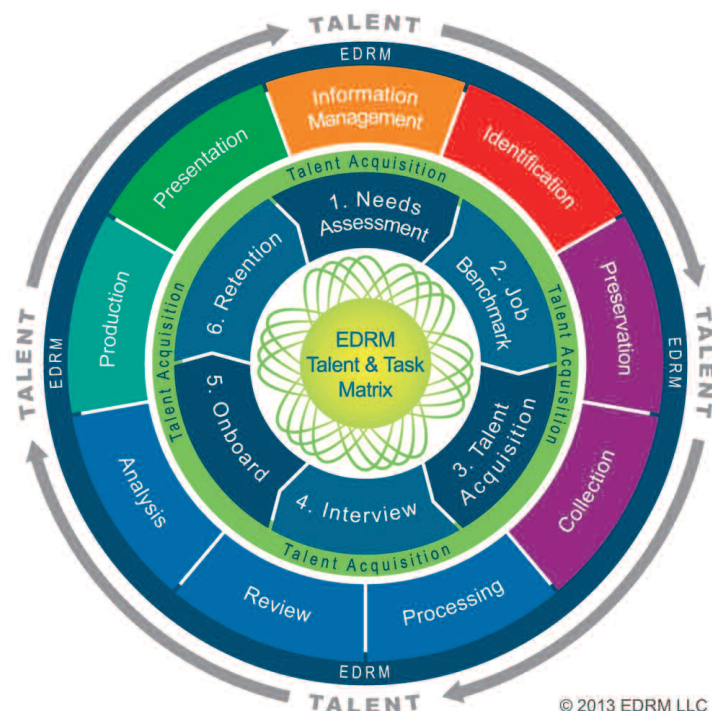
in-house legal departments, and contract eDiscovery companies all need individuals to design and implement information management and eDiscovery systems.

If you are an information professional who has never considered the legal field as a career option, you may be doing yourself a disservice. As eDiscovery rules and regulations continue to expand and evolve, it is up to information professionals to understand their own importance to this industry and market themselves as the most desirable candidates for eDiscovery positions. 🌊

*Ariana Ross (ariana.ross@mail.mcgill.ca) is a recent McGill MLIS graduate with a concentration in Archival Studies.*

Notes

1. “Title V: Disclosures and Discovery. Rule 34: Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes,” *Federal Rules of Civil Procedure*, December 1, 2009, 52, accessed March 10, 2014, <http://www.uscourts.gov/uscourts/rulesandpolicies/rules/cv2009.pdf>.
2. “About EDRM,” accessed March 10, 2014, <http://www.edrm.net/what-is-edrm>.





## Following the Trail of Breadcrumbs: Your Health from Data to Decisions

### Did you get your flu shot this year?

**Perhaps you're wondering** if you really *need* the flu shot. You've probably heard all kinds of anecdotal evidence from people suggesting that, following their flu shot, they became ill. Or perhaps the exact opposite: your neighbour didn't get the influenza vaccination and also managed not to get sick. Regardless, those comments are single accounts of an individual's experience, which is not always the best evidence for every person. If you were going to take advice from someone regarding this decision, who should it be? Any of the following possibilities could be influential: government institutions, public health services, your physician, your mom or friend, a blog you found, or even a celebrity. Behind every decision you make about your health, often in collaboration with your physician or other allied health professional, there is hopefully some evidence to guide the process.

In a perfect world, your physician would be able to incorporate an evidence-based medicine approach into your care. This approach is "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."<sup>1</sup> Whether this evidence is elicited from filtered sources (evidence-based summaries or point-of-care tools) or unfiltered sources (the primary literature from randomized control trials to case studies) is irrelevant. It all originated from data procured from individual patients such as yourself.

### Creating and managing data

Health data is accumulated in order to be analyzed to support medically sound decisions. The process is long and involves issues surrounding the privacy of the patient. However, the benefit of obtaining this evidence to help anyone, from policy-makers to patients, make decisions regarding their health and the health of others cannot be underscored enough. It all starts with an inquiry, a gap in knowledge if you will, into what is likely an answerable question. It may originate from a clinician in practice or from the research community surveying the available evidence on a topic.

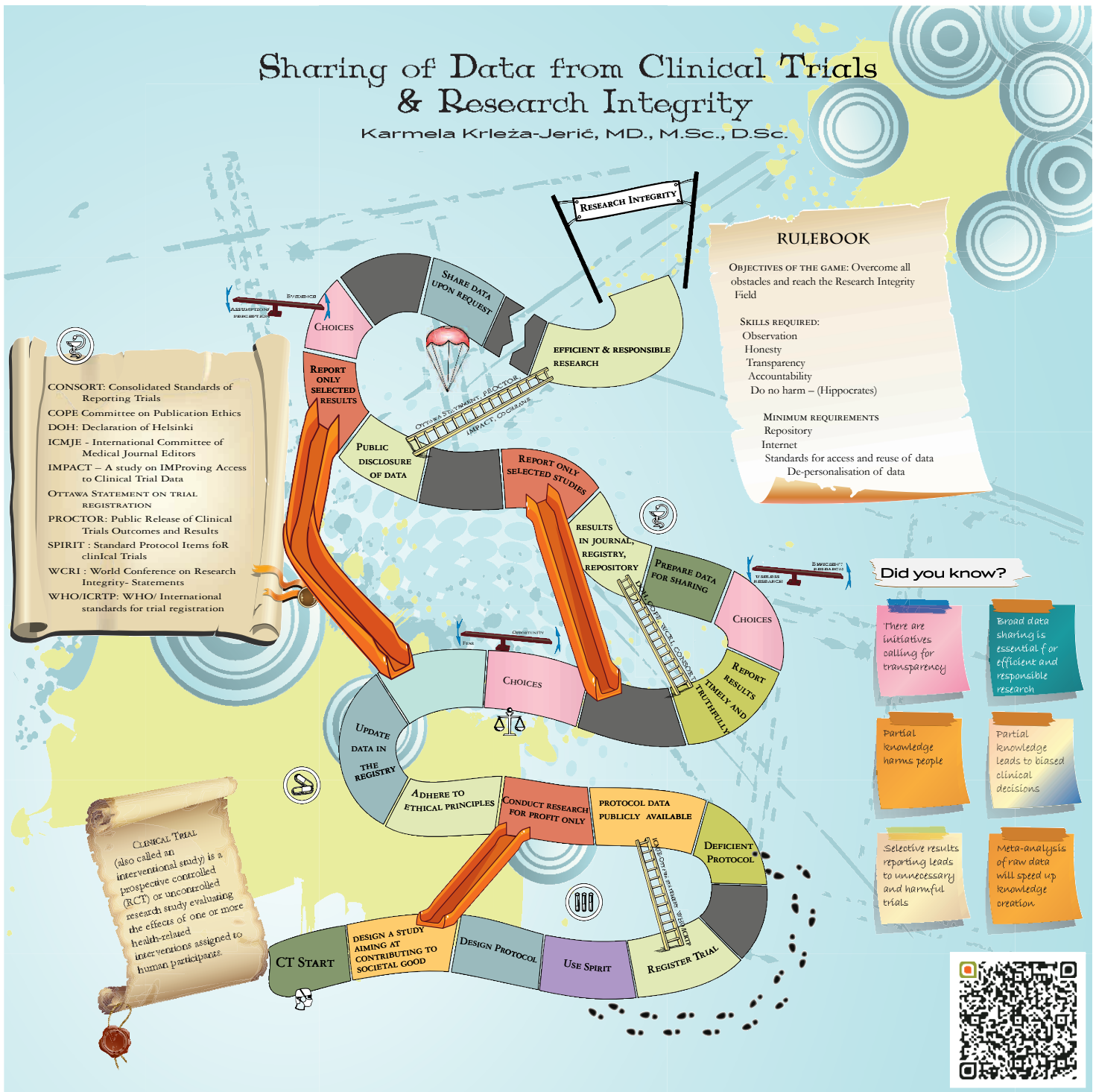
Once the gap is identified, a study is designed and proposals for funding are written. The Canadian Institutes of Health Research (CIHR) is the main funder of health research in Canada. Data needs to be gathered, and a fair proportion of health data comes from clinical trials. A clinical trial "is any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes. Interventions include but are not restricted to drugs, cells and other biological products, surgical procedures, radiological procedures, devices, behavioural treatments, process-of-care changes, or preventive care."<sup>2</sup>

Once funding has been secured, a rigorous ethics approval process must be undertaken and a protocol designed to reduce the risk of bias. If all criteria are met, a research group will initiate a clinical trial, which often involves randomization of the participants. The unique strength of randomization is that, if successfully accomplished, it prevents selection bias in allocating interventions to participants.<sup>3</sup> It is important to understand that because the trial involves people, the number of participants is very likely to change throughout the course of the trial. Some participants may withdraw from the trial, while others may be added. This particular issue results in some trials either not being completed or being completed but not published. The latter case generates what is referred to as "unpublished data," which by itself is likely to be inconclusive; however, if it can be harmonized with another data set, it may be of use. Transparency is highly important throughout this entire process, in order to ensure that the trial is replicable.

Throughout the trial, researchers should be managing the data they collect carefully, through curation. Curation involves a similar process in any discipline, though the dissemination of health data is somewhat complicated due to its confidential nature, which is related to patient privacy. It is thought to be a significant barrier for some researchers, while others have proven that there is a viable solution: de-identification of data, which refers to a set of methods

# Sharing of Data from Clinical Trials & Research Integrity

Karmela Krleža-Jerić, MD., M.Sc., D.Sc.



CREATIVE CONCEPT & DESIGN BY APROPO MEDIA

that can be applied to data to ensure that the probability of assigning a correct identity to any given record is very low.<sup>4</sup>

Only some of the data collected in research projects is reported and manipulated into statistics. The term “missing data” refers to the data available to original researchers but not reported in published abstracts or

full-text publications. In some other disciplines related to health, such as proteomics, authors are asked to make their data openly available at the time of publication, often as supplementary material. This extra level of transparency provides another check and balance to ensure that data is reproducible and reliable. Clinical trial data is rarely made openly accessible, meaning it cannot be reused for further analysis or verified by



another researcher. The net result: not being able to leverage these opportunities increases the cost of research and ultimately means that more participants are required in order to gather data.

### How can we manage all of this data?

Research data management “covers all aspects of handling, organizing, documenting and enhancing research data, and enabling their sustainability and sharing.”<sup>5</sup> It ensures that the research being conducted is replicable and has integrity, while minimizing the risk of data loss. It is also fundamental that this data be accessible in the long-term in order to be reused by future generations.

In fact, many granting agencies and research funders are demanding to see research data management plans at the beginning of research projects. While Canadian funding agencies, such as CIHR, are currently navigating the details for data management, other agencies in countries such as the United Kingdom and the United States have already established guidelines and plans.


Effective data management is required during the initial planning stages of the research project. This planning needs to specify how the data will be managed during the project, as well as how the data will be made accessible upon the completion of the project. Many steps concerning how the data can be shared, along with limitations (such as privacy concerns) and steps to remove these limitations, need to be discussed.

There are several key issues to consider when planning data management:

- Ethical and legal considerations surrounding data, involving participants, colleagues, funders, and institutions
- Designing data management according to the purpose of the research
- Assigning responsibilities according to relevant involvement in the research
- Incorporating data management methods throughout the entire research project as it progresses, reviewing as necessary
- Maintaining consistency of practice implementation methods

Mitigating the issues involved in research data management involves multiple steps, managed by trained researchers diligently trying to ensure that their methodology is rigorous. Figure 1 represents a visual representation of the many challenges and opportunities associated with sharing clinical trial data to ensure research integrity.

### Using your newfound knowledge

At this point, after this discussion of the vast amounts of data behind all health information, you may be thinking about the actress Jenny McCarthy and her anti-vaccination stance. What evidence does she have to support her claims? Equipped with your newfound knowledge, you are now able to make an informed decision regarding interventions for your health (such as the flu shot), as you understand the process involved, from data-gathering to your decisions. Since your health is a shared decision-making process between you and your healthcare provider, you are now well equipped to contribute to that conversation. 

*Lee-Anne Ufholz, BSc, MLIS (Lee-Anne.Ufholz@uOttawa.ca) is the Director of the Health Sciences Library at the University of Ottawa and the Vice-President of the Canadian Health Libraries Association (CHLA). She has worked in academic libraries, special libraries and archives.*

*Lindsey Sikora, BSc, MIST (lindsey.sikora@uottawa.ca) is the Health Sciences Research Liaison Librarian at the University of Ottawa, in Ottawa, Canada, and is the President of the Library Association of the National Capital Region (LANCR). She has worked collaboratively with many research groups on scoping, rapid and systematic reviews within the areas of medicine and health sciences.*

*Acknowledgments: many thanks to Dr. Karmela Krleza-Jeric for kindly agreeing to share Figure 1.*

### Notes

1. David L. Sackett, William M. Rosenberg, J. A. Muir Gray, R. Brian Haynes, and W. Scott Richardson, “Evidence Based Medicine: What It Is and What It Isn’t,” *British Medical Journal* 312, no. 7023 (1996), 71-72.  
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2349778/pdf/bmj00524-0009.pdf>.
2. World Health Organization, “Clinical Trials,” 2014, accessed March 17, 2014, [http://www.who.int/topics/clinical\\_trials/en/](http://www.who.int/topics/clinical_trials/en/).
3. Julian P. T. Higgins and Sally Green, eds., *Cochrane Handbook for Systematic Reviews of Interventions* (Chichester, England: Wiley-Blackwell, 2008).
4. Khaled El Emam, *The Case for De-identifying Personal Health Information* (Ottawa, Ont.: CHEO Research Institute, 2011).
5. UK Data Archive, “Create and Manage Data: Planning for Sharing,” 2014, accessed March 17, 2014, <http://data-archive.ac.uk/create-manage/planning-for-sharing/getting-started>.



## Metadata Management on a Budget

### What is metadata management in a library?

Libraries typically manage metadata records that act as surrogates for electronic and print resources. Each record has a life cycle, during which it is:

- 1) captured (via original transcription or copy cataloguing);
- 2) organized/classified;
- 3) maintained and used; and
- 4) disposed of, when it is no longer an active record.

I think libraries often excel at numbers one and two in that lifecycle. However, things can often fall apart when it comes to number three. While we focus on finding the financial resources for initial acquisition and description, we don't always budget for ongoing maintenance.

However, what happens when someone identifies problems in metadata records? What happens if those twenty thousand new articles were actually catalogued as videos by mistake? When users are searching the catalogue for articles, they might think there is nothing there. Ultimately, this means that the records and resources aren't being used and we're not giving our users the service they expect and deserve.

If we're lucky, we have a dedicated systems librarian who will modify our data. Otherwise, we might pay our ILS vendor or a freelance contractor to do the work. Sometimes, we manually correct it or put it on a to-do list until we can get cheaper labour to manually correct it for us. More often than not, the problems get forgotten and the data remains problematic, as we don't have the money or the time to handle it. These are the typical scenarios that I've encountered time and again in the seven or eight years that I've worked in libraries.

Honestly, though, I think these scenarios can often be avoided. Your data needs to be fixed, but you shouldn't have to pay out the nose to fix your data, nor should you have to change each record by hand or leave it in someone else's task queue. Even though the

ultimate goal is valuable, it's a waste of time and money. You can do it yourself right now for less.

### Enter metadata management tools

First, are you able to export records from your ILS (or other library software)? If so, in what format? MARC? MARCXML? CSV?

Second, are you able to re-import and overwrite your existing records? This will be essential, as we don't want to load in duplicate records by mistake. It has been a few years since I used a proprietary ILS, but exporting and overlaying records is 100 percent possible with an open source ILS such as Koha or Evergreen.

Third, what standards do you use? Once you've exported your data, you will want to review your metadata standards. This is vital since you want to be fixing your records, not making them worse. I've been working with MARC for years in a technical services capacity and I'm still learning new things from this quite inscrutable standard. It is always worthwhile to do your research and check reference sources, especially when you're unsure about a particular detail.

Fourth, what tool will you use? After your refresher course, you will need to find a tool to manipulate your data. If you're a Windows user, you will probably find MarcEdit to be an easy and more than adequate tool. Originally created by Terry Reese at the University of Oregon in 1999 to aid in a massive library database cleanup, the tool is still actively maintained by its creator. It is quite powerful, installation is easy, and the learning curve is negligible. It is the first tool that I used for bulk metadata management, and while I don't use it anymore, it saved my library at the time at least \$2,000 (as quoted by a vendor) when I was able to transform our CSV records into MARC records. Not bad at all, seeing that it only cost the library maybe \$100-200 for my student wages.

Of course, there are a few downsides. First, MarcEdit is designed for Windows. While there is an

installation package for Linux/Other, I haven't heard of anyone successfully installing the program on anything but Windows. Second, as far as I know, it is closed source. Amongst other things, that means that only Terry Reese can work on it. If he were to be struck by lightning or decided to stop working on the tool, that would be the end of MarcEdit updates.

Fortunately, there are many open source programming libraries that exist for the management of metadata, especially for MARC/MARCXML. Code4Lib provides an excellent collection of links to these libraries. Personally, I favour MARC/Perl and File\_MARC (the PHP library) because these are the two programming languages I use the most, and MARC/Perl has a reputation for being a very effective tool.

### **But I'm not a programmer..**

Lack of programming skills is quickly becoming a thin excuse for not using programming languages. More and more disciplines, not just librarianship, are including technical skills in their everyday toolkit. This is partially because we are all working more and more with digital data, and thus need to do more with that data. It's also partially due to the fact that it is actually quite easy to get started programming. It isn't magic. There are plenty of interactive tutorials online that can teach a person some basics, and forums abound to answer more specific questions. Believe me, if you have a question, someone has already asked it and multiple people have already answered it.

In reality, the hardest part about programming on one's own can be getting a software development environment set up for the first time. You might have had to cajole your IT folks into letting you install MarcEdit, so it might be significantly more difficult to get a web server (e.g., Apache or IIS) and a programming language (e.g., PHP or Perl) installed on your workstation.

Now, don't panic when I say *software development environment* or *web server* or *programming language*. These are just tools and you don't need to understand them completely. Baby steps.

While many software developers use Linux, you can actually use Windows fairly easily (if you're so inclined). Try asking your IT department if you can play around on a computer off the network. After all, you're not going to try hacking the organization's servers or offer a new web service. You just want to manipulate some data.

If they're still stonewalling you, try using a computer at home to experiment. Once you're confident that you know what you're doing (mostly), you might have an easier time persuading IT to give you the software that you need. Alternatively, you could use a LiveUSB or LiveDVD to boot up your workstation using an operating system – usually Linux – with all the software libraries you need already on it.

If you don't want to try the "Live" methods, try installing WAMP or Virtualbox or Cygwin at home.

WAMP is a software bundle that includes an Apache web server, a MySQL database, and a programming language (usually PHP). These three tools are all you need to start creating quite powerful programs for manipulating your metadata. (It's worth noting that my first experience with programming, after using MarcEdit, was using the Linux equivalent XAMP.)

Virtualbox is a virtualization software package that allows you to install a "virtual machine" (VM) on your computer. You will need to install a guest operating system such as Linux (which is a good learning experience and will change how you think about open source vs. proprietary software) on this VM, but this will give you a good sandbox for writing and executing your programs. (I still use Virtualbox to create a disposable software development environment when I'm on a Windows computer.)

Cygwin is a Linux emulator that gives you access to a lot of Linux software within Windows. It doesn't require you to fully install Linux (like Virtualbox does), nor does it necessarily require administrator privileges to install (also like Virtualbox does). If you're lucky, you can actually install this one on your work computer without any help from IT.

### **Breathe**

You might be feeling overwhelmed at the moment. That's OK. Change is difficult. However, you have the ability to use these tools. You went to university for at least six years. You are an intelligent person. You are more capable than you know. You are also allowed to make mistakes. You can admit a lack of knowledge and ask for help. Consider reading the short article "The Self-Education of Systems Librarians," by Mark Jordan (see list of resources). While it may be geared towards current and aspiring systems librarians, I'm sure you'll

*Metadata Management continued on page 29*





## Beyond the Hype: Data Management and Data Governance

In both the business and information management worlds, the past couple of years brought much industry buzz over data, Big Data, data analytics, and sexy data scientists.<sup>1</sup> My organization was no exception to this trend, and I had the opportunity to transfer from the information management department at Canada's largest oil and gas company to a new data governance team in January 2013. The year 2013 presented a steep learning curve for this librarian as I learned the ropes of data management, and it is a real privilege to recap by sharing some key takeaways of this discipline with my CLA peers.

### What are data management and data governance?

Special librarians and other information professionals working in corporate environments enable organizations to leverage information so as to reduce corporate risks or maximize business benefits. "Information" is a broad term with a myriad of definitions, but it often connotes the idea of physical or electronic documentation. "Data" also has a variety of definitions, but is often used with reference to raw numerical or textual values stored within a structured database system. As increasingly publicized in the information professions, data, like information, can be rigorously managed in order to derive increased value.

*Data management* is the business function that develops and executes the acquisition, control, protection, delivery, and enhancement of data.<sup>2</sup> Like the broader information management domain, data management is premised on the proposition that data is an organizational asset that has business value. This principle implies that data stores, such as a company's financial system or customer database, can have the same critical value as a physical asset, such as a factory or a petroleum upgrader. Investing in maintaining data assets, through defined data management processes and data quality monitoring and improvements, then becomes an integral part of supporting the organization's core business objectives.

*Data governance* is the organization and implementation of accountabilities for managing data.<sup>3</sup> Data governance includes the roles for managing data as well as the plans, policies, and procedures that control—in essence *govern*—data.

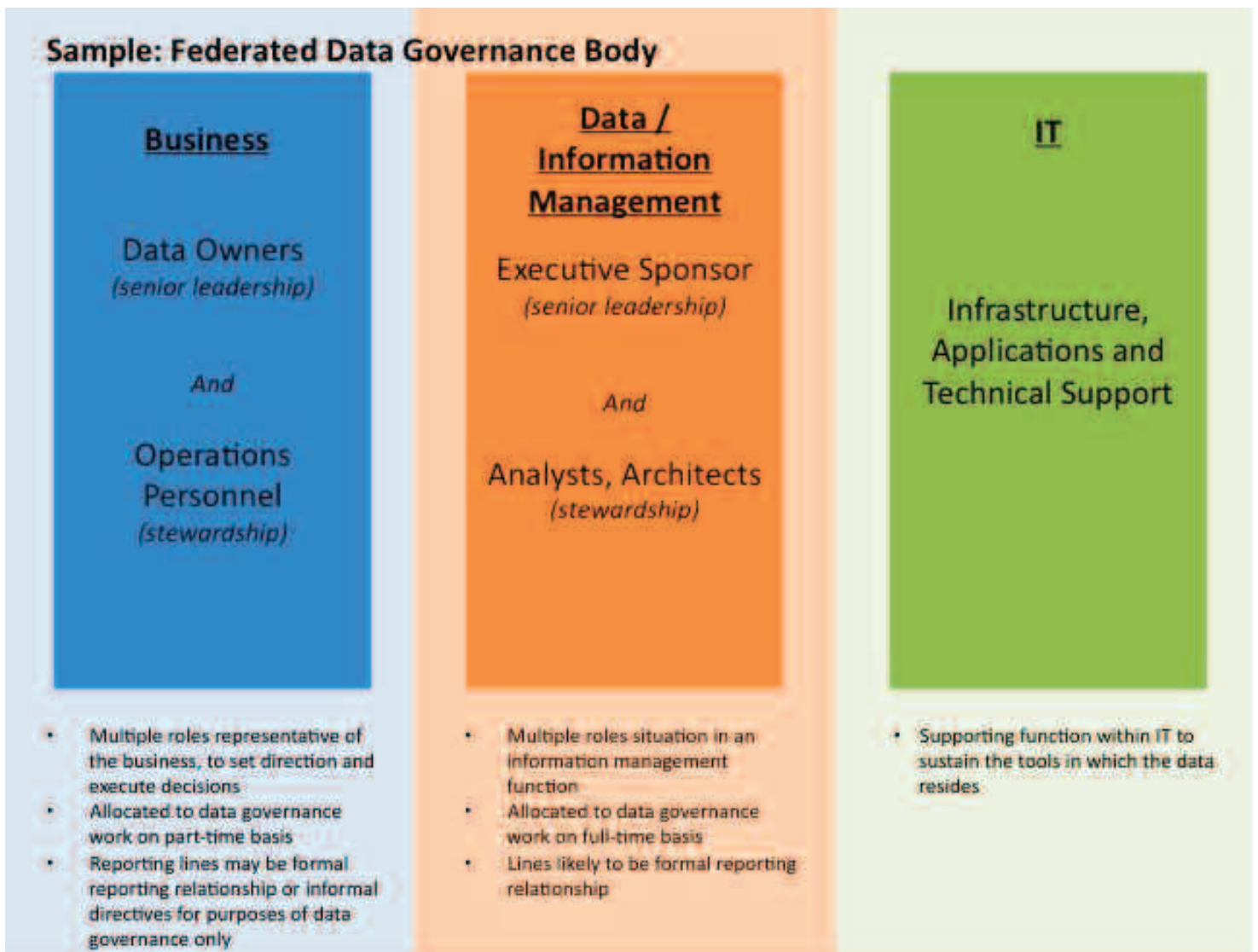
### Business cases for data management

As librarians, we easily comprehend, if not the detailed technical workings of databases, at minimum the principles behind data systems and the foundations of quality within these systems. Having worked with catalogues and circulation databases, we may also have an easier time grasping the consequences of poor data management. When it comes to establishing data governance, understanding an organization's business need with respect to data is a concept transferrable from managing end-user needs with respect to information. Librarians are trained to seek out the connections between a person or group's objectives and the information or data that supports it.

In the corporate sector, whether public or private, not all personnel bring this background to the table. While libraries have, on some level, always functioned as data centres, many other organizations are not so strongly rooted in their systematic information practices and have only recently recognized the value inherent in the information and data that frames their broader objectives.

In my own working experience, I've found that industries focused on physical assets, such as oil and gas or manufacturing, are looking at data management and data governance with novel interest. The business case for "why manage our data" stems from recognizing the pain points in an organization and tying these operational deficiencies back to the data.

For example, assume an industrial scenario in which plant maintenance is continually delayed, causing machine failures to occur. If this happens because inventory systems do not correctly show the spare parts residing in the warehouse, and parts are re-ordered



instead of taken from the supply, this is ultimately a data problem. Allocating resources to examine the very business case for data management then forms the first steps of data governance.

### Where to invest

An early challenge of data management is discovering where to start. Per the concept of “Big Data,” data is now being created and subsequently stored (though not necessarily managed or governed) at an exponential rate. Even the more conservative estimates propose that data grows at a rate of 35 to 40 percent per year.<sup>4</sup> Organizations need to manage this in feasible ways. When there is so much data to manage, there’s no way to “do it all” – at least not all at once.

Many organizations opt to focus on master data over transactional data. *Master data* describes the people, places, or things involved in running the organization and is used by multiple business processes, while transactional data describes an event or transaction that has occurred.<sup>5</sup> For retail organizations, “Customer” is a common master data object. The record for Customer X will be used in every transaction with that customer. *Transactional data*, on the other hand, is typically used just once, but it often includes references to master data (i.e., Sale 1234 was completed with Customer X.). By prioritizing the management of master data, which is reutilized by multiple business processes, transactional data will ultimately benefit. High-quality master data can also be analyzed to help target business areas that need

improvement, such as retailers using customer data to target their marketing.

Within master data, particular data objects or entities (again, such as “Customer”) need to be selected for initial governing focus. As data improvements are made and sustained, this master data governance can expand across other data objects. Selecting and prioritizing these objects depends on the pain points identified in the business case for data management; focusing on the data objects that create expensive business problems yields the greatest return on investment.

## Forming data governance

Because data is created and used by all parts of an organization, data governance is often set up in a manner that federates different business units with a central data management function. This structure helps build common understanding of the impacts of data objects across different business areas, while also including a fully dedicated team of information specialists to execute data management activities and oversee the program. While data governance is highly collaborative, hierarchy is necessary. Senior leadership needs to be involved in order to escalate and execute data decisions that ultimately impact the business.

Often, data governance functions will be set up to mirror pillars of an org chart, like the example below. Keep in mind that the images on the left-most side represent part-time roles; these people are still situated in full-time jobs in business operations. However, they have dedicated some of their availability to data management activities and are of key importance when data-quality issues arise or when decisions must be made.

## Governing with data management standards

In addition to developing strategy and documenting decisions, the data governance body oversees the profiling of data for particular quality aspects. Having accepted the business case for data management, the organization has in principle agreed to invest in data quality. The active data governance body needs to take this to the next level of detail. It is one thing to state “We need better data” and another to define what exactly “good” means in order to apply tangible, quantifiable improvements to the data.

Different branches of data theory define different dimensions of data quality.<sup>6</sup> While the dimensions and their definitions may vary from author to author, there are some common quality requirements:

- **Completeness:** Does the data have all its required or conditionally required fields completed?
- **Uniqueness:** Are there duplicate data objects or database records, or near-duplicates (i.e., referring to the same physical asset or person, but under a different name)?
- **Integrity:** Are completed fields constructed in a way that maintains the structure of the data so that it can be effectively used (e.g., are North American phone numbers consistently transcribed with ten digits rather than seven)?
- **Accuracy:** Is the data correct, corresponding accurately to the thing to which it refers? This quality dimension is often the most difficult to manage, as it requires validation (sometimes manual) against a source of truth.


Just as data governance bodies make decisions about which data types and data objects to focus on, they also determine which data quality aspects will be prioritized in their data management program. To do this, policies or standards are produced to set corporate-level requirements regarding which data objects are included in scope and which quality aspects are applied to those objects. Again, because data governance matures and expands over time, more standards will be written for more objects, and more data quality aspects will be addressed within existing standards.

With standards in place, data is profiled and monitored against quality requirements, performance metrics are collected on both the data itself and the business processes that utilize it, and the execution of the standards (and the effectiveness of the data governance team) is demonstrated.

## In closing

There are many drivers behind data management: some organizations want to increase productivity and subsequent revenue, some want to increase reliability of their operations and ensure that their staff is working in the most efficient and effective manner, others want to reduce legal risk and protect their corporate reputation.



From the perspective of librarianship, data management and data governance comprise another information service to support organizations in their broader goals. Librarians offer invaluable skill sets to data governance bodies through their ability to understand the impacts of data to cross-functional business processes, their knowledge of the logical workings of database systems, their collaborative approach to building business relationships, and their training to address information needs from multiple points of view. Data is nothing new to librarians or to business, but trending interests and hype about data shed new light on opportunities for organizations to improve their performance and for librarians and information professionals to contribute to that success.<sup>7</sup> 

**Melanie Sucha**, MLIS, ([msucha@suncor.com](mailto:msucha@suncor.com)) is a Process Analyst – Data Governance at Suncor Energy in Calgary. She teaches Information Management via distance education in the Faculty of Information and Media Studies at Western University.


## Notes

1. Thomas H. Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21<sup>st</sup> Century," *Harvard Business Review*, October 2012, 70-76.

2. Michael Brackett, *The DAMA Guide to The Data Management Body of Knowledge* (Bradley Beach, NJ: Technics Publications, 2009), 17.
3. Danette McGilvray, *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information* (Burlington MA: Elsevier, 2008), 52.
4. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers, *Big Data: The Next Frontier for Innovation, Competition and Productivity* (Washington DC: McKinsey Global Institute, 2011), 20, accessed January 7 2014, [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)
5. McGilvray, *Executing Data Quality Projects*, 42. 6. Ibid., 30.
7. Author's Note: I am indebted to my colleague Mona Pomraning, a data consultant of Greater Seattle Area, Washington, for her invaluable insight into the suitability of librarians in data management roles, and for the many, many conversations that informed my data management views. I would also like to thank Elizabeth Burton of Suncor Energy Inc. for her assistance in refining ideas presented in this article.

## Metadata Management continued from page 25

find it inspiring. It highlights the fact that technology is open and accessible, and that if you're curious, you can learn how to harness its potential. The only limitation is your ambition.

Well, maybe I'm fibbing a little bit. We always have limitations on our time and money. However, consider this professional development. Invest a few hours or days in learning how to use these tools. It will invariably save you time and money when you need to modify your library's metadata. You can write a short yet effective metadata management Perl program using the MARC::Record module in forty lines or less. Your catalogue will be an accurate representation of your collection in no time. 

## Resources

- <http://www.loc.gov/marc/>  
<http://marcedit.reeset.net/about-marcedit>  
[http://wiki.code4lib.org/index.php/Working\\_with\\_MARC](http://wiki.code4lib.org/index.php/Working_with_MARC)  
<http://summit.sfu.ca/system/files/iritems1/427/self-education-systems-librarians.pdf>

**David Cook** ([kingandcaptain@gmail.com](mailto:kingandcaptain@gmail.com)) is a Canadian systems librarian currently living and working in Sydney, Australia, where he helps develop the open source ILS Koha. He has almost completely taught himself everything he knows about Linux (specifically the Debian distro), programming, databases, and web servers in the past two years—with the wonderful assistance of some very patient and knowledgeable Koha colleagues. He fervently believes in life-long learning and always has something new he is trying out at home or at the office.



## Know Thy Vendor: Getting the Best in Off-Site Records Storage

*Librarians and other information professionals should be careful about where they store their records. Some off-site vendors are better—and more honest—than others. Much depends on reviews of a vendor’s site, reputation, and indexing system.*

I did six years in a federal penitentiary that reminds me a lot of the warehouse I run. It’s an old concrete pile and it smells damp. You know the place. You store your records here: dozens of boxes crammed with files and miscellaneous materials. Much of this stuff is junk. You’ll never need to look at these old correspondence files again, or those cartons of brochures entitled *Hot New Fiction for 1983* and *Welcome to the World of DOS*.

Why are you saving so much obsolete stationery? How about these boxes of letterhead with the logo that your library stopped using two decades ago? Where did these antique computer tapes come from? And what’s so important about these mouldering newspapers and telephone directories? You may not remember these items, because so much time has passed since you packed them up and sent them my way. The only reminder I give you is a monthly invoice, which you pay without question. I love this business.

But not everything that you store off-site is worthless. For example, consider these original building plans for your central library and all of your branches: not only the structures, but also the plumbing, electrical, and heating systems. This material is essential for the purposes of facility maintenance. It even includes warranties for the roofs and other building components. You wouldn’t want to lose these documents, would you? If you did, it would cause a lot of inconvenience for any contractor employed to work on your buildings. The cost of the work could increase a great deal. And without the warranties for the roofs, you would have to pay far more for repairs.

### City Hall

“What people tell me is that they’re not worried about their building plans because somebody else has copies of them,” says Brian, a BC construction contractor with thirty years of experience. “Departments connected

to a municipality assume that City Hall keeps vital records and is prepared to share them at short notice. But often those records are lost, or misplaced, or simply unavailable. I’m obliged to work without them, and my company has to spend more time – and more of the customer’s money – figuring out what’s in the plans so that we can proceed with our project. I’d advise libraries not to take the storage of their records for granted.”

Your trustees will suggest that your valuable files need better storage. You ran out of space in your branch basements years ago, so you need an off-site storage vendor who will treat your information with respect. But first, what about your records management?

If you want to control costs by cutting down on useless paper, you’ll need an inventory of the records in all of your locations. Your inventory should contain the most basic data regarding the kinds of documents you keep, their inclusive dates, quantities, and current storage location. If you compile an inventory from scratch, you’ll probably discover tons of material that you don’t really need. Your inventory can turn out to be an effective cost-saving tool, since it will indicate opportunities to free up space and reduce storage expenses.

Next, you need a records retention schedule, which tells you how long you must retain various documents. Does your accounting department already have a retention schedule? That’s good, but is it up to date? Does it indicate retention periods for all of your library’s documents? You must regularly audit and update your retention schedule so that documents from all departments are controlled. Without a well-maintained retention schedule, you’ll pay a lot for the storage of paper that should have been sent to a recycler.

## Procrastination

“People often neglect to keep their retention schedules up to date,” says Lauren, a records manager and corporate librarian in Toronto. “There’s simply too much to do, and some tasks are more urgent than others, and the retention schedule can be ignored for another week, or month, or year. And eventually it’s so out of date that you must spend a lot of time updating it. But without it, managing your records—particularly in off-site locations—can be more difficult and a lot more expensive.”

When you decide to store records off-site, make sure that your off-site vendor is reputable. It’s amazing how often organizations in all sectors send sensitive information to vendors with questionable backgrounds. You should know better than the Attorney General’s office that stored sensitive records concerning criminal investigations in a warehouse owned by mobsters. Or the trust company that kept its mortgage files in the basement of an office tower managed by an offshore company that specialized in shady property development. Or the insurance company that sent confidential medical files to a paper destruction company that sold the files to a movie production studio.

You should know better, but do you? Have you actually checked out the reputation of your off-site vendor, asked for references, interviewed the warehouse manager, and made inquiries about the vendor’s reliability? If not, you might be keeping people like me in business. If your confidential files are used for unscrupulous purposes, you might never find out. Worse, the illicit use of your records might be exposed, and you’ll have a crisis to deal with. Your trustees will not like the idea of your sensitive information falling into the wrong hands. And how would you and your co-workers like the idea of that happening to your HR files? You might stop doing business with me, but I won’t suffer much. I may deal with a lot of paper, but I never leave a paper trail. Moreover, there are always new customers for a vendor who has storage space for old files.

## Pigeons unacceptable

You should also inspect any off-site space in which you intend to store records. Be sure that the warehouse is well maintained. Look for signs of water ingress and moisture damage. Check the security procedures and fire control systems. And examine the shelving: is it stable, moored to a supporting surface, and rust-free? Do you see any sign of rodents or pigeons? Don’t put up with sloppy warehouse maintenance or substandard



Photo by Deborah Johnson

*A paperless office? When?*

security, and don’t accept a vendor’s promises that a warehouse’s weaknesses will be rectified soon. That vendor might assume that you intend to visit the warehouse only once. Usually he’s right. And many customers don’t even make an initial visit.

An off-site warehouse may be in top condition, with state-of-the-art security, but what about the vehicles that transport your documents to and from your sites? Some customers prefer unmarked vans with alarms and advanced locking systems. Others will tolerate any old truck that pulls up to the loading bay. Do crooks in your area break into vehicles? Do they steal anything on wheels? If so, ask your vendor to use an inconspicuous and fully alarmed vehicle when transporting your files.

“There are numerous cases in which boxes of documents go missing in transit,” says Drew, a records manager in Alberta. “I’ve heard about boxes flying out the back doors of vans on the way to storage warehouses. The boxes hit the highway and paper flies everywhere and spreads for miles. That is a literal paper trail as well as a records manager’s nightmare.”

To ensure that your files aren’t misplaced in an off-site location, check the vendor’s indexing system. A responsible vendor won’t accept your materials for storage unless they are clearly labeled with your library’s name and other identifiers. Large vendors use bar coding to track each document box. Ideally, the vendor will be able to tell you exactly where your records are at any time. Some customers use their vendor’s indexing





Photo by Deborah Johnson

*We will digitize everything ... tomorrow, and tomorrow, and tomorrow*

system to compile a records inventory, and work with the vendor to develop retention schedules.

### Additional services

Once you find a vendor you can trust, consider the ways that you can work together to manage your company's records. Vendors can often provide a lot more than storage space for documents. Some will destroy and recycle old paper, or arrange for its conversion to digital formats. Some will assist you with your business continuity planning, and set up hot sites and alternative space in case your administrative offices are damaged or inaccessible. Many offer records management auditing services to discover any inefficiencies or security weaknesses.

Large vendors maintain data storage systems for clients who need to back up their electronic information off-site. Beginning in the 1970s, customers would store their tapes and disks in the vendor's vaults, and some still do. But the growing trend is to transmit data directly to an off-site system in a secure location. That location may be in the same city as the client, or across the

country, or on another continent. These days collocation technology and the Internet allow clients to have multiple back-up sites at a reasonable price.

Off-site storage of electronic data demands comprehensive security measures. Your firewalls and encryption must be current, high-quality, and frequently audited if you are to avoid losing data to hackers. Your off-site data security systems should be strictly confidential, and known to only your senior IT managers. Materials related to security – for example, manuals – should be carefully locked away. You can be sure that hackers would enjoy browsing through them.

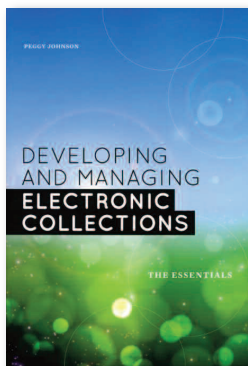
### Contract or con man

Whatever you demand from an off-site vendor, include your terms in a contract. If you're a new client, feel free to sign a contract for a term of no more than a year, to see how well your vendor performs. Don't agree to a longer period until you're satisfied with the vendor's service, facility maintenance standards, and prices. Insist on a cancellation clause. Above all, don't compromise your library's standards because a vendor won't agree to improving security measures at a warehouse or testing an electronic back-up system. Remember that you and your vendor should work together so that you can derive the greatest benefit from your off-site storage agreement.

Or you can take your chances with a vendor like me. You've never visited my warehouse or asked me about its security systems. You don't care about the water stains on the ceiling, or the rats, or the occasional loss of a document box. It doesn't matter if a few records go missing, especially if they're garbage. But sometimes they're worth their weight in gold. Look at these borrower records you sent me accidentally. I've got a friend who puts together mailing lists—but you don't want to hear about that. 🐼

*Guy Robertson (guy\_robertson@telus.net) is an Instructor in the Library Technology Program at Langara College, an Instructor in the School of Public Safety & Security at the Justice Institute of BC, and an Adjunct Professor at the School of Library, Archival, and Information Studies at the University of British Columbia. In 2009 CLA published his Unofficial Wisdom: Selected Contributions to Feliciter 1995 – 2009. In 2014, Woodhead Publishing (UK) will publish his Disaster Planning for Libraries: Process and Guidelines. Note: The names of persons quoted in this article have been changed to ensure privacy. And no, Mr. Robertson has not spent time in a federal penitentiary. He remains at large.*





### Developing and Managing Electronic Collections: The Essentials

Peggy Johnson. Chicago, IL: ALA Editions, 2013. ix, 186 pages. softcover, \$65.00 US. ISBN-13: 978-0-8389-1190-7 (paper), 978-0-8389-9651-7 (PDF), 978-0-8389-9616-4 (ePub), 978-0-8389-9617-1 (Kindle)

**This is an excellent** brief introduction to a very broad topic. The book is well written and logically organized, with a glossary of terms at the end, and each chapter is not only documented with notes but also includes a substantial list of “Suggested Readings” relevant to the subtopic. There are eight chapters overall, starting with an “Overview of the Evolving Electronic Collections Environment” and ending with comments on the “Future of E-content in Libraries,” with chapters in between on selection, evaluation, order placement, licensing, dealing with vendors, “Working across Organizational Units to Acquire and Manage E-resources,” and budgets. The first chapter is masterful, covering the key events in some sixty years of the development of e-resources without being cursory, and in fact highlighting some interesting details.

In the heart of the book, Johnson covers a large range of the most important topics, including current issues of concern to managers of e-resources, such as reconsidering participation in Big Deals, “the complexity of troubleshooting,” and the frustrating swath of problems associated with buying and managing e-books. She says correctly that “the business model for e-books is in flux and volatile,” but offers hope for some aspects of the problem in her final chapter: “Publishers and distributors will make their books readable on more than one device, paralleling the move that Apple made when it began selling DRM-free tracks that could be played on various devices (not just iPods) and when Amazon.com began selling DRM-free music.” E-resources are the primary workload of those who work in collections in academic libraries, and Johnson predicts that this will not change: “Delivering licensed content will continue to be libraries’ top priority.” She also predicts—provocatively, I think, but it is easy to make a strong case for it—that “patron-driven acquisitions will become the primary approach to collection building in academic libraries.”

I would highly recommend this book for anyone wanting a short, informed overview of e-resources. There is the odd thing in it that is not up to the overall quality of the book (e.g., the case study about free digital content (p. 46), and the somewhat obvious and lame advice on effective meetings with sales reps (p. 96)), but these are tiny faults in an exceptional book.

Reviewed by Wayne Jones ([wayne.jones@carleton.ca](mailto:wayne.jones@carleton.ca)), Associate University Librarian (Collections and Library Development) at Carleton University in Ottawa, ON.



### 50+ Library Services: Innovation in Action

Diantha Dow Schull. Chicago, IL: ALA Editions, 2013. 335 pages, softcover, \$55.00 US. ISBN-13: 978-0-8389-1119-8

**There has been** much discussion in library literature in recent years about how the “boomer revolution” will necessitate change in libraries’ traditional approach to services for seniors.

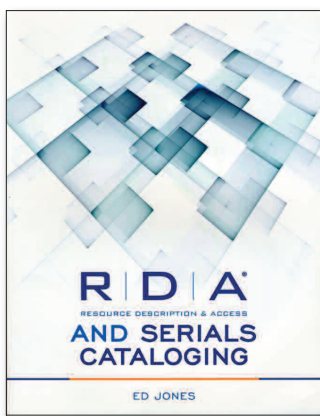
This comprehensive overview details how many libraries have risen to this challenge, developing innovative programs and service models. Schull begins with an overview of evolving service models for adults aged fifty and older (“50+ library services”), and profiles library systems recognized for leading-edge innovations in this area, describing how they have developed their services, with information on funding and partnerships and useful “keys to success.”

The core of the book is a compendium of program descriptions from American libraries, arranged thematically around nine subject areas, such as “Work, Careers and Service” and “Reflections and Transitions.” Each theme section begins with an outline of the trends in this area relating to adults 50+ and the implications of these trends for library service. Several detailed program descriptions are presented for each theme, as well as an “in brief” listing of many other library programs with links for additional information. The book concludes with a discussion of how library planning

and design can make the physical space more appealing for 50+ customers.

Although the focus of the book is on new and innovative services, some of the programs described will be familiar, such as “lunch and learn” series and computer classes. The point here is that these “traditional” programs can be tweaked and marketed effectively to the active 50+ audience, and also that many programs designed for a wider age range will attract this group as well. The funding sources and other resources noted in the book are American, but this does not detract from its value to Canadian librarians looking for inspiration in developing new services for boomers and beyond. There is truly a wealth of ideas here! Highly recommended for all public libraries.

*Reviewed by Heather MacKenzie ([mackenh@Halifax.CA](mailto:mackenh@Halifax.CA)), Manager, Diversity and Accessibility, Halifax Public Libraries, Halifax, NS.*



## **RDA and Serials Cataloging**

*Ed Jones. Chicago, IL: ALA Editions, 2013. 215 pages, softcover, \$60.00 US. ISBN-13: 978-0-8389-1139-6*

*RDA and Serials Cataloging* is a practical and timely publication that will undoubtedly become essential reading for cataloguers as they move beyond the basics of FRBR and RDA cataloguing and into the realm of applying RDA instructions in the challenging real-life context of serials cataloguing.

Jones has identified the audience for this publication as “serials cataloguers who are new to RDA” and “monograph cataloguers who are new to both serials cataloguing and RDA” (p. vii). The author assumes that the reader has a working knowledge of both AACR2 and MARC 21. Given that assumption, it is not surprising to find that the publication dives into the finer details of

RDA-compliant coding of MARC 21 serials records with few references to non-MARC applications of RDA. A substantial amount of detail and a well-rounded treatment of MARC 21 serials cataloguing has been packed into 215 pages. The first chapter provides an introduction to serials cataloguing and thoughtfully describes challenges posed by the dynamic nature of serial publications while outlining the history of serials cataloguing. While many readers may be tempted to bypass this chapter, reading it is highly recommended. As someone who has been involved with serials cataloguing since 1979, Jones brings a rich perspective to the topic.

While chapters two and three provide essential background information, most cataloguers will likely find chapter four to be indispensable. In this chapter, Jones discusses various elements of serials records and relevant RDA instructions and provides multiple examples. When it’s relevant, readers are directed to an appropriate source for further information such as the RDA instructions, the MARC 21 standard, or LC-PCC guidelines. Jones’s involvement with CONSER since 1981 and his recent work with FRBR, FRAD, and RDA is evident in the depth and balance of practical and theoretical detail given.

Jones’s readable writing style and lighthearted approach to discussing common cataloguing dilemmas makes what could have been a dry topic into a compelling and surprisingly interesting read.

*Reviewed by Donna Frederick ([dem124@campus.usask.ca](mailto:dem124@campus.usask.ca)), Metadata Librarian, University of Saskatchewan, Saskatoon, SK.*



## CLA EXECUTIVE COUNCIL 2014

### **President**

*Marie DeYoung*  
(902) 420-5532  
marie.deyoung@smu.ca

### **Vice-President/ President-Elect**

*Sandra Singh*  
(604) 331-4007  
sandra.singh@vpl.ca

### **Treasurer**

*Mike Ridley*  
(519) 824-4120 ext.56747  
mridley@uoguelph.ca

### **Councillor-at-Large**

*Lorna E. Rouke*  
(519) 884-8111 ext. 28271  
lerourke@uwaterloo.ca

### **Councillor-at-Large**

*Marc Truitt*  
(506) 364-2567  
mtruitt@mta.ca

### **Councillor-at-Large**

*Jane Schmidt*  
(416) 979-5000  
jschmidt@ryerson.ca

### **Ex Officio, Executive Director**

*Valoree McKay, CAE*  
(613) 232-9625 ext. 306  
vmckay@cla.ca

## CLA STAFF CONTACTS

*Beverly Bard*  
**Desktop Publisher &  
Book Review Coordinator**  
bbard@cla.ca ext. 324

*Valoree McKay, CAE*  
**Executive Director**  
vmckay@cla.ca ext. 306

*Chris Culhane*  
**Program Coordinator, Young Canada Works (YCW)**  
ycw@cla.ca ext. 321

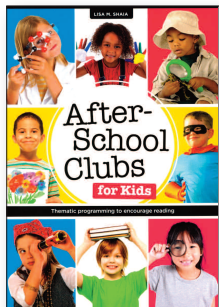
*Wendy Walton*  
**Conference & Events Manager**  
wwalton@cla.ca ext. 302

*Anita Fortier*  
**Orders Administrator**  
orders@cla.ca ext. 310

*Penny Warne*  
**Web & IT Infrastructure Manager**  
pwarne@cla.ca ext. 320

*Geraldine Hyland*  
**Member Services Manager**  
ghyland@cla.ca ext. 301

**Telephone: 613.232.9625 • Fax: 613.563.9895 • [www.cla.ca](http://www.cla.ca)**



## **After-School Clubs for Kids: Thematic Programming to Encourage Reading**

Lisa M. Shaia

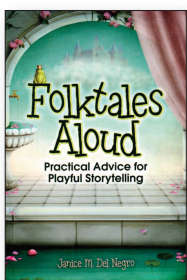
Learning that takes place "after hours" in a club setting is often an undersupported component of children's education. After-school clubs built around books encourage independent, recreational reading, which in turn has a positive impact on the rest of a child's day. In this book, Shaia offers a year's worth of ready-to-implement program ideas for librarians and educators. Her month-by-month calendar of themed clubs is conveniently divided by grade level (K-2, grades 3-4, grades 5-6) to allow club leaders and organizers to closely align activities and book selections to the ages of the club's participants. This handy book:

- Offers programming on such themes as science, math, animals, mysteries, art, fairy tales, and more, all of which can be adapted for either a half-hour or forty-five-minute time slot
- Details age-specific bibliographies and suggestions for read-alouds, music, and craft or game activities, with different ideas for each week in a month

- Presents information on publicizing the activity, community outreach, display ideas, set-up, supplies, management, and evaluation

With its inviting approach to book-based programming, Shaia's start-to-finish guide helps promote reading as a fun, engaging activity for kids.

CLA Member Price: \$48.60 • 208 pages • 7" x 10" • Softcover • 2013 • ISBN-13: 978-0-8389-1202-7



## **Folktales Aloud: Practical Advice for Playful Storytelling--eEditions e-book**

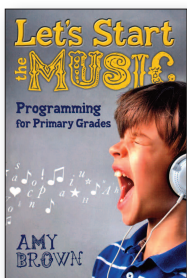
Janice M. Del Negro

A good folktale triggers the imagination, connecting children to a wider world as well as increasing their vocabulary and comprehension skills. In this delightful and easy-to-use book, teacher and storyteller Del Negro gives librarians, teachers, and parents the keys to storytelling success. Including more than a dozen original adaptations of folktales from around the world, tailored specifically for library and classroom use, she:

- Reviews storytelling basics such as selecting a tale and learning the story
- Offers tips for dealing with stage fright and reluctant listeners
- Presents a bibliography of recommended online and print resources, steering readers to more wonderful tales to tell

For young listeners the folktale is a perfect gateway to the exciting worlds of culture and literature, and Del Negro's book invites their engagement with proven techniques and original story scripts that can be used by experienced as well as beginning tellers.

CLA Member Price: \$55.00 • 224 pages • 6" x 9" • Softcover • 2014 • ISBN-13: 978-0-8389-1135-8



## **Let's Start the Music: Programming for Primary Grades**

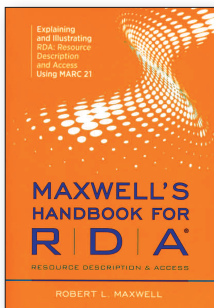
Amy Brown

Music programs have been scaled back or eliminated altogether from the curricula of many schools. Luckily, storytimes offer ideal opportunities for music and songs. In this collection of easy-to-use, easy-to-adapt library programs for children in grades K-3, Brown connects songs and musical activities directly to books kids love to read. Offering several thematic programs, complete with stories, songs, and flannelboard and other activities, her book includes

- Music activities, lists of music-related books, mix-and-match activities, and additional web resources
- Terrific tips on how to teach songs to young children
- Ways to develop original songs and rhythms to enliven children's books

Even if you can't carry a tune in a bushel basket, this handy resource has everything you need to start the music in your storytimes.

CLA Member Price: \$52.65 • 184 pages • 6" x 9" • Softcover • 2014 • ISBN-13: 978-0-8389-1166-2



## **Maxwell's Handbook for RDA: Explaining and Illustrating RDA: Resource Description and Access Using MARC21**

Robert L. Maxwell

In this clear and comprehensive resource, cataloging expert Robert Maxwell brings his trademark practical commentary to bear on the new, unified cataloging standard. Designed to interpret and explain RDA: Resource Description and Access, this handbook illustrates and applies the new cataloging rules in the MARC21 environment for every type of information format. From books to electronic materials to music and beyond, Maxwell:

- Explains the conceptual grounding of RDA, including FRBR and FRAD
- Addresses the nuances of how cataloging will, and won't, change in the MARC21 environment
- Shows catalogers how to create and work with authority records of persons, families, corporate bodies, geographic entities, works, and expressions
- Explores recording relationships, working with records of manifestations and items, and more

- Provides numerous sample records to illustrate RDA principles

A guided tour of the new standard from a respected authority, this essential handbook will help catalogers, LIS students, and cataloging instructors navigate RDA smoothly and find the information they need efficiently.

CLA Member Price: \$89.10 • 912 pages • 7" x 10" • Softcover • 2014 • ISBN-13: 978-0-8389-1172-3

*The Canadian Library Association is the exclusive Canadian source of all American Library Association publications*

**Order from:** Canadian Library Association, 1150 Morrison Drive, Suite 400, Ottawa, ON K2H 8S9

Tel: 613-232-9625, ext. 310 • Fax: 613-563-9895 • [orders@cla.ca](mailto:orders@cla.ca) • or shop CLA at [www.cla.ca](http://www.cla.ca)

