A knowledge management system framework for an open biomedical repository: communities, collaboration and corroboration

Lisa Kruesi, Frada Burstein and Kerry Tanner
(Human-Centred Computing, Faculty of Information Technology, Monash University)
Published in the *Journal of Knowledge Management*, 2020

Purpose

To assess the opportunity for a distributed, networked open biomedical repository (OBR) using a knowledge management system (KMS) conceptual framework. An innovative KMS conceptual framework is proposed to guide the transition from a traditional, siloed approach to a sustainable OBR.

Methodology

The paper reports on a cycle of action research, involving literature review, interviews and focus group with leaders in biomedical research, open science and librarianship, and an audit of elements needed for an Australasian OBR; these, along with an Australian KM standard, informed the resultant KMS framework.

Findings

The proposed KMS framework aligns the requirements for an OBR with the *people, process, technology* and *content* elements of the KM standard. It identifies and defines nine processes underpinning biomedical knowledge—discovery, creation, representation, classification, storage, retrieval, dissemination, transfer and translation. The results comprise an explanation of these processes and examples of the *people, process, technology* and *content* dimensions. While the repository is an integral cog within the collaborative, distributed open science network, its effectiveness depends on understanding the relationships and linkages between elements and achieving an appropriate balance between them.

Implications

Adoption of the KMS framework for a distributed, networked OBR will facilitate open science through reducing duplication of effort, removing barriers to the flow of knowledge and ensuring effective management of biomedical knowledge.

Value

The framework demonstrates the dependencies and interplay of elements and processes to frame an OBR KMS.

Keywords biomedical knowledge management, knowledge dissemination, knowledge reuse, open science, libraries, open biomedical repositories

Paper type Research paper

Introduction

Global pandemics, as experienced in 2020 as a result of COVID-19 and throughout history, severely disrupt and can have a devastating impact upon human existence (LePan, 2020). Management of biomedical knowledge is key to achieving satisfactory levels of healthcare. The US National Library of Medicine (NLM) is a world leader in biomedical knowledge management (KM) (US National Library of Medicine, 2019a). The NLM works with the National Institutes for Health to make biomedical knowledge openly accessible from PubMed, an aggregator database, and the repository PubMed Central International (PMCI), which comprises the US PMC and Europe PMC (US National Library of Medicine, 2018). This research was motivated following interest by stakeholders to determine the viability of an OBR for this region, such as an Australasia PMC *<reference withheld for blind review>*.

The ultimate aim of the OBR is to facilitate the route from basic research results to effective healthcare solutions; in this sense, OBRs are KM vehicles–ensuring what is necessary to derive the greatest benefit from knowledge resources (Becerra-Fernandez and Sabherwal, 2015). The structure and content of repositories has evolved from knowledge databases that were originally systems accessed predominantly by specialists to web-based systems driven by end users. A digital repository is a set of systems and services that ingest, store, manage, display, retrieve and allow reuse of digital objects. (Pinfield *et al.*, 2014). Throughout the world, institutional, aggregating and disciplinary repositories co-exist (University of Nottingham (UK), 2005-).

Open science is about making activities in the discovery process fully and openly available, creating transparency, and advancing future discovery by allowing others to build on existing work (Watson, 2015). With the development of open science, repositories have been established to make knowledge accessible, discoverable, mineable, interoperable and permanently available. With the significant investment by Australian taxpayers in biomedical research and mandates to open up publicly funded research, an Australasia OBR could become a formal member of an international repository system. Whilst university institutional repositories exist, the Australasian region lacks a consolidated OBR.

KM involves getting the right knowledge to the right user and using this knowledge to improve organizational and/or individual performance (Jennex *et al.*, 2009), and doing what is needed to fully exploit available knowledge resources (Becerra-Fernandez and Sabherwal, 2015). In developing a knowledge management system (KMS), there is a need for a holistic approach beyond just an IT focus (Chhim *et al.*, 2017). A KMS is a combination of knowledge management practices, such as a set of methods to support learning and organizational processes of KM development, and KM tools, such as IT-based systems supporting the practices (Centobelli *et al.*, 2019). The proliferation of repositories causes confusion for researchers in terms of which repository to consult or use as a platform to disseminate their research output. There are currently no validated approaches for researchers or administrators to compare repositories objectively or systematically. In highlighting the essential elements and processes for an OBR, the KMS framework presented here provides a way to identify and address the gaps to a sustainable open science system.

Removing the barriers to accessing research content through the establishment of a regional biomedical KM repository could help to avoid the vast amount of duplicate effort that occurs between organizations, and in academic and research libraries. Importantly, a coordinated approach will help break down the silos, dissolve historical organizational boundaries, cement vital connections required to create biomedical knowledge, and address the challenges of building an OBR (Arlitsch and Grant, 2018, Joo *et al.*, 2018). A large collaborative, distributed network is proposed whereby the repository is a vital cog in the scientific scholarly communications cycle. This network needs to achieve quality and reproducibility throughout the complex biomedical knowledge system of open science communities.

This research applies a *people, process, technology* and *content* model derived from the Australian standard, *Knowledge Management–a guide (Standards Australia, 2005).* This standard established that the organization of knowledge is an ecosystem that consists of a complex set of interactions between these four elements and that a balance between the elements is essential—one element

should not be developed at the expense of another (Halbwirth and Sbarcea, 2005, Standards Australia, 2005).

Motivation for this research into developing a KMS-based framework for planning an OBR came from a desire to remove knowledge from silos, and to improve the poor coordination of *people, process, technology* and *content* in the biomedical information profession. These factors result in a costly disconnect through to the pipeline that delivers the point-of-care evidence for patient care (Australian Living Evidence Consortium, 2018). Insights gained from the researcher's extensive experience working in leading Australian research universities reinforced the view that an institutional approach to repository management leads to unnecessary duplication of effort and major inefficiencies.

To address these issues, the literature review outlines the key research on the evolution of knowledge databases for biomedical research, the transition from information silos to open scholarship and OBR. This sets the scene for the remainder of the paper, which has a focus on KM and its relationship with open scholarship. In particular, KMS models that align with biomedical research activities are introduced. The results and discussion sections expand on the *people, process, technology* and *content* elements of a repository, which are detailed for each of the nine KM processes identified. The discussion draws implications of this research for the future management of biomedical knowledge, the opportunities for professional practice and the potential future direction of this research.

Literature Review

Evolution of knowledge databases for biomedical research

With the flurry of internet technologies and database advancements over recent decades, there has been a proliferation in the number of search platforms, repositories and databases for accessing knowledge. PubMed linked to PubMed Central (PMC) makes research evidence in the life sciences accessible throughout the world. PubMed is an aggregator database, the precursor of which was the printed Index Medicus that began in 1879 (Greenberg and Gallagher, 2009).

The international reliance upon PubMed knowledge is evident from the widespread reuse of PubMed citations. When formal arrangements were necessary, more than 500 licences to MEDLINE, PubMed's subset, were issued, with 200 of them to providers outside the United States. There are 28 freely available biomedical online tools reliant on PubMed content (Lu, 2011). These figures are likely to be conservative, as after 2016 the US National Library of Medicine discontinued licensing the system and opened PubMed up freely to all (US National Library of Medicine, 2016).

Thousands of proprietary databases that index and abstract journal articles have been made available since the 1970s (Regazzi, 2015). A few of the key biomedical databases that index the top world journals and grey literature include MEDLINE, Embase, CINAHL, Ovid Emcare, Biosis and Cochrane Library. Some of the databases abstract and index the same publications, though each will have a distinguishing feature: for example, the Embase database has a focus on drug and pharmaceutical research. All of the different biomedical databases provide a unique perspective and search features. For example, when undertaking a systematic review there are search standards

such as those published in the Cochrane Handbook, that indicate it is mandatory for researchers to search The Cochrane Central Register of Controlled Trials (CENTRAL) and MEDLINE, together with Embase if available, when undertaking a Cochrane Review (Lefebvre *et al.*, 2019).

The quality of biomedical literature that is deposited in biomedical repositories, is concerned with excluding research publications that do not achieve and maintain set publishing standards. Bodies such as the Committee on Publication Ethics (COPE) and the International Committee of Medical Journal Editors (ICMJE) define best practice in the ethics of scholarly publishing and assist editors and publishers to achieve this (Committee on Publication Ethics, 2020, International Committee of Medical Journal Editors, 2020). Achieving a consistent quality approach involves setting out to avoid promulgation of misinformation that can occur because of inadequate peer review or research fraud. Research practices need to be tailored to the needs of the discipline, along with services and tools created for reusability as part of daily researcher work (Chen *et al.*, 2019). It is argued that predatory journals must be denied the "legitimacy afforded by inclusion in prestigious databases like PubMed" (Harvey and Weinstein, 2017).

From information silos to open scholarship

Bibliographic databases have many limitations, and some of these have impacted on the design of present knowledge repositories. Proprietary bibliographic databases are usually organized by publisher preferences; they vary in design, with some requiring individual login. Most of these databases are standalone systems and their usage is restricted by subscription and licensing conditions; navigating database silos is challenging and such resources can require a high level of expertise in order to find relevant content (McLean and Lynch, 2003). Open scholarship seeks to address some of these limitations by making research output discoverable.

Based on existing trends, it is estimated that by 2025, 44% of all journal articles will be available as open access and 70% of article views will be to open access articles (Piwowar *et al.*, 2019). "The declining relevance of closed access articles is likely to change the landscape of scholarly communication in the years to come" (Piwowar *et al.*, 2019). The debate on who owns research output is being hotly disputed (Piwowar, 2019). Various social networking services for researchers, such as ResearchGate and Academia.edu now challenge traditional approaches to disseminating research.

Open Biomedical Repositories

A disciplinary repository like PMC meets the *process* requirement of the open access policies of major bodies such as the Wellcome Trust, US National Institutes of Health (NIH), the National Health and Medical Research Council (NHMRC), the Australian Research Council and the European Union as a means to disseminate research findings (US Department of Health and Human Services National Institutes of Health, 2008, Wellcome Trust, 2020, Australian Research Council, 2013, National Health and Medical Research Council (Australia), 2018, Science Europe Working Group on Open Access, 2015). Most researchers retrieve PMC articles from searches undertaken using internet search engines like Google. PMC has more than a billion articles retrieved from the NLM website each year, and according to the PMC Project Manager, this demonstrates how important this repository is for research discoverability (NLM Program Manager, 2018).

An Australasia OBR could be a platform to find related research data. "Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process" (Wilkinson *et al.*, 2016). Europe PMC has made numerous biomedical knowledge databases discoverable and their services for researchers are constantly improved. The Europe PMC model splits costs over 31 different funders; this is a strength. Publishers make funder-attributed research available through the repository and this allows services to be built on top. For example, Europe PMC biostudies reports are created to extract the underlying data. "Much more is achieved than just a repository of articles" (Kiley, 2018). A deeper insight of systems, such as Europe PMC which incorporates information from disparate big data sources, can better leverage the potentiality to enhance existing value generation means or stimulate new value (Marques Júnior *et al.*, 2020).

KM processes

Whilst recent research has explored the relationship of institutional repositories with KM practices in relation to people, technology and content, no identified research has been found on disciplinary repositories (Sabharwal, 2019).

The theoretical foundations of KM are well documented (Baskerville and Dulipovici, 2006, Dalkir, 2017). This research focuses on the KM processes related to biomedical research and innovation. According to Tuomi, when we explicitly address processes that underpin the establishment of shared understanding, it is then we develop KMS (Tuomi, 1999). Justification for the focus on effective knowledge processes is highlighted by Wiig, who recommends the need for systematic KM (Wiig, 1997). Whilst synergies exist with the work of Bhatt, who examines knowledge in the knowledge development cycle, this KMS framework extends beyond organizational boundaries (Bhatt, 2000). The focus of Wong and Aspinwall is also from an organizational perspective though their findings that the 'people' aspect, being management leadership and support, are key to successful KM, in addition to technology being a tool and not an absolute answer, are factors in common with this research (Yew Wong and Aspinwall, 2005).

KM processes are important throughout key research activities (Saito *et al.*, 2007). The conduct of clinical trials is a major biomedical research activity and requires registration of data and report planning early in the research. KM is not directly concerned with data, though the exception to this is when knowledge discovery occurs through data mining techniques (Becerra-Fernandez and Sabherwal, 2015). KM is concerned with the discovery of tacit and explicit knowledge from data and information or from the synthesis of prior knowledge (Becerra-Fernandez and Sabherwal, 2015). The discovery activity of research involves iterations of searching and reading (Kramer and Bosman, 2017).

Each of the knowledge creation modes identified by Nonaka are vital to the transformation of research by health practitioners, industry, or consumers, to adopt the findings as knowledge. A social process, which often involves checking with other practitioners and gaining insight from a range of sources occurs to form part of a 'mindline,' the knowledge in context that is used in practice. This social activity underpins the constant and repeated process to transform research into knowledge (Gabbay and le May, 2010).

Information systems that support collaboration, coordination and communication processes can increase a researcher's contact with colleagues. These information systems underpin knowledge creation activities (Alavi and Leidner, 2001). Conception, planning and commencement of research (National Health and Medical Research Council (Australia), 2016) and preparation and analysis (Kramer and Bosman, 2017) are key activities in knowledge discovery and creation processes.

KM processes align with biomedical knowledge creation activities. For example, there is a strong correlation of the KM storage and retrieval processes with the biomedical research activities of data collection, processing, analysis, storage and management (National Health and Medical Research Council (Australia), 2016) and writing and publication (Kramer and Bosman, 2016). The four groupings of KM processes and their alignment with biomedical research activities are reported in the literature *<reference withheld for blind review>*.

To develop a framework for an Australasian OBR, KM processes were mapped against biomedical research activities, to yield a set of nine key knowledge processes. These include: discovery, creation, representation, storage, classification, retrieval, dissemination, transfer and translation.

This framework goes beyond previous frameworks, such as the Institutional and Development (IAD) framework, as it is standards-based and operationalizes a comprehensive system of knowledge (Hess and Ostrom, 2006).

Research design

This research reports on Cycle Two of an action research project. Cycle One of this research achieved confirmation of stakeholder interest in an investigation into the concept, viability and potential for an Australasian OBR *<reference withheld for blind review>*.

Cycle Two, undertaken from 2018-2019, addressed the proposition that KM provides an effective theoretical framework for developing, analyzing and evaluating alternative designs for repositories that support the advancement of open scholarship. The Australian KM standard was adopted as the underlying theoretical lens for this research (Standards Australia, 2005). Action research is suited for practitioner research because it involves those who are experiencing the organizational or social challenges being addressed (Elden and Chisholm, 1993). The dual aims of action research are for practical problem solving, and for testing and/or potentially developing a model or theory (Burstein and Gregor, 1999, Williamson, 2013). The KM system framework developed is an outcome of the second action research cycle; it both contributes to theory and has practical applications.

Cycle 3 is testing the viability of this approach to developing an OBR.

Figure 1 provides an overview of the Action Research Cycles. Besides literature review, the data collection research techniques included semi-structured interviews and a focus group session. A university human research ethics committee approved the study and the associated documentation.

Figure 1 Action Research Cycles

Cycle One 2017-2018

Cycle Two 2018-2019

Cycle Three 2020
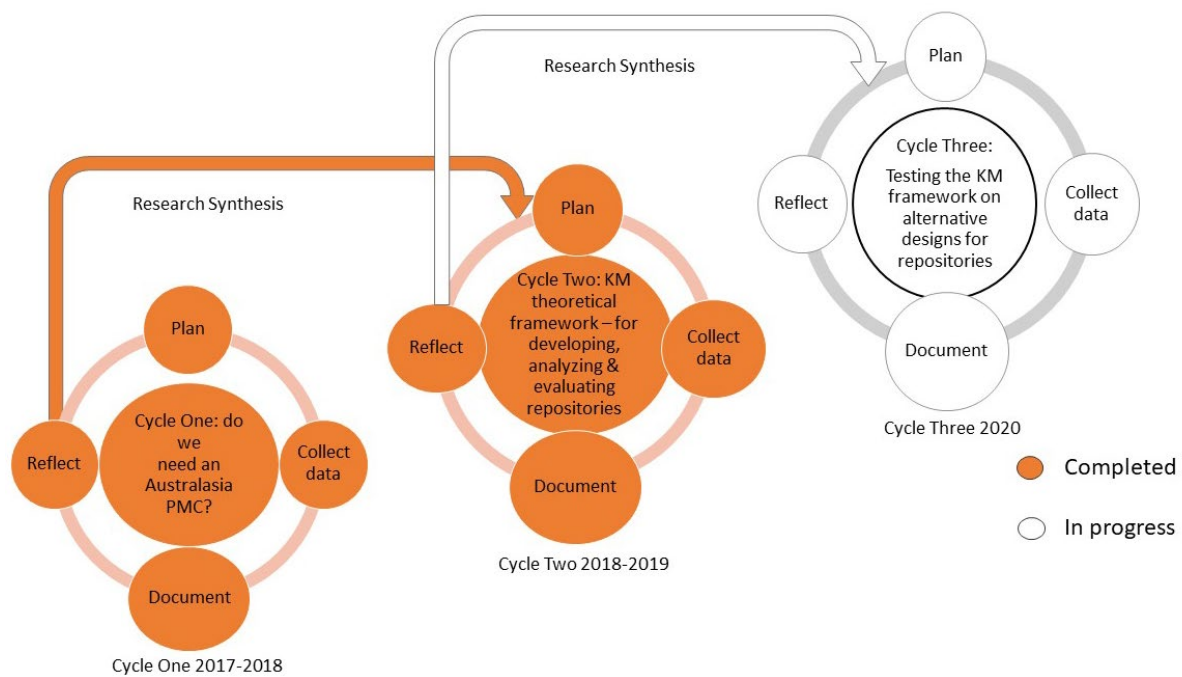
Interview participants were experts in OBR from the US National Library of Medicine (NLM), Europe PMC and PMC Canada. Two senior Australian academic library directors were also interviewed and a focus group was conducted with senior executive staff from the National Library of Australia (NLA). The interviews with Europe PMC and PMC Canada took approximately one hour each and were undertaken using the Zoom video-conferencing system. In person, one-hour interviews were held with a Program Manager from the US NLM and Australian academic library directors. Four library executives provided input in a focus group setting held at the NLA. Consent authorization for the interviews and focus group sessions was obtained. Transcripts of the sessions were captured and analyzed.

Interviews held with the US NLM, Europe PMC and PMC Canada and were semi-structured. The *people, process, technology* and *content* aspects of each PMC repository were the focus of the interviews; the approach from the Australian KM standard was adapted with the elements being defined as *people* addressing the *who*, the *process* the *how*, the *technology* the *tools* and the *content* the *what* (Standards Australia, 2005). Table 1 provides definitions of the elements.

An early version of the Australasian OBR framework was introduced at subsequent interviews with each of PMC entities. The emergent KMS framework for an Australasian open biomedical repository was also presented at the focus group and the interviews with two Australian academic library directors. Feedback from the research participants was recorded in interview transcripts and the key findings, along with the literature review, formed the basis of the final KMS OBR framework. This work is summarised in the results and discussion section.

Table 1 Definitions of the elements, adaption from the Australian Standard on KM

Element	Definition
People	The 'who' such as researchers, practitioners, professional staff, support staff, publishers, editors, and consumers. Includes the culture and environmental aspects.
Process	The 'how' and includes regulations, standards, rules, guidelines, plans, priorities, checklists, codes, instructions, taxonomies, protocols, policies, procedures and other explicit knowledge sources.
Technology	The 'tools' such as software, hardware, storage, digital systems, platforms, databases, websites and expert systems.
Content	The 'what' such as research data, metadata, database records, classification schemes, articles, videos, graphs, maps, visualizations, reports, and other digital objects.

Results and discussion

This section provides an analysis of the interviews, focus group results and the published research in the field. The essential *people, process, technology* and *content* elements of an OBR are detailed for each of the nine KM processes. The framework comprises elements that exist independently from a repository platform though are required to achieve a sustainable system. As the descriptions focus on the essential components for a sustainable OBR, not all of the elements, *people, process, technology* and *content,* feature within each of the nine processes. The framework is recommended to plan for an OBR and transition towards Open Science. Figure 2 illustrates the framework.

Figure 2 Towards Open Science: KM Processes for a Sustainable OBR
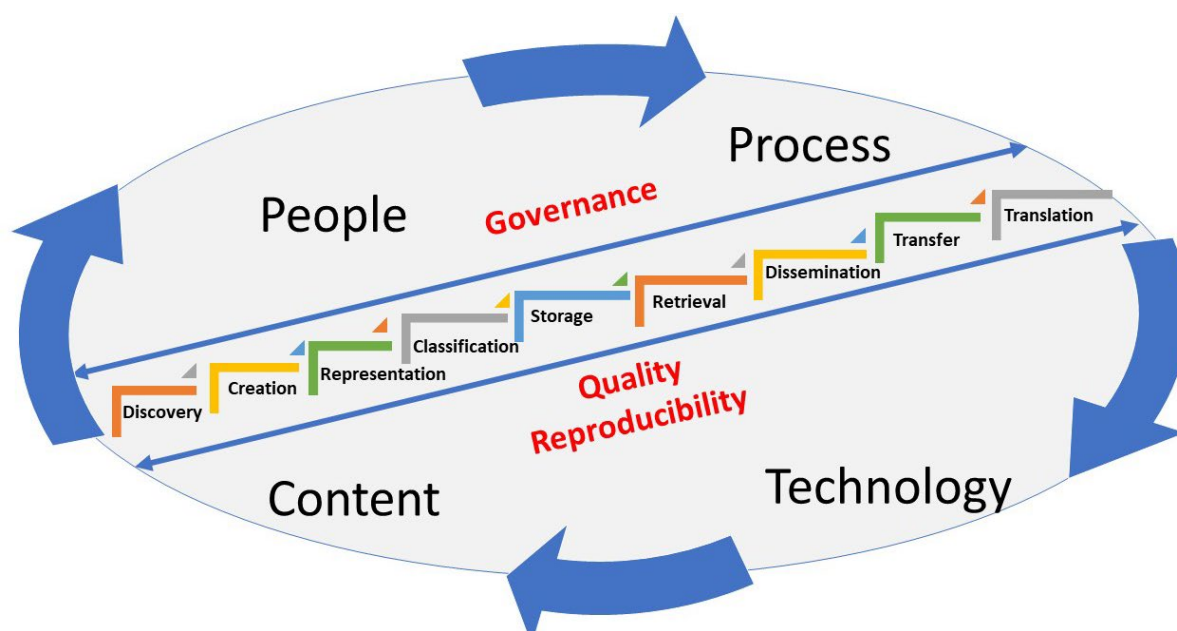
Discovery

The process of knowledge discovery occurs with the development of new tacit or explicit knowledge from data and information or from the synthesis of prior knowledge (Becerra-Fernandez and Sabherwal, 2015). Researchers, *the people* aspect of discovery in biomedicine, review existing knowledge by undertaking searches in biomedical databases and using search engines to identify gaps in knowledge. Knowledge (*the content*) is recorded in bibliographic databases, such as MEDLINE as well as various specialized molecular biology databases, such as the Entrez series of databases by the US NLM National Center for Biotechnology Information. These resources provide an opportunity and a need for developing advanced methods and tools for computer-supported knowledge discovery (*the technology*). For example, it is possible to search for genes that cause a particular disease or for drugs that treat that disease (Hristovski *et al.*, 2005). Analysis of scientific texts through text-mining systems are in common approaches to help with the discovery aspect of biomedical research (Jensen *et al.*, 2006).

Creation

New knowledge is created through a combination and exchange of diverse and overlapping knowledge inputs, generated when researchers interact (Polanyi, 1966, Schumpeter, 1934). Knowledge creation involves the generation of facts, relationships, and insights that are new to the existing body of knowledge (Arrow, 1962). New knowledge is typically intangible when it is created, but it can be converted into new products, patents, publications, and other tangible forms (Nonaka and Takeuchi, 1995). Over the past three decades, open scholarship mandates and data management principles have evolved. The open universal approaches on knowledge creation are being guided by the FAIR (Findability, Accessibility, Interoperability and Reusability) principles (Wilkinson *et al.*, 2016).

The *people* element of creation relates to researchers undertaking data analysis and other investigation activities. It also includes the role of funding bodies who steer and underpin creation efforts (Europe PMC, 2020). Others include those who create the databases and repositories, such as PMCI, and the publishers who copy-edit and provide platforms for research output.

Creation *processes* are influenced by priorities set by government and research organizations in response to public needs. Excellence in Research for Australia (ERA) is a periodic government *process* that determines research priorities and disciplines strengths for Australian universities. The ERA *process* has a significant impact on creation. Higher Education Research Data Collection is undertaken by Australian universities on reporting requirements to obtain research and development income data. This *process* also influences research output creation activities in the Australian higher education sector. Funding bodies lead creation, for example, Europe PMC has 31 funders that expect that research outputs they fund will be made freely and readily available (Europe PMC, 2020). These funders administer a *process* to drive behaviour and steer creation in the research sector.

Technology for creation involves access to a wide array of repositories to support research investigations. In particular, researchers (*people*) use (*technology*) such as open biomedical literature repositories like PMCI, along with (*content)* data repositories such as DataMed and Dryad Data Repository (Roberts *et al.*, 2017, Ohno-Machado *et al.*, 2017). Such repositories link to databases such as ENA, PDB, ArrayExpress, UniProt, RefSNP, OMIM, Pfam, RefSeq, Ensembl, InterPro, Bioproject, Biosample, EMDB, PXD, EGA, and TreeFam (Kim, 2015). *The content* to support creation is made available in research protocols, research data and research objects; all of these help with determining the novelty of the research and its contribution to existing knowledge.

Representation

Representation comprises explicit knowledge in the form of digital scholarly objects. Tuomi argues that structured knowledge becomes information when assigned a fixed representation and it is a standard interpretation as data (Tuomi, 1999). Knowledge representation may be in the form of publication, which is joint authoring, structuring, contextualizing and release of knowledge elements supported by workflows (Maier, 2007). The *people* element includes copywriters, editors and graphic designers working for publishers, who have a major role in determining publishing styles.

In biomedical research, *process* is achieved from international protocols and *people* networks to continually develop and improve research output standards and practices. For example, the Equator Network promotes transparent and accurate reporting of health research findings to improve the impact and reliability of biomedical research articles (UK Equator Centre, 2020). The International Committee of Medical Journal Editors (ICMJE) aims to improve the quality of medical science and its reporting. The ICMJE editors make recommendations for the conduct, reporting, editing and publication of scholarly articles in medical journals (International Committee of Medical Journal Editors, 2020). The ICMJE and the Equator Network are bodies that influence the biomedical knowledge representation *process* in the form of instructions, standards, checklists and codes; examples include the Consolidated Criteria for Reporting Qualitative Research (COREQ), and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (UK Equator Centre, 2020). The Committee on Publication Ethics (COPE) defines best practice in the ethics of scholarly

publishing and assists editors and publishers to achieve this (Committee on Publication Ethics, 2020). There are other *processes*, such as the US NLM expecting publishers to have at least a two-year history of quality scholarly publishing in the life sciences prior to their consideration of their journal for PMC; this serves to set rigorous standards for inclusion of only quality biomedical *content* (US National Library of Medicine, National Institutes of Health, 2019).

The *people* and *technology* elements of representation are achieved in an OBR, such as PMC, in the following ways:

- by publishers, some of who make all of their content available at the time of publication (for example, PLOS, BMC and eLife),
- by publishers who make individual articles available at the time of publication (in hybrid journals),
- by authors who self-archive the author manuscripts in PMCI, and
- by publishers who deposit the peer-reviewed manuscripts for free on behalf of authors (for example, Nature Publishing Group) (Europe PMC, 2019).

Classification

Taxonomies, also called classification or categorization schemes, are considered to be knowledge organization systems that serve to group objects together based on a particular characteristic. For example, keywords to describe research output are assigned by researchers. PubMed articles are assigned descriptors by librarians from the controlled and hierarchically-organized vocabulary published by the NLM, known as Medical Subject Headings (MeSH) (US National Library of Medicine, 2020a).

Classification *processes* involve rules for naming and describing research output. For example, rules for naming may be governed by international bodies such as the International Association for Plant Taxonomy (IAPT), which governs plants and the International Commission of Zoological Nomenclature (ICZN), which governs the naming of animal taxon. *(Note: Agricola, an agricultural database, is available from Europe PMC).*

Technology is used to provide automated classification systems, for example PubMed uses Solr, an open-source enterprise for document indexing. Unified Medical Language System (UMLS) integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records (US National Library of Medicine, 2019b).

Classification schemes and ontologies are used by repositories to allow *users (the people element)* to navigate the *content*, for example, Europe PMC researchers use GO, UniProt, EFO, ChEBI, NCBI Taxonomy and UMLS as ontologies in the biological *content* to "achieve a common understanding of the categories of objects described in life sciences data and the labels used for those categories" (Stevens *et al.*, 2010).

Storage

Storage comprises computer components and recording media used to retain digital data. Databases and repositories require storage for metadata and content. In the 21st century, the speed of mass data production and deposition necessitates creative solutions for data storage and computing infrastructure (Cook *et al.*, 2018). For repositories, storage is disk space available using a file system on top of storage hardware. Storage usually defines where the content resides.

Open biomedical literature repositories accommodate human data entry and publisher entry (the *people* element*)* of metadata and full text content. An OBR has an archival role. For example, the PMC manuscript submission system applies a standards-based approach for content preservation and has adopted the XML format to allow text mining. The preservation of associated data, such as clinical trials, is a priority (NLM Program Manager, 2019). Continuous improvement of the *technology* infrastructure takes place. A strict privacy policy is administered by PMC and no individual system user details are made publicly available. Standards-based systems are adopted to help ensure ongoing preservation of *content* and to enable text mining.

Since 2006, when funding bodies such as the NIH, NHMRC, ARC and the European Union established open access publishing mandates requiring their funded research be stored in repositories, the transformation of scholarly publishing through these *processes* was set in motion. During the period that PMC Canada was active, 2009–2018, the *process* allowed manuscripts to be submitted to either an institutional repository or the PMC repository and neither process was mandatory. It was stated that "when you give many options people often take the path of least resistance, which might be super-positive or might be just doing nothing" (NLM Program Manager, 2019). The failure of PMC Canada to provide clear *process* contrasts with the success of the clear, although at times cumbersome, *processes* of the US PMC. For example, in 2008 the US Congress mandated submission of certain funder research output to US PMC. The *process* was strengthened further in 2013 when every research output detailed in a government grant application had to be made openly available in PMC. This *process* helped to achieve 90% compliance with the open access policy. In an interview, the Manager of PMC Canada gave the definition that PMC Canada is exclusively about *technology*, and the repository's KM role is primarily to store what is being discovered and it is not involved in the creation of knowledge. The lack of acknowledgement of the interrelationship of elements within a KM system and their interplay with biomedical research activities and open science may be one of the factors contributing to the failure of PMC Canada (Landa, 2018).

The 'Group of Eight' top research universities in Australia, and other international research university networks have formed the Sorbonne Declaration on Research Data Rights. The declaration calls on governments to develop laws/policies to **"**avoid a 'lock-in effect" from commercial platforms and data services to ensure the openness and the reusability for research data" ("Sorbonne declaration on research data rights," 2020). Storage *processes* are based on international standards, such as preservation and interoperability standards (*Digital preservation handbook*, 2015, National Archives of Australia, 2018, International Organization for Standardisation, 2017).

Storage *process* requires adherence to open standards, for example, PMC submissions must be in XML format that conforms with an acceptable journal article DTD (Document Type Definition). PMC

stores content in eXtensible Markup Language (XML), which represents the structure and meaning of a document in a human-readable form (US National Library of Medicine, 2020b). All PMC content is converted to and stored in the NISO Z39.96-2015 JATS XML format. This is the commonly used archival format for journal articles (US National Library of Medicine, 2020b).

Research data is stored in a format to meet the FAIR (Findable, Accessible, Interoperable and Retrievable) principles that have been established to share and reuse data (Wilkinson *et al.*, 2016). *Content* needs to be stored in a standard way that can be efficiently migrated to future systems. Standards such as the OAI-ORE (Open Archives Initiative, Object Reuse and Exchange) have helped to transform *content* management in repositories. OAI-ORE can bind knowledge objects into publications and allow the reuse of knowledge objects by storing these items in collections (Tarrant *et al.*, 2009).

Retrieval

Two broad types of information retrieval are the *pull* model, that involves search for and retrieval of information based on specific user queries, and the *push* model, where information is automatically retrieved and delivered to the potential user based on some predetermined criteria (Alavi and Leidner, 2001). Retrieval is enhanced by search support functions, such as online classification schemes, for example MeSH (Medical Subject Headings), search tips and Boolean search options (Maier, 2007). Information retrieval entails finding research references based on search algorithms that interrogate internet or database metadata or full-text articles. Researchers, funders, industry, research support staff and consumers (*the people element*) apply tacit and explicit knowledge to create search strategies to retrieve research output. PubMed provides a search and retrieval platform for biomedical and life sciences literature with the aim of improving health—both globally and personally (US National Library of Medicine, 2020c). The retrieval interface for PubMed received a major overhaul in 2020, demonstrating the importance of constant improvement *processes* and the *technology* (Collins, 2019). The new release of PubMed included improvements to advanced search features, term mapping based on an algorithm to retrieve the best match, new cite links and additional search filters. Plain language summaries for research articles are made available by publishers. The new PubMed links to secondary source sites such as ClinicalTrials.gov, GenBank, Figshare and Dryad when these sources are available from a research article.

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) standard underpins the development and promotes interoperability standards that help to facilitate the efficient dissemination of *content*. The fundamental technological framework and standards are a means to open up access to a range of digital objects ("Open Archives Initiative Protocol for Metadata Harvesting," n.d.)

The *process* for achieving gold standard quality levels of systematic review searches are defined, for example by the Cochrane Handbook 2019, the Campbell Methods Guides 2016, the CEE Guidelines and the Standards for Environmental Evidence synthesis 2018 (Collaboration for Environmental Evidence, 2018, Lefebvre *et al.*, 2019, Kugley *et al.*, 2017). It is not possible to rely upon open systems alone for search retrieval, as there is "no [avoiding] proprietary search systems if one attempts a rigorous systematic review"(Gusenbauer and Haddaway, 2020).

OBR *technology* for retrieval of research output combines interoperable systems that aggregate *content* (publications and data) from other sources; for example, US PMC and Europe PMC have APIs for reuse of *content* where permitted. OpenAIRE, Unpaywall, PubMed Linkout are linking tools bringing together disparate *content* and have revolutionized access to the effective retrieval of research output beyond organizational boundaries (Artini *et al.*, 2015).

International biomedical research resources exist, such as the Entrez Programming Utilities (E-utilities) at the US National Center for Biotechnology Information (NCBI) for biomedical research objects and national approaches such as the NLA's TROVE for multi-disciplinary coverage; these are two examples that offer open *content* for retrieval of biomedical knowledge (Sayers, 2010-). In addition, there are meta-search platforms (*technology and content elements*) for retrieval of vast biomedical research objects, e.g., Accessss, Epistemonikos and Trip (Health Sciences Library, 2020). Other directories of repositories to widen the net for sourcing biomedical research *content* are the Directory of Open Access Repositories (OpenDOAR) and SATORI (Semantic AnnoTations and Ontological Relations Interface) (University of Nottingham (UK), 2005-, Lekschas and Gehlenborg, 2018).

Dissemination

Dissemination ensures knowledge is available to those who need it (Kingston, 2012). Presentations at conferences, personal communications and systems, such as social media, are used to disseminate research objects. Informal mechanisms, such as unscheduled meetings or seminars may be effective in promoting socialization but may preclude wide dissemination (Holtham and Courtney, 1998). Repositories may be most effective means for disseminating knowledge that can be readily generalized to other contexts (Alavi and Leidner, 2001).

The *process* of dissemination is often detailed in policies, procedures and guidelines in relation to submissions for OBR. Organizations such as the NHMRC, ARC, NIH, universities, other research bodies and publishers specify the timing and dissemination requirements in their open access policies. The US Public Access Policy had a significant influence on depositing of research in the US PMC by researchers (*people*) by imposing penalties for failure to comply with the requirement to publish their research openly.

One billion articles are retrieved each year from US PMC by researchers (*people* and *technology*), which demonstrates the dissemination success of this biomedical repository (NLM Program Manager, 2019). The Manager of PMC Canada reported that their site had four million article downloads and that most of these were made by users in China. In relation to dissemination and as a security measure, bulk downloads from the US PMC and Europe PMC are restricted due to copyright. Social media sites, such as ResearchGate and Academia.com foster online communities of practice and rely upon *content* from repositories such as PMC International.

Bibliographies and reference lists are effective at setting the format for the dissemination of research objects (*content*) and these are based on referencing standards, which are detailed in library guides. These guides are managed outside of repositories, although they are integral to the quality and consistency of reporting biomedical knowledge.

Transfer

Transfer is the conveyance of knowledge from one place, person, organization to another such entity (Major and Cordey-Hayes, 2000). The most important aspect of KM in the organizational setting is the transfer of knowledge to the location where it is needed and can be used (Major and Cordey-Hayes, 2000). The transfer of public research resources from government to universities is intended to generate common knowledge, provide instruction, and serve the public (Stevens and Bagby, 2001). Transfer involves clarification of the terms and conditions between relevant parties in relation to use of the *content*. Transfer involves transmission (sending or presenting knowledge to a potential recipient) and absorption by that person or group. "Transmission and absorption together have no value if the new knowledge does not lead to some change in behaviour, or the development of some idea that leads to new behaviour" (Prusak and Davenport, 1998).

Executive and senior staff (*people*) in research organizations and industry refer to repository output to identify *content* relevant to transfer agreements. The knowledge transfer process takes place through patenting, licensing, contracts, trade secrets, joint ventures with inventors and commercial spin-offs (Stevens and Bagby, 2001).

The Mind the Gap report details systems *(technology)* for open knowledge databases and search platforms (Maxwell *et al.*, 2019). Text and data mining systems are key *technologies* to aid the technology transfer process. Integration of repositories with Current Research Information Systems (CRIS) occurs in institutions as a means to optimize research knowledge (Summers and Evans, 2020?, euroCris, 2020). *Content* can be transferred and interoperable between systems based on international standards. For example, the FHIR (Fast Healthcare Interoperability Resources) is available for the transfer of healthcare information, including research articles (Health Level Seven International, 2019).

The *people* element dominates in the transfer KM process, as identified during creation of the framework for an OBR. It is evident that the interoperable characteristic of the research data (*content)* needs to seamlessly move to interdependent systems (*technology)*, such as CRIS, as a key aspect of the transfer *process*.

Translation

Knowledge translation involves "the synthesis, dissemination, exchange, and ethically sound application of knowledge to improve health, provide more effective health services and products, and strengthen the health care system" (Straus *et al.*, 2009). The conundrum for biomedical researchers (*people*) can be in applying findings from the research literature as a result of irreproducible findings. Biomedical researchers from drug companies have reported that approximately one-quarter of high-profile papers are reproducible. The gravity of this problem is indicated by the cost of irreproducible published results, which is estimated to be US$28 billion (Freedman *et al.*, 2015). Contrastingly, "big data analytics for the medical field, is viewed as a potential panacea that will potentially save more than $300 billion per year in US health-care costs" (Luo *et al.*, 2016). The future challenge for OBR is to achieve the *process* and *technological* developments necessary to make quality reproducible research *content* available for translation.

The translation of research is the time-lag between biomedical research and its adoption in healthcare and by the wider society. In relation to OBR, translation of research is informed through the linking *process* of the research grant details to their research output. For example, US PMC and Europe PMC (*technology*) allows principal investigators (*people*) to link their articles to their grant information (*content*). *Technology* systems such as Impactstory are available for researchers to generate reports on the online impact of their research (Priem and Piwowar, 2012). Data is extracted from the repository for analytical reporting in systems such as ImpactStory that are *technology* systems complementary to open *content* repositories.

Summing up the KM system framework

The audit of the *people, process, technology* and *content* elements throughout the nine KM open biomedical processes reveals that the majority of the elements are connected, although some elements are independent from a repository platform. The foremost elements of a repository include: the *technology* in representation; the *people, technology* and *content* within storage; and the *technology* and *content* within dissemination.

The empirical data collected demonstrated that applying the KM system framework has strong support from library practitioners and researchers as a means for the Australasian biomedical and health sciences research sectors to increase their collaboration beyond the organizational silos that presently restrict their impact. Establishing an Australasia biomedical KM approach would help to avoid the vast amount of duplicate effort that occurs in managing institutional repositories.

For biomedical researchers, the openly available PubMed, MEDLINE and PMC are the foundation, primary research repositories. As funders, such as the ARC and NHMRC, have open access policies directing researchers to openly publish articles, considering an Australasian OBR for reporting on research performance is a means to achieve funding body compliance. Institutional executives and open access leaders view the PMC International system as a means to manage and review the output of biomedical research linked to grant details, to help avoid duplication of research and link related findings.

A proposed Australasian OBR, as a member or regional node of PMC International has the potential to achieve a number of goals. These include: reducing the duplication of effort and the fragmented and incomplete access to health research output that presently exists with institutional repositories; preserving health research and associated data for present and future generations of users throughout the world; and producing quality metadata which is widely discoverable.

Study scope, limitations and future research

The current research has focused on biomedicine. This is, firstly, because the field generates a prolific amount of world research output and Australia is in the top twenty countries with the most biomedical publications (National Science Board, 2019, Xu *et al.*, 2015). Additionally, in matters of life and death, access to health research output should have no barriers. This research builds on the worldwide effort to reduce barriers, in particular paywalls to health knowledge. There are internationally established classification and information management schemes available to build upon. Further to this, an open biomedical repository proof of concept can potentially expand to other disciplinary areas to achieve a future regional multi-disciplinary repository.

Testing the knowledge management system framework on other types of repository, Cycle Three of this research, will identify any gaps in this approach. Future research is also required to reflect on the strengths and weaknesses of the *people, process, technology* and *content* elements from adoption of the framework.

Conclusion

The aim of open repositories is to facilitate the route from basic research results to healthcare solutions and to provide existing and new knowledge at the point of need. This research makes a unique contribution to the field of KM, as no previous studies have reported on the intersection between open science and a KMS. In the Australasian region, biomedical research is governed at the information management level, whereas to achieve informed healthcare decisions, synthesized knowledge is required to support action. This research confirms that production of biomedical knowledge transcends organizational boundaries and can benefit if conceptualized as a KMS.

Achieving Australia's National Digital Health Strategy requires networks of healthcare academics, researchers and supporting information professions to collaborate on a system that is respected by health consumers as safe, seamless and secure (Australian Digital Health Agency, 2018). The proposed KM system framework is a powerful tool that can be used to align the components that underpin the effectiveness and sustainability of an open biomedical repository. It is a tool to assist the transition from organizational silos to an open science environment.   Adoption of the framework can ultimately assist research communities to foster world class collaboration and corroboration through systematic and coordinated effort informed by KM theory and practice.

References

Alavi, M. and Leidner, D.E. (2001), "Review: knowledge management and knowledge management systems: conceptual foundations and research issues", *MIS Quarterly,* Vol. 25 No. 1, pp. 107-136.

Arlitsch, K. and Grant, C. (2018), "Why so many repositories? examining the limitations and possibilities of the institutional repositories landscape", *Journal of Library Administration,* Vol. 58 No. 3, pp. 264-281.

Arrow, K. (1962), "Economic welfare and the location of resources for invention", in *The rate and direction of inventive activity*, Princeton University Press, Princeton, NJ, pp. 155-173.

Artini, M., Atzori, C., Bardi, A., La Bruzzo, S., Manghi, P. and Mannocci, A. (2015), "The OpenAIRE literature broker service for institutional repositories", *D-Lib Magazine,* Vol. 21 No. 11/12, available at: https://doi:10.1045/november2015-artini

Australian Digital Health Agency (2018), "Australia's National Digital Health Strategy: safe, seamless and secure: evolving health and care to meet the needs of modern Australia", available at: https://conversation.digitalhealth.gov.au/australias-national-digital-health-strategy (accessed 19 May 2020).

Australian Living Evidence Consortium (2018), "Medical Research Future Fund: 2018-2020 priorities", available at: https://australia.cochrane.org/sites/australia.cochrane.org/files/public/uploads/LivingEvidence/aust_living_evience_consortium_mrff_priorities_2018-2020_final_for_web.pdf (accessed 23 April 2020).

Australian Research Council (2013), "ARC Open Access Policy Version 2017.1", available at: https://www.arc.gov.au/policies-strategies/policy/arc-open-access-policy (accessed 18 August 2020).

Baskerville, R. and Dulipovici, A. (2006), "The theoretical foundations of knowledge management", *Knowledge Management Research & Practice,* Vol. 4 No. 2, pp. 83-105, available at: http://10.1057/palgrave.kmrp.8500090.

Becerra-Fernandez, I. and Sabherwal, R. (2015), *Knowledge management: systems and processes*, 2nd ed., Ebooks Corporation, New York, NY.

Bhatt, G.D. (2000), "Organizing knowledge in the knowledge development cycle", *Journal of Knowledge Management,* Vol. 4 No. 1, pp. 15-26, available at: https://doi.org/10.1108/13673270010315371.

Burstein, F. and Gregor, S. (1999), "The systems development or engineering approach to research in information systems: an action research perspective", in *Proceedings of the 10th Australasian Conference on Information Systems*, Victoria University of Wellington, New Zealand, pp. 122-134.

Centobelli, P., Cerchione, R. and Esposito, E. (2019), "Efficiency and effectiveness of knowledge management systems in SMEs", *Production Planning & Control,* Vol. 30 No. 9, pp. 779-791, available at: https://doi.org/10.1080/09537287.2019.1582818.

Chen, X., Dallmeier-Tiessen, S., Dasler, R., Feger, S., Fokianos, P., Gonzalez, J.B., Hirvonsalo, H., Kousidis, D., Lavasa, A. and Mele, S. (2019), "Open is not enough", *Nature Physics,* Vol. 15 No. 2, pp. 113-119.

Chhim, P.P., Somers, T.M. and Chinnam, R.B. (2017), "Knowledge reuse through electronic knowledge repositories: a multi theoretical study", *Journal of Knowledge Management,* Vol. 21 No. 4, pp. 741-764.

Collaboration for Environmental Evidence (2018) *Guidelines and standards for evidence synthesis in environmental management,* version 5, Pullin, A.S., Frampton, G.K., Livoreil, B. and Petrokofsky, G., eds., available at: http://www.environmentalevidence.org/information-for-authors [accessed 1 May 2020].

Collins, M. (2019) "The new PubMed is here", *NLM Tech Bull*, Nov-Dec, (431), e3, available at: https://www.nlm.nih.gov/pubs/techbull/nd19/nd19_pubmed_new.html [accessed 18 August 2020].

Committee on Publication Ethics (2020), "Promoting integrity in scholarly research and its publication COPE [Home page]", available at: https://publicationethics.org/ (accessed 29 April 2020).

Cook, C.E., Bergman, M.T., Cochrane, G., Apweiler, R. and Birney, E. (2018), "The European Bioinformatics Institute in 2017: data coordination and integration", *Nucleic acids research,* Vol. 46 No. D1, pp. D21-D29.

Dalkir, K. (2017), *Knowledge management in theory and practice*, 3rd ed., Cambridge, Massachusetts : The MIT Press.

Digital preservation handbook, (2015) 2nd ed., Beagrie, N. and Kilbridie, W., eds., Digital Preservation Coalition, available at: https://www.dpconline.org/handbook [accessed 8 May 2020].

Elden, M. and Chisholm, R.F. (1993), "Emerging varieties of action research: introduction to the special issue", *Human Relations,* Vol. 46 No. 2, pp. 121-142, available at: https://search.proquest.com/docview/1474345838?accountid=12528

euroCris (2020), "Why does one need a CRIS [Current Research Information Systems]?", available at: https://www.eurocris.org/why-does-one-need-cris (accessed 4 May 2020).

Europe PMC (2019), "About Europe PMC", available at: https://europepmc.org/About (accessed 29 April 2020).

Europe PMC (2020), "Funders", available at: https://europepmc.org/Funders/ (accessed 27 April 2020).

Freedman, L.P., Cockburn, I.M. and Simcoe, T.S. (2015), "The economics of reproducibility in preclinical research", *PLoS Biol,* Vol. 13 No. 6, p. e1002165, available at: https://journals.plos.org/plosbiology/article%3Fid=10.1371%252Fjournal.pbio.1002165.

Gabbay, J. and le May, A. (2010), *Practice-based evidence for healthcare clinical mindlines*, Taylor & Francis, Hoboken, NJ.

Greenberg, S.J. and Gallagher, P.E. (2009), "The great contribution: Index Medicus, Index-Catalogue, and IndexCat", *Journal of the Medical Library Association* Vol. 97 No. 2, pp. 108-113.

Gusenbauer, M. and Haddaway, N.R. (2020), "Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed and 26 other resources", *Research Synthesis Methods,* Vol. 11 No. 2, pp. 181-217, available at: https://doi.org/10.1002/jrsm.1378.

Halbwirth, S. and Sbarcea, K. (2005), "The spotlight on knowledge management, joint presentation", paper presented at *NSW KM Forum*, 7 September 2005, Sydney, Standards Australia.

Harvey, H.B. and Weinstein, D.F. (2017), "Predatory publishing: an emerging threat to the medical literature", *Academic Medicine,* Vol. 92 No. 2, pp. 150-151, available at: https://doi:10.1097/ACM.0000000000001521.

Health Level Seven International (2019), "Overview - FHIR v4.0.1", available at: https://www.hl7.org/fhir/overview.html (accessed 4 May 2020).

Health Sciences Library, McMaster University, (2020), "Resources for evidence-based practice: the 6S pyramid", available at: https://hslmcmaster.libguides.com/ebm (accessed 27 March 2020.

Hess, C. and Ostrom, E. (2006), "A framework for analyzing the knowledge commons", in Hess, C. and Ostrom, E. (Eds.), *Understanding Knowledge as a Commons: from Theory to Practice*, MIT Press, Cambridge, MA, pp. 41-81.

Holtham, C. and Courtney, N. (1998), "The executive learning ladder: a knowledge creation process grounded in the strategic information systems domain", in *Proceedings of the Fourth Americas Conference on Information Systems*, Baltimore, MD, pp. 594-597.

Hristovski, D., Peterlin, B., Mitchell, J.A. and Humphrey, S.M. (2005), "Using literature-based discovery to identify disease candidate genes", *International Journal of Medical Informatics,* Vol. 74 No. 2-4, pp. 289-298.

International Committee of Medical Journal Editors (2020), "About ICMJE", available at: http://www.icmje.org/about-icmje/ (accessed 23 April 2020).

International Organization for Standardisation (2017), "ISO 20614:2017", available at: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/85/68562 .html (accessed 29 April 2020).

Jennex, M.E., Smolnik, S. and Croasdell, D.T. (2009), "Towards a consensus knowledge management success definition", *VINE,* Vol. 39 No. 2, pp. 174-188.

Jensen, L.J., Saric, J. and Bork, P. (2006), "Literature mining for the biologist: from information retrieval to biological discovery", *Nature reviews genetics,* Vol. 7 No. 2, pp. 119-129.

Joo, S., Hofman, D. and Kim, Y. (2018), "Investigation of challenges in academic institutional repositories: a survey of academic librarians", *Library Hi Tech*, available at: https://doi.org/10.1108/LHT-12-2017-0266.

Kiley, R. (2018) "Europe PMC: interview with the Development Lead, Open Research, Wellcome Trust", *Interview date* April 2018.

Kim, J.H. (2015), "Europe PubMed Central and linked data", paper presented at *Biohackathon EMBL-EBI*, Nagasaki, available at: https://tinyurl.com/y3ueoorm. (accessed 27 April 2020).

Kingston, J. (2012), "Choosing a knowledge dissemination approach", *Knowledge and Process Management,* Vol. 19 No. 3, pp. 160-170, available at: https://doi.org/10.1002/kpm.1391.

Kramer, B. and Bosman, J. (2016), "Innovations in scholarly communication - global survey on research tool usage", *F1000Res,* Vol. 5, p. 692, available at: http://doi:10.12688/f1000research.8414.1.

Kramer, B. and Bosman, J. (2017), "Wheel of Open Science practices (image)", available at: https://figshare.com/articles/Wheel_of_Open_Science_practices_image_/4628014 (accessed 27 April 2020).

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A.-M.K., Hammerstrøm, K. and Sathe, N. (2017), "Searching for studies: a guide to information retrieval for Campbell", *Campbell Systematic Reviews,* Vol. 13 No. 1, pp. 1-73, available at: https://doi.org/10.4073/cmg.2016.1.

Landa, E. (2018) "PMC Canada: interview", *Interview date* January 2018.

Lefebvre, C., Glanville, J., Briscoe, S., Littlewood, A., Marshall, C., Metzendorf, M.I., Noel-Storr, A., Rader, T., Shokraneh, F., Thomas, J. and Wieland, L.S. (2019) "Searching for and selecting studies", Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.0*, Cochrane, Chichester UK, available at: https://training.cochrane.org/handbook/current/chapter-04#section-4-7 [accessed 14 August 2020].

Lekschas, F. and Gehlenborg, N. (2018), "SATORI: a system for ontology-guided visual exploration of biomedical data repositories", *Bioinformatics,* Vol. 34 No. 7, pp. 1200-1207, available at: https://doi.org/10.1093/bioinformatics/btx739.

LePan, N. (2020), "A visual history of pandemics", *World Economic Forum*, available at: https://www.weforum.org/agenda/2020/03/a-visual-history-of-pandemics/ (accessed 21 April 2020).

Lu, Z. (2011), "PubMed and beyond: a survey of web tools for searching biomedical literature", *Database,* Vol. 2011, p. baq036, available at: https://doi.org/10.1093/database/baq036.

Luo, J., Wu, M., Gopukumar, D. and Zhao, Y. (2016), "Big data application in biomedical research and health care: a literature review", *Biomedical informatics insights,* Vol. 8, available at: https://doi.org/10.4137/BII.S31559.

Maier, R. (2007), *Knowledge management systems information and communication technologies for knowledge management*, 3rd ed.*,* Springer, Berlin.

Major, E. and Cordey-Hayes, M. (2000), "Knowledge translation: a new perspective on knowledge transfer and foresight", *Foresight-The journal of future studies, strategic thinking and policy,* Vol. 2 No. 4, pp. 411-423.

Marques Júnior, E., Gobbo, J.A., Fukunaga, F., Cerchione, R. and Centobelli, P. (2020), "Use of knowledge management systems: analysis of the strategies of Brazilian small and medium enterprises", *Journal of Knowledge Management,* Vol. 24 No. 2, pp. 369-394, available at: https://doi.org/10.1108/JKM-06-2019-0334.

Maxwell, J.W., Hanson, E., Desai, L., Tiampo, C., O'Donnell, K., Ketheeswaran, A., Sun, M., Walter, E. and Michelle, E. (2019) *Mind the gap: a landscape analysis of open source publishing tools and platforms*, The MIT Press, Cambridge, MA, available at: https://mindthegap.pubpub.org/.

McLean, N. and Lynch, C. (2003), "Interoperability between information and learning environments–bridging the gaps: a joint white paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information [Draft]", available at: https://pdfs.semanticscholar.org/81f8/1577d6b59bcfee598d39cc031d8997cfeeb8.pdf (accessed 26 April 2020).

National Archives of Australia (2018), "Digital Preservation Policy", available at: https://www.naa.gov.au/about-us/our-organisation/accountability-and-reporting/archival-policy-and-planning/digital-preservation-policy (accessed 29 April 2020).

National Health and Medical Research Council (Australia) (2016), "NHMRC statement on data sharing", available at: https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing (accessed 28 September 2017).

National Health and Medical Research Council (Australia) (2018), "Open Access Policy", available at: https://nhmrc.gov.au/about-us/publications/open-access-policy (accessed 16 August 2020).

National Science Board, National Science Foundation, (2019) "Publication Output: U.S. Trends and International Comparisons", *Science & Engineering Indicators 2020*, National Science Foundation, Alexandria, VA, available at: https://ncses.nsf.gov/pubs/nsb20206/publication-output-by-region-country-or-economy [accessed 16 August 2020].

NLM Program Manager (2018) "Update on PMC", *Interview date* 21 May 2018.

NLM Program Manager (2019) "Update on PMC", *Interview date* 12 April 2019.

Nonaka, I. and Takeuchi, H. (1995), *The knowledge-creating company: How Japanese companies create the dynamics of innovation*, Oxford University Press, Oxford, UK.

Ohno-Machado, L., Sansone, S.-A., Alter, G., Fore, I., Grethe, J., Xu, H., Gonzalez-Beltran, A., Rocca-Serra, P., Gururaj, A.E. and Bell, E. (2017), "Finding useful data across multiple biomedical data repositories using DataMed", *Nature Genetics,* Vol. 49 No. 6, pp. 816-819, available at: https://doi.org/10.1038/ng.3864.

"Open Archives Initiative Protocol for Metadata Harvesting", (n.d.), available at: https://www.openarchives.org/organization/ (accessed 1 May 2020).

Pinfield, S., Salter, J., Bath, P.A., Hubbard, B., Millington, P., Anders, J.H. and Hussain, A. (2014), "Open access repositories worldwide, 2005–2012: past growth, current characteristics, and future possibilities", *Journal of the Association for Information Science and Technology,* Vol. 65 No. 12, pp. 2404-2421.

Piwowar, H. (2019), "The breakout moment for open repositories is now. How can we build the best future for our users? [Closing plenary keynote]", paper presented at *Open Repositories, June 10-13*, Hamburg, Germany, available at: https://lecture2go.uni-hamburg.de/l2go/-/get/v/24979 (accessed 26 April 2020).

Piwowar, H., Priem, J. and Orr, R. (2019), "The Future of OA: A large-scale analysis projecting Open Access publication and readership", *bioRxiv*, p. 795310, available at: https://doi.org/10.1101/795310.

Polanyi, M. (1966), *The tacit dimension*, Doubleday and Company Inc., Garden City, NY.

Priem, J. and Piwowar, H. (2012), "The launch of ImpactStory: using altmetrics to tell data-driven stories", *Impact of Social Sciences Blog*, available at: https://blogs.lse.ac.uk/impactofsocialsciences/2012/09/25/the-launch-of-impactstor/ (accessed 16 August 2020).

Prusak, L. and Davenport, T. (1998), *Working knowledge: how organizations manage what they know*, Harvard Business School Press, Boston, MA.

Regazzi, J.J. (2015), *Scholarly communications: a history from content as king to content as kingmaker*, Rowman & Littlefield, Lanham.

Roberts, K., Gururaj, A.E., Chen, X., Pournejati, S., Hersh, W.R., Demner-Fushman, D., Ohno-Machado, L., Cohen, T. and Xu, H. (2017), "Information retrieval for biomedical datasets: the 2016 bioCADDIE dataset retrieval challenge", *Database,* Vol. 2017, available at: https://doi.org/10.1093/database/bax068.

Sabharwal, A. (2019) "The transformative role of institutional repositories in academic knowledge management", Gunjal, B. (Ed.) *Managing knowledge and scholarly assets in academic libraries*, IGI Global, Hershey, PA, available at: http://doi:10.4018/978-1-5225-1741-2.ch006.

Saito, A., Umemoto, K. and Ikeda, M. (2007), "A strategy-based ontology of knowledge management technologies", *Journal of Knowledge Management,* Vol. 11 No. 1, pp. 97-114, available at: https://doi.org/10.1108/13673270710728268.

Sayers, E. (2010-) *A general introduction to the E-utilities. In Entrez Programming Utilities help [Internet]*, US National Center for Biotechnology Information, Bethesda MD, available at: https://www.ncbi.nlm.nih.gov/books/NBK25497/.

Schumpeter, J.A. (1934), *The theory of economic development: an inquiry into profits, capital, credit, interest and the business cycle*, Harvard University Press, Cambridge, MA.

Science Europe Working Group on Open Access (2015), "Science Europe principles on open access to research publications", available at: http://www.scienceeurope.org/wp-content/uploads/2015/10/SE_POA_Pos_Statement_WEB_FINAL_20150617.pdf (accessed 23 April 2020).

"Sorbonne declaration on research data rights", (2020), available at: https://go8.edu.au/wp-content/uploads/2020/01/Sorbonne-declaration.pdf (accessed 17 August 2020).

Standards Australia. (2005), *Knowledge management: a guide AS5037*, Standards Australia International, Sydney.

Stevens, J.M. and Bagby, J.W. (2001), "Knowledge transfer from universities to business: returns for all stakeholders?", *Organization,* Vol. 8 No. 2, pp. 259-268, available at: http://journals.sagepub.com/doi/abs/10.1177/1350508401082012.

Stevens, R., Rector, A. and Hull, D. (2010), "What is an ontology?", *Ontogenesis*, available at: http://ontogenesis.knowledgeblog.org/66 (accessed 29 April 2020).

Straus, S.E., Tetroe, J. and Graham, I. (2009), "Defining knowledge translation", *CMAJ,* Vol. 181 No. 3-4, pp. 165-168, available at: https://doi.org/10.1503/cmaj.081229.

Summers, B. and Evans, J. (2020?), "A staged path to a fully integrated CRIS & repository [Presentation]", available at: https://dspacecris.eurocris.org/bitstream/11366/1009/1/Haplo%20EuroCris%20Tech%20Case%20Study.pdf (accessed 17 August 2020).

Tarrant, D., O'Steen, B., Brody, T., Hitchcock, S., Jefferies, N. and Carr, L. (2009), "Using OAI-ORE to Transform Digital Repositories into Interoperable Storage and Services Applications", *Code4Lib Journal,* No. 6, available at: https://tinyurl.com/w2cfm38.

Tuomi, I. (1999), "Data Is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory", *Journal of Management Information Systems,* Vol. 16 No. 3, pp. 103-117, available at: https://doi.org/10.1080/07421222.1999.11518258.

UK Equator Centre (2020), "The EQUATOR Network: Enhancing the QUAlity and Transparency Of health Research", available at: https://www.equator-network.org/ (accessed 29 April 2020).

University of Nottingham (UK) (2005-), "Directory of Open Access Repositories (OpenDOAR)", available at: http://v2.sherpa.ac.uk/opendoar/information.html (accessed 3 May 2020).

US Department of Health and Human Services National Institutes of Health (2008), "NIH Public Access Policy ", available at: https://publicaccess.nih.gov/policy.htm (accessed 18 August 2020).

US National Library of Medicine (2016) "Changes to the NLM Data Distribution Program", *NLM Tech Bull*, Nov-Dec, (413), b4, available at: https://www.nlm.nih.gov/pubs/techbull/nd16/brief/nd16_data_distrib.html.

US National Library of Medicine (2019a), "About the NLM", available at: https://www.nlm.nih.gov/about/index.html (accessed 22 January 2020).

US National Library of Medicine (2019b), "Unified Medical Language System (UMLS)", available at: https://www.ncbi.nlm.nih.gov/pubmed/ (accessed 29 April 2020).

US National Library of Medicine (2020a), "MeSH", available at: https://www.ncbi.nlm.nih.gov/mesh (accessed 17 August 2020).

US National Library of Medicine (2020b), "PMC overview", available at: https://www.ncbi.nlm.nih.gov/pmc/about/intro/ (accessed 17 August 2020).

US National Library of Medicine (2020c), "PubMed overview", available at: https://pubmed.ncbi.nlm.nih.gov/about/ (accessed 17 August 2020).

US National Library of Medicine, National Institutes of Health (2018), "PMC International", available at: https://www.ncbi.nlm.nih.gov/pmc/about/pmci/ (accessed 17 August 2020).

US National Library of Medicine, National Institutes of Health, (2019), "How to include a journal in PMC", available at: https://www.ncbi.nlm.nih.gov/pmc/pub/addjournal/ (accessed 29 April 2020).

Watson, M. (2015), "When will 'open science' become simply 'science'?", *Genome biology,* Vol. 16 No. 1, pp. 101-101, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4436110/.

Wellcome Trust (2020), "Open access policy", available at: http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm (accessed 17 August 2020).

Wiig, K.M. (1997), "Knowledge management: an introduction and perspective", *Journal of Knowledge Management,* Vol. 1 No. 1, pp. 6-14.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B. and Bourne, P.E. (2016), "The FAIR guiding principles for scientific data management and stewardship", *Scientific data,* Vol. 3 No. 160018, available at: https://doi.org/10.1038/sdata.2016.18.

Williamson, K. (2013), "Action research: theory and practice", in Williamson, K. and Johanson, G. (Eds.), *Research methods : information, systems and contexts*, Tilde University Press, Prahran, Vic., pp. 188-202.

Xu, Q., Boggio, A. and Ballabeni, A. (2015), "Countries' biomedical publications and attraction scores. A PubMed-based assessment ", *F1000Research,* Vol. 3 No. 292, available at: https://f1000research.com/articles/3-292/v2.

Yew Wong, K. and Aspinwall, E. (2005), "An empirical study of the important factors for knowledge-management adoption in the SME sector", *Journal of Knowledge Management,* Vol. 9 No. 3, pp. 64-82, available at: https://doi.org/10.1108/13673270510602773.