Education for Information 36 (2020) 281–299
DOI 10.3233/EFI-190259
IOS Press

Changes in academic libraries in the era of Open Science

Stefka Tzanova
York College, City University of New York, New York, USA
E-mail: stzanova@york.cuny.edu

In this paper we study the changes in academic library services inspired by the Open Science movement and especially the changes prompted from Open Data as a founding part of Open Science. We argue that academic libraries face the even bigger challenges for accommodating and providing support for Open Big Data composed from existing raw data sets and new massive sets generated from data driven research. Ensuring the veracity of Open Big Data is a complex problem dominated by data science. For academic libraries, that challenge triggers not only the expansion of traditional library services, but also leads to adoption of a set of new roles and responsibilities. That includes, but is not limited to development of the supporting models for Research Data Management, providing Data Management Plan assistance, expanding the qualifications of library personnel toward data science literacy, integration of the library services into research and educational process by taking part in research grants and many others. We outline several approaches taken by some academic libraries and by libraries at the City University of New York (CUNY) to meet necessities imposed by doing research and education with Open Big Data – from changes in libraries' administrative structure, changes in personnel qualifications and duties, leading the interdisciplinary advisory groups, to active collaboration in principal projects.

Keywords: Open science, open data, open big data, data science, research data management, institutional repositories, academic libraries, science literacy, data science literacy

1. Introduction

Academic libraries have been at the core of academic institutions for centuries. Their evolution is exclusively related to the evolution of parent institutions, as the information services provided by their libraries are uniquely related to the specific needs of education and research communities at the University. Academic libraries "can be viewed as information systems that both reflect and influence, and even help to create, paradigms and authority, for they set limits in various ways on the ideas and information available to users" (Dain, 1990), and evolve with the University in order to provide information services to support the Open Science model. Consequently, it is very important for the academic libraries to understand the nature and scope of Open Science in order to adapt and expand their services to a new research and educational paradigm. In that process the academic libraries seek to maximize the research and educational potential of digital technologies and communications by providing open and universal access to data sets in repositories related to particular projects in addition to their traditional role of providing access to primary and secondary resources (such as monographs, serials, databases, government documents, etc.). It is important to emphasized that Open Science is not "free for all", rather it is "science on nominal cost" or "science without overhead", because expenses for building up and supporting the institutional repositories, various databases (local and global), maintaining the educational resources, etc., are all jointly covered by institutional budgets, grants, and public funds. Thus the rate of adoption of Open Science by academic libraries is contingent on institutional policies and consequently on institutional financial commitment.

Even if is clear that the ultimate goal of Open Science is to maximize the research output by removing barriers and promoting collective science, a single universal definition of Open Science does not exist. Some authors discuss Open Science only as establishment of Open Access Publishing and Open Source and identify the development of Internet and easy cross boundary communications as main driving force (Grand et al., 2012). Others analyze Open Science in a framework of communication theory (Kulczycki, 2016). In a context of the science-society relation, Open Science has been also analyzed as a cultural change, aiming to expand the economic impact of science onto society by removing the barriers in front of knowledge exchange and by changing the ways of creation, dissemination, storage, and delivery of scientific data (David, 2008). Similarly, Seaz and Martinez-Fuentes (2018) recognize Open Science as a "global movement that brings up socio-cultural and technological change, based on openness and connectivity on how research is designed, conducted captured and accessed". As a starting point in this paper we adopt the quite general Fecher's and Friesike's definition of Open Science as an "umbrella term encompassing a multiple assumptions about the future of knowledge creation and dissemination" (Fecher & Friesike, 2014).

Kraker et al. (2011) define the four instruments of Open Science: Open Access (OA), Open Source (OS), Open Data (OD), and Open Methodology (OM), all representing the application of the concept of openness toward each step in a research workflow. In the educational context, particularly in learning and teaching, Open Educational Resources (OER) have joined the movement, too. It is important to mention that despite the fact that Open Science is currently most visible in the area of "hard sciences" (due to large data sets generated by high throughput experiments and simulations) it is not limited to only the STEM fields, but is also applicable to other types of scientific research. For example Open Data project "Brain Research through Advancing Innovative Neuro Technologies" (BRAIN) explores perspective of Open Science in psychology (Hesse, 2018). The Open Access and Open Data both are the most "subject independent" components of Open Science because they appear at any study despite of its topic and scope. Consequently, and as a global trend the academic libraries emphasize on building up of a support for Open Access and OD in addition to traditional information services they provide. In particular, many academic libraries speed up their own development toward offering "non-traditional" data driven and data oriented services in order to support research requirements, scenarios and workflows typical for collaborative and highly communicative open science projects. These new data services necessitate the use of Internet based models of communication and utilization of large set of complex digital technologies. Accordingly, gaining of a set of new qualifications and skills from library staff become mandatory for the 21st century librarians (Affelt, 2015).

In this paper, we study the changes in the role of academic libraries triggered by Open Science with focus in particular on Open Data and Open Big Data and their impact on academic library services. By drawing on examples of library initiatives at the City University of New York (CUNY), we illustrate how academic libraries are facilitating the adoption and implementation of Open Science concept at CUNY. This paper is organized as follows: first we will discuss Open Science as a manifestation of the efforts to recover the "public nature of science" but in a context of changed science-society relation triggered by technological achievements. Second, we will outline those aspects of Open Science which are relevant from an academic library perspective. Third we will discuss the new roles that academic libraries have taken on under the model of Open Science with emphasis on specific new role(s), the librarians and supporting staff have taken on in a context of Open Data and Open Big Data beyond Open Access. Since OD/OBD denotes open access to raw research data (or databases with raw data), scripts and methodologies, OD requires a different and specific skills set from librarians. We explore different approaches and we will propose practical recommendations aiming to help librarians working in this space to adapt quickly to Open Science realm. Lastly we will discuss the various new roles that academic libraries have taken on under the model of Open Science – from educator, enabler, research diversity promoter to mediator and adviser. In addition we will discuss the specific new role(s) of libraries in a context of Big Data, and we will describe how librarians have successfully adapted to and taken on new roles, providing a roadmap for those who would like to follow.

2. Open Science – revolution or evolution in scientific research

Scientific research can be defined as planned, organized, and systematic collection, interpretation and analysis of data done with the purpose of contributing to global scientific knowledge. In other words, scientific research has intrinsic public nature – as the French physiologist Claude Bernard once said, "Art is I, science is we" (Bernard, 1957). Ensuring open and reproducible research has become a main goal across scientific communities and is supported by political circles and funding organizations (Boulton, 2016). The understanding is that open and reproducible research practices enable scientific re-use, accelerating future projects and discoveries in any discipline (Chen et al., 2019).

However, the current system of dissemination of scientific knowledge does not serve the public nature of science. The subscription based model, professed by journal publishers and their for-profit system of dissemination based on marketing, does not support the research process. The obstacle is purely financial – almost 75% of published scientific articles are behind paywalls therefore accessible to only those who work at institutions able to afford the steep subscriptions (Tennant et al., 2016). Unfortunately subscription to all peer-reviewed journals is not affordable for a single individual, research institute or university, meaning that the potential impact of published research is never fully reached due to financial limitations. Second, additional barriers are posed by the wide-spread disagreement regarding data and curated samples availability and their corresponding metadata, especially in the field sciences (McNutt et al., 2016). Third obstacle is data collection and data itself. Data and metadata practices of researchers often appear incomplete or deficient because data acquisition processes are different for different sciences (Van Tuyl & Whitmire, 2016). For example for laboratory scientists, data are usually computer generated, hence in digital format, therefore they can be automatically uploaded in repositories with little or no human intervention. For field sciences however, (ecologists, archeologists etc.) data collected on a filed are later recovered with large degree of human improvisation before being incorporated into data repositories (Gitleman, 2013). Finally many scientists are unwilling to share their data due to fears of exploitation of data sets, rich enough to produce several publications (Molloy, 2012).

Under subscription based journal model, sharing research data is possible into the scope of particular paper, but that does not necessarily cover the complete data sets for particular research. Indeed many journals offer mechanisms to upload research data (Stuart et al., 2018), but the sets of raw experimental and modeling data, details about non-traditional or unique experimental methodologies, results from failed experiments, results from failed theories and many other research byproducts, are rarely or never published in subscription journals.

Open Science aims to alleviate most of above problems by changing the ways knowledge is both created and disseminated across society. Increased access to research outputs might help foster a culture of greater scientific education and literacy, which in turn could have a direct impact on public policy (European Commission, 2012; Zuccala, 2010), particularly in domains such as climate change and global health, as well as increasing public engagement in scientific research. It is important to emphasize that openness and sharing the data will affect not only knowledge creation and dissemination, but will also increase effectiveness of education and data and knowledge processing. As Stodden et al. (2016) pointed out the "access to the computational steps taken to process data and generate findings is as important as access to the data themselves". However there is significant confusion about what open research data should look like and about compliance of these data with Open Knowledge/Open Data definition (Molloy, 2012). That is rational, since very little scientific content is created outside the scientific communities (Fecher & Friesike, 2014). No doubt the Open Science movement is an effort to make scientific data a public good in contrast to the expansion of intellectual property rights over knowledge.

Indeed Open Science propagation is facilitated by the development of digital technologies and the exponential growth of data produced by the global scientific community. Due to the advancement of information technologies and computers, scientific experiments generate unprecedented enormous amounts of data which can be made accessible at any place/country by any researcher via the World Wide Web. Open Science is also a direct result from changes in the research process and the increasing need of collaborative and interdisciplinary research. However it is important to mention that Open Science as global phenomenon requires as well significant socio-cultural changes at all levels along with harmonizing legislation systems and political support. In Europe the European Science Cloud (EOSC) is an umbrella for academic and research libraries, universities and research centers with the goal to provide solutions for the scientific community in the context of Open Science (Mons et al., 2017). In the US the Open Science Chain (OSC), a project in progress funded by National Science Foundation (NSF), aims to develop a cyberinfrastructure platform that would allow researchers to make available metadata and verification information about their scientific datasets and update this information as the datasets change over the time.

3. Open Science as a cultural and social phenomena

From the perspective of knowledge dissemination, Open Science goes beyond transmission of knowledge, facts, ideas or information among participants in communication channels to remapping social relations and creating new sets of social interactions. In other words the openness influences the entire process of knowledge creation. For instance scientific publications as a high quality final product have little or no socio-cultural dimension. By making all steps of the research process visible to society, open science creates the conditions to involve different and often wider social groups in the creativity process. Therefore, Open Science is more of a social and cultural phenomenon aiming to recover the founding principles of scientific research rather than an alternative form of knowledge exchange. Thus, as an object of study, Open Science should not be modeled via a simple provisional communication model, but rather with constitutive models (Craig, 1999) because Open Science is not defined only by the processes of communication between scientists even if is true that openness in science depends on application of various new communication technologies which aid both scholarly communications and research impact evaluations. Indeed, Open Science rises up on a base of economic and technological developments (for example, the ability to propagate knowledge free of charge, existence of common platforms for sharing information not limited in time and space, global social networks, Internet 2, etc.), and also goes beyond these technological achievements and changes the understanding of the value of scientific knowledge across society. As such, Open Science manifests itself as a form of social organization built on a base of technology which aims to maximize the rate of accumulation of knowledge into the society and consequently to maximize the rate of growth induced by research activities. The driving force behind Open Science are not-for-profit organizations, scientists and their professional organizations, informal groups, public libraries, academic libraries, universities, foundations, government agencies, and individuals.

4. Open Science – an academic library perspective

As mentioned above, Open Science has several constructing components that aim to support high quality reproducible science. Open Science is often described as a multifaceted notion encompassing open access to publications, open research data, open source software, open collaboration, open peer review, open notebooks, open educational resources, open monographs, citizen science, or research crowdfunding (FOSTER, 2017) in order to remove barriers in the sharing of scientific research output and raw data. In this section we will focus on OA and OD because these two components inspire the largest changes in academic libraries' services and operations under Open Science model. Historically, the driving force behind OD and OBD originated from scientific communities and not-for-profit organizations but are now the result of governments' efforts (and consequently are requirements from institutions) to set up mediated data repositories and formulate the rules and policies for sharing the research data coming from all publicly funded projects. For example, numerous pilot initiatives such "H2020 Programme" (European Commission, 2011) in Europe requires any research funded by public sources to be published in Open-Access journals and data to be stacked in Open-Access data collections. In the US, the two main research funding agencies – National Science Foundation (NSF) and National Institutes of Health (NIH) – have similar mandatory requirements in accordance with FASTR (Fair Access to Scientific and Technology Research Act), approved by the Congress in 2013, instructing all U.S. science funding agencies to provide public access to federally supported research outputs.

As defined by Open Society, Open Access (OA) is a publishing and distribution model that makes scholarly research literature – much of which is funded by taxpayers around the world – freely available to the public online, without restrictions. In the context of Open Science OA references free access to scientific publications and databases with results from studies in a particular scientific discipline(s) (e.g., metabolomics databases). More precisely the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* (Max Planck Gesellschaft, 2003) defines OA as "a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community". Many important and rich data sets are result of projects which rely on data collected from non-professional scientists (citizen science projects). As Groom (2017) points out: "citizen science data sets comprise 10% of data sets on GBIF (Global Biodiversity Information Facility), but account for 60% of all observations". Due to large costs of citizen science projects and "contrary to what many people assume, data sets from volunteers are among the most restrictive in how they can be used". Typically these open data sets are accessible to library patrons via registration. Some examples are data sets accumulated for urban microbiome project – AREM – led by the City University of New York or eBird database in ornithology. In addition, citizen science data sets are often in different formats and are almost always web based. To ensure access to these resources the academic libraries have to develop and expand their metadata services.

That includes, but is not limited to metadata consultation services to patrons, hands on tutorials and manuals.

Open Data (OD) refer to free access to repositories with raw data, workflow scripts and methodologies used in the course of actual studies. OD is rooted first in the need of scientific communities to share the tools needed for scientific research and second to ensure the reproducibility of science (Gewin, 2016). However, disseminating sets of raw data of different formats is a more complex process which goes far beyond simple repositories like BitBucket and GitHub which works quite well for OS. The engagement with OD and OBD obliges the libraries to become involved in data curation, preservation and management of digital scientific content and data sets in various formats (Heidorn, 2011). The rationale is that OD and OBD require understanding of advanced technologies in order to ensure effective re-use and sharing of raw data. The term OD is often used interchangeably with OBD but there are significant differences between them. The OD is defined by its impact since even small amount of data when made public can inspire, define or invigorate the research activity. In contrast OBD is defined by both its impact and its size. The availability of OBD has direct effect on the number and quality of collaborative projects which place the OBD in the core of both data driven research and the modern "hybrid" research model. OBD does not need to originate exclusively from government supported research. Common examples are OD sets made available through global organizations (e.g. World Bank), local governments, municipalities, institutions, NGO, electronic social media (Facebook, Twitter etc.) which are used to conduct research in the social sciences. For example the recent multi-disciplinary research project conducted by Schwartz and his team (Schwarz et al., 2013) utilizes OBD sets from Facebook to develop a framework for describing uniquely the correspondences and dissimilarities among groups of people in terms of their different languages used. OBD's nature as big rapidly changing and multifaceted object requires academic librarians to acquire knowledge and skills beyond the traditional boundaries of library science and to move toward information science. In this sense academic librarians face the challenges to understand the complexity of data science workflows and consequently to have fundamental understanding of computer logic, programming paradigms, database searching and sophisticated information retrieval tools.

Open Standards are broadly defined as standards "independent of any single institution or manufacturer, and to which users may propose amendments" (Pountain, 2003). Academic libraries recognize Open Standards as a vehicle to ensure long term preservation of digital content and interoperability between systems, to solve obsolescence problems caused by advances in computer hardware and consequent changes in specifications, storage formats and access mechanisms. In particular, academic libraries engage actively with the Open Archives Institute (OAI) Protocol for Metadata Harvesting (OAI-PMH). The latter inaugurates an application-independent interoperability framework based on metadata harvesting. In addition, academic libraries do utilize various open standards for information retrieval such as Open-URL and Dublin Core Metadata Initiative (Corrado, 2005).

5. Academic libraries and Big Data

Modern research methods and sampling techniques generate large data sets. Data driven projects necessitate Research Data Management (RDM) strategy during all stages of the project – initial planning, collection of data from different sources, identification and labeling of data sets, processing these data sets and preservation and sharing of the results and raw data with the research community (Cox & Pinfield, 2014). It is important to mention that RDM is a complex activity, which incorporates data curation. The latter is an important part of archiving and preserving data for re-use. Academic librarians are investing resources to curate research data, especially Big Data (Akers, 2014). The latter are characterized by 3 Vs – high volume, high velocity, and high variety. In the context of the research cycle, Big Data is also characterized by its veracity. The latter refers to the quality of Big Data, understood in terms of accuracy (reliable methods of data acquisition), completeness of data (are there duplicates or missing data), consistency (are measurements and unit conversions accurate), uncertainty about its sources, and model approximations (Lukoianova & Rubin, 2014).

It is worth mentioning that Big Data typically needs application of a set of specific data cleaning, data validating and analytical tools in order to become valuable. From that perspective, the support for open data demands that librarians expand their qualifications toward data science and accept more data-centric roles (Hoy, 2014; Federer, 2016) in order to provide research data services. These include, but are not limited to, data sharing, data reuse, data collection, data visualization, data preservation and data curation. The technical side of the process requires the ability to interface with stored large data sets, and the technical ability to predispose data in a format suitable for decision making (Affelt, 2015). Within the scope of OD the volumes of data shared increase dramatically, which will transform libraries into Big Data libraries. However, the academic and public libraries with large collections are well positioned to make the transition toward OD and Open Big Data (OBD) libraries by cross sharing their collections and using Big Data Analytics (BDA) to identify similarities and consequently combine resources effectively. Such an approach is possible through the use of advanced data sharing cloud based technologies, when resources become virtual resources in a cloud (public or private) and thus become fully open for the research community. In addition BDA requires application of high performance algorithms and specific interfaces in order to make knowledge extraction effective. On the library side, the reference interview skills aiming to understand customer needs should be enriched with BD knowledge so librarians can offer OBD solutions as well (Hoy, 2014). Librarians' roles should expand to cover at least part of data scientist duties to curate, clean, remove duplicates, and maintain data. That includes complex metadata related activities, content development, classification and other activities in more complex research domains defined by cross disciplinary and interdisciplinary research. In these complex domains, however, the existing expertise accumulated for well-defined social science data repositories, bioinformatics repositories, and georeference data repositories is of little help because new multimodal strategies take place and because the cross disciplinary and multidisciplinary research generates large data volumes of different types of data. Variability and volume of data shifts the focus toward effective data management, security, protecting privacy of individual researchers, preserving sensitive data according to federal or government regulations, anonymization and many others. With the current lack of standard data quality assessment protocol for OD and OBD, libraries are facing the challenges of ensuring the quality of resources in terms of accessibility, reusability, and trustworthiness. In this regard, academic libraries should take a leading role in developing (with IT collaboration) extensive metadata policies and implementation methodologies that enable the use of the same data by different researchers or groups. Nevertheless, the variability and volume of OD and OBD require academic libraries to take steps in expanding, tuning and justifying roles of librarians in realm of OD and OBD by requesting basic data science competencies in addition to the traditional library science expertise and qualifications.

5.1. Research data management

In 2011, the U.S. National Science Foundation, one of the main research funding agencies for the country (along with National Institutes of Health – NIH), implemented a game changing rule requiring all grant proposals to include a data management plan (DMP). Consequently, academic libraries began to offer support for research data management by offering repositories, DMP services, training and education to campus researchers (Fearon et al., 2013). Most researchers are willing to share data, but are either unaware about specific steps and models they should follow or their funding is below the cutoff ($500,000 USD for NSF) required for having a data management plan. That caused academic libraries to shift their focus from formal DMP services to more integral services by providing instruction on best practices in data management and facilitating the use of disciplinary data repositories (in addition to institutional repositories). In addition, libraries have taken on the role to work with researchers to adequately describe their datasets prior to submission to data repositories (Akers, 2013). Support for the whole spectra of RDM however differs among institutions depending on their specific needs (Kouper et al., 2013), and their strategy to develop workforce equipped to support the needs of researchers (Soehner et al., 2010). Different approaches give the different presence of the libraries in development of the research management services. In some instances the leadership role in establishment of research data management was given to academic libraries and the supporting team was composed of IT specialists. For example, Cornell University formed a team of multi-specialists led by the library, which included library and management experts, as well as researchers from Cornell Advanced Computing and Cornell Institute for Social and Economic Research (Akers et al., 2014). At Johns Hopkins University, the librarians developed the whole data infrastructure and launched the JHU Data Management Services (Shen & Varvel, 2013). Another practice is to establish a separate body especially designed to address "challenges in data organization, description, dissemination, and discovery" (Akers et al., 2014). That approach was followed by Purdue University where the Distributed Data Curation Center was established under the leadership of the Dean of Libraries (Akers et al., 2014). At some institutions the library is the first to recognize the importance of RDM and consequently to create the administrative structure to support it. For example, the library at the University of Illinois at Urbana-Champaign created a new leadership position (Director of Data Services) and several library positions specifically designed to provide data support, including a life sciences data services librarian and an engineering data services librarian. This new personnel was heavily involved in RDM initiatives, although more recently a digital humanities librarian has also taken an active role in supporting the production and preservation of research data (Akers et al., 2014). A similar approach was taken by the University of Michigan where the library issued a report detailing the extent of its role and taking on shared responsibilities by forming task force groups (Akers et al., 2014).

5.2. Cornerstone projects

Other aspects of OBD are technological, institutional, managerial, and financial. From a technological perspective, academic libraries are not only nominal end-users of technological infrastructure, but also a generator of ideas aimed at ensuring interoperability and quality control of the data. For example, the Purdue University Research Repository and Distributed Data Curation Center launched DMPTool, embedded into PURR, a scientific collaboration tool built on the HUBZero platform. PURR facilitates scientific collaboration, publishes and assigns DOIs to datasets, and provides Ask-a-Librarian chat or e-mail "data reference" services. Liaison librarians work closely with the repository manager to incorporate the use of repository services during data management planning (Akers et al., 2014). In a similar fashion, libraries are involved in projects for development and implementation of metadata typically in collaboration with metadata specialists. For example, at Purdue University the subject librarians get support from data and metadata group which include a metadata specialist and several data specialists. The metadata group is under library leadership and reports to three library divisions (Akers et al., 2014).

In order to ensure flexibility, especially research involving interdisciplinary projects, the specialists from academic libraries quite often consult/collaborate with IT professionals in the implementation of data sharing resources for the University. The library at Purdue University launched, hosts, and supports Databib, an online annotated bibliography of data repositories useful for data producers and users, librarians, funding agencies, and publishers (Akers et al., 2014). Such an approach is justified by the need of the academic library to be in sync with technological advances and to convey specific national, social, regulatory, educational, institutional, and other regulations and requirements back to IT teams. However, it requires expansion of the qualifications of the librarians beyond library science and subject degrees toward statistics, data science and information technologies. As Tammaro and Casarosa (2014) outlined, librarians must have knowledge and understanding as to how the curation of digital resources differs from that of traditional materials and how to manage them. They must understand subject vocabulary, concepts and culture and to be able to understand the possible implications (in technical, institutional, economic and legal terms) of assuming the responsibility for long-term digital curation, providing guidance and support. Lastly, librarians need to master the necessary practical skills, e.g., to be able to manage projects and organize digital collections in order to guarantee that digital materials remain accessible and usable between domains (Tammaro & Casarosa, 2014). On an institutional level the role of the academic libraries enriches with active support of a variety of institutional models and consequently with development of set of tailored data sharing approaches aimed towards specific needs of the research community. To meet these projects' goals the libraries need personnel with relevant qualifications. According to the Association of Research Libraries' Profile for 2010 (Potter et al., 2010), only 10% of the members of Association of Research Libraries (ARL), a nonprofit organization of research libraries and research institutions in the USA and Canada, express knowledge in data-curation or e-science. In order to close that gap in recent years some academic libraries assign specifically trained personnel with both MLS and Data Science qualifications to develop data management standards based on users' workflows that can be leveraged by all researcher communities, and thus to incorporate the newly developed applications with the existing resources to ensure interoperability with current systems, such as documents, data sets, and codes (Akers, 2013, 2014).

6. New and expanded roles of academic libraries in the Open Science framework

6.1. Educator

Academic libraries take a leading role by fostering the culture of openness across research communities. Despite the fact that historically their primary role is to provide resources that support the curriculum, academic librarians view instruction as a vital part to their professional identities (Medaille, 2011). The nature of instructional work covers the whole spectrum – from semester-long courses for credit, one-shot information literacy classes, one-on-one instruction, research consultation, or featured resources presentations (Grigas et al., 2016). Being strong advocates for Open Access, academic libraries contribute to releasing the research potential of digital technologies and communications by promoting openness and providing easy and universal interfaces to data sets in the institutional repositories, in addition to their primary role. In the classroom OA morphs into Open Educational Resources

(OER), which is a new paradigm. OER are not only a pedagogical shift, but also a new leadership opportunity for academic libraries (Jensen & West, 2015). According to the Hewlett Foundation, "Open Educational Resources are teaching, learning, and research materials in any medium – digital or otherwise – that reside in the public domain or have been released under an intellectual property license that permits no-cost access, use adaptation and redistribution by others with no or limited restriction". OER typically refers to educational resources released under Creative Commons license and are typically available in electronic format. Open Educational Resources (OER) came as a continuum Massive Online Open Courses (MOOC) and Open Course Ware (OCW) launched by MIT in 2001, but quickly gained momentum and went ways beyond. Academic librarians are using their skills and copyright expertise to support, promote and create OER. Okamoto (2013) broadly categorizes academic libraries' activities as "advocacy, promotion, discovery, evaluation, collection, preservation, and access; curation and facilitation; and funding." The City University of New York (CUNY) embraced the OER initiative in 2010 when six community colleges were the first to establish "zero textbook" courses. Later a $4 million grant from the State of New York followed and in 2017 CUNY Libraries at all 24 campuses joined the OER movement and started converting educational resources used in high enrollment courses to OER materials. The short-term goal was to reduce costs for students and accelerate their progress through college; while the long-term goal is to create "Zero Textbook Cost" degree programs (CUNY Libraries Open Educational Resources, 2018). If proven successful the groundbreaking impact would be changing the culture and creating structures that connects curriculum and pedagogy in order to improve student learning outcomes.

6.2. *Enabler*

Academic libraries are traditionally a central body engaged in dissemination of knowledge and information for the academic community. Libraries go beyond formal education and play a role in sustaining science literacy by providing access to information in a variety of formats. Under the framework of Open Data dissemination, libraries have to support diverse and geographically disjoint communities. With 24 campuses across all five borrows in New York CUNY (City University of New York) is the perfect example. Their open access institutional repository Academic Works, coordinated by the Office of Library Services, is dedicated to collecting and providing access to the research, scholarship and creative work of the CUNY community. Hosted on a secure server and given a persistent URL (to ensure long-term access), Academic Works provides for the preservation and dissemination of a full range of faculty scholarship (CUNY Academic Works, n.d.). The central challenge becomes re-use of reliable data which academic libraries solve by enabling and supporting research data curation along with adopting new digital technologies capable to work on Open Big Data, expanding their digital services and tuning their educational programs to be sound to Open Science enabled type of research. Indeed massive data sets are already aggregated in research portals, for example, GBIf.org or IDigBio.org. Combined with increased diversity of research teams (Adams, 2012), the massive web accessible repositories oblige different level of coordination among academic libraries and data collectors (Knapp et al., 2013). Increase in diversity in research teams along with cross-disciplinary projects generate a wide range of types and formats of research data; from traditional numerical data, through unstructured data (e.g. videos) or spreadsheets or even handwritten notes and lab reports (Sumbal et al., 2017). Academic libraries meet the challenge of providing access to their institutional repositories holding many sets of data in various formats by supporting various interfaces, developing and supporting metadata systems for various data streams, supporting all ecosystems of software and data products for reformatting and reusing large and complex data, and support for data sets identifiers including specific ones (e.g., NCBI identifiers). Metadata, defined as data about data, provide a context for further re-use of the data (White et al., 2013), but the type of metadata depends on a file type. For instance simple text files may contain metadata in JSON or XML format, while some other formats such as NetCDF or HDF5 are self-documenting, and many video formats contain their own embedded metadata. Supporting a wide variety of metadata in different formats along with support of sophisticated software for data transformation permit libraries to link relevant data, therefore to address the problem of interoperability when variable names are assigned to data standards (defining how to describe different types of data for the domain, e.g., observations, specimens, samples, etc.) due to use of the same or subsets of the same data from different teams. Indeed, the data standards itself are domain specific (e.g., Darwin Core Standard for Biology or Climate Forecasting Convention) but one and the same data can and should be used in different context in different projects.

6.3. *Promoter of research diversity*

Some open data do exist already in grey literature, social media, blogs, social networking sites (e.g. ResearchGate), but those data are rarely research-curated or validated and thus they are not suitable for re-use in scientific research. By taking on the function of research data curation (along with open access), academic libraries guarantee the reliability of open diverse data sets for the community and consequently establish the conditions to contest citation bias and publication bias common in scientific research. As several empirical studies have shown (Czarnitzki et al., 2015), publication bias stems from editors' selection of the works to be published based on criteria not always driven by research quality, from researchers' willingness to pick up topics based on the political conjuncture set up by journal editors, and from researchers' willingness to publish only selected parts of their research. All those biases could limit the scope and directions of further scientific research. Similarly, citation bias, resulting from researchers' willingness to publish only in subscription journals with high impact factor, or researchers' willingness to abandon parts of their research which they don't believe to be "highly citable", affects and limits the choice of next research topic(s) as well. As a result of citation and publication bias a significant part of the data remains hidden, in this way defeating the purposefulness of data as a generator of research ideas, research topics or implementations.

By building and maintaining open access institutional repositories, academic libraries can host the entire (published and unpublished) research output of their home institutions including both scientific publications and curated research data. Those data that haven't been included in published works can be re-used for verification and reproducibility purposes. Data curation "involves maintaining, preserving and adding value to digital research data throughout its lifecycle" (Digital Curation Centre, 2014) and thus goes a step further than digital preservation which ensures long-term integrity only but not accessibility for immediate scientific re-use. The challenge there is not the technological capacity of such repositories but creating adequate metadata and policy ensuring sustained access to those curated data. For CUNY Academic Works, such submission policy and practices are already established – all current faculty, students and staff can submit their work to the repository (CUNY Academic Works, n.d.). Included works are selected and deposited at individual campuses where library coordinators provide consultation on a need to know basis in accordance with the submission policies, pending approval by the Scholarly Communications Librarian at the Office of Library Services who is the gatekeeper of the repository.

6.4. Resources assessment

Under OD, academic libraries take on a new role to evaluate and grade the data storage resources, both private and public. In 2010 only life sciences repositories numbered more than 1,000 (Marcial & Hemminger, 2010). Currently, there are more than 2,000 open research data repositories of different types and with many different policies (Kindling et al., 2017). The journal *Nature* (2016) recently published a list of recommended repositories as "Repositories included on this page have been evaluated to ensure that they meet our requirements for data access, preservation and stability". In Europe, RE3DATA (re3data.org), which is the most comprehensive source of reference for research data repositories, allows for the searching for research repositories utilizing over 40 criteria, such as subject, domain, content type, country, etc. (Pampel et al., 2013). It is very important to mention that research data management is discipline specific and thus the most suitable data repository should be carefully selected based on a set of criteria, which cover the whole process of data archiving and reuse in the discipline, and also within a framework of local and national legislative constraints and limitations defined by funding agencies. For example the answers of the following questions (Hart et al., 2016) might help to provide an integral understanding of suitability of a data repository for the particular type of research activity: Does the facility provide data curation or not? What type of data are accepted (structured or/and unstructured)? Are data backed up and if so how often? What is the data retention policy? And so on. The criteria for evaluating data repositories are usually developed by academic libraries with collaboration from IT departments. At City University of New York (CUNY) in addition to their institutional repository Academic Works, the CUNY High Performance Computing Center (HPCC) provides storage space which holds long term data for different research projects across the University (CUNY HPCC, n.d.). These data are annotated by the researchers (metadata) and are backed up on centralized data silo once every 24 hours. Currently the project data repository holds about 600 Tera Bytes research data from several NSF and NIH funded projects. Currently CUNY is building up research data cloud with capacity to store all types of research data deposited from different research groups across the university.

6.5. Mediator and advisor

The abundance of research data is a desired output from Open Science, but it also causes significant changes in the types and scope of research projects; in addition, data variability promotes the proliferation of cross-disciplinary and multidisciplinary projects. At research universities where those types of projects are predominant, the librarians have the opportunity to take on a new role as mediator of agreements between research groups and individuals both inside and outside of the institution. That includes preparing documentation about formal roles of institutional bodies (library, departments, centers, etc.), publication agreements and – in terms of modes of data sharing – defining the additional specific policies for open but sensitive data restricted under federal or local legislation. Due to their practical nature, the output of such projects is typically welcomed by the industry, since it decreases the investment for applied research and development made by manufacturers. For the academic libraries, however, the partnership of the research institution with the industry (grants, centers of excellence, tech incubators, etc.) aimed to support OD model not only ensures easy access to state of the art hardware, and advanced data management technologies, but also stimulates the adjustment of library services to take advantage of these technologies.

Open Science research not only uses OD, but also generates large amounts of "new" data which have to be stored, curated and made public. Storage of the "new" data demands specific data center capabilities and brings the new role to the academic librarians in large research universities – as research advisors. Academic libraries are well placed to take on such a role, because librarians can offer their expertise about funding opportunities and guidance during the exploratory stage of research (Howse et al., 2006). Consequently librarians may advise patrons about OD processing – from most suitable data storage through sharing and re-use. In this new role the academic library personnel relies on a combination of traditional library science expertise, data science knowledge and IT proficiency.

7. Conclusion

In this paper we discussed the changes in services and development of new roles in academic libraries inspired by Open Science and especially changes motivated by Open Data and Open Big Data. The role of Open Data as generator of new research increased dramatically in the last decade so providing integral support for Open Data driven explorations becomes a central challenge for academic libraries. In order to support open data driven research, academic libraries re-invent themselves by launching expansions of traditional library services, adoption of new data science roles and expanding the library's educational and mediator functions. These processes have led to deep transformation in libraries themselves making them more technologically savvy, data oriented and active participants in the research process. Indeed academic institutions choose different approaches to ensure the support for Open Science, but in all instances the academic librarians are entitled to play a central role by providing leadership, information services, research data management services and even collaborating in research projects in their institutions. By sharing examples from CUNY we illustrate how librarians can help to incorporate the Open Science concept in teaching, learning and research processes at the University. This has ramifications for those who are training the next generation of academic librarians (i.e., graduate library and information science programs), as well as for those who periodically predict the disappearance of libraries – perhaps libraries and librarians will not only survive but thrive by adapting to and taking on the opportunities that arise as a result of the new roles that come along.

References

Adams J. (2012). Collaborations: The rise of research networks. *Nature*, *490*, 335-336.

Affelt, A. (2015). *The accidental data scientist: big data applications and opportunities for librarians and information professionals*. Information Today.

Akers, K.G. (2014). Going beyond data management planning: Comprehensive research data services. *College & Research Libraries News*, *75*(8), 435-436.

Akers, K.G., Sferdean, F.C., Nicholls, N.H., & Green, J.A. (2014). Building support for research data management: Biographies of eight research universities. *International Journal of Digital Curation*, *9*(2), 171-191.

Akers, K.G. (2013). Looking out for the little guy: Small data curation. *Bulletin of the American Society for Information Science and Technology*, *39*(3), 58-59.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.

Bernard, C. (1957). *An introduction to the study of experimental medicine*. Dover Publications.

Boulton, G. (2016). Reproducibility: International accord on open data. *Nature*, *530*, 281.

Chen, X., Dallmeier-Tiessen, S., Dasler, R., Feger, S., Fokianos, P., Gonzalez, J.B., & Rodriguez, D.R. (2019). Open is not enough. *Nature Physics*, *15*, 113-119.

Corrado, E.M. (2005). The importance of open access, open source, and open standards for libraries. *Issues in Science and Technology Librarianship*, (42).

Cox, A.M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4), 299-316.

Craig, R.T. (1999). Communication theory as a field. *Communication Theory*, *9*(2), 119-161.

CUNY Libraries. 2018. Open Educational Resources: New York State open educational resources funds CUNY year one report. Retrieved https://www2.cuny.edu/libraries/open-educational-resources/.

CUNY Academic Works. n.d. Retrieved from https://academicworks.cuny.edu/.

CUNY High Performance Computing Center. n.d. Retrieved from https://www.csi.cuny.edu/academics-and-research/research-centers/cuny-high-performance-computing-center.

Czarnitzki, D., Grimpe, C., & Pellens, M. (2015). Access to research inputs: Open science versus the entrepreneurial university. *Journal of Technology Transfer*, *40*(6), 1050-1063.

Dain, P. (1990). Scholarship, higher education, and libraries in the United States: Historical questions and quests. *Libraries and Scholarly Communication in the United States: The historical dimension*. Greenwood Press.

David, A. (2008). The historical origins of "Open Science": An essay on patronage, reputation and common agency contracting in the scientific revolution. *Capitalism and Society*, *3*(2), 1-103.

Digital Curation Centre. (2014). What is digital curation? Retrieved from https://escholarship.org/uc/item/32q2z1c9.

European Commission. (2011). Horizon 2020 – The Framework Programme for Research and Innovation.

Fearon, D., Gunia, B., Sherry, L., Pralle, B.E., & Sallans, A.L. (2013). ARL SPEC Kit 334: Research data management services (July 2013).

Fecher, B., & Friesike, S. (2014). Open Science: One term, five schools of thought. In *Openning Science*. Springer, Cham.

Federer, L. (2016). Research data management in the age of big data: Roles and opportunities for librarians. *Information Services & Use*, *36*(1-2), 35-43.

FOSTER. (2017). What is Open Science? Introduction. Retrieved from https://www.fosteropenscience.eu/content/what-open-science-introduction.

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature*, *529*(7584), 117-119.

Gitelman, L. (2013). *Raw data is an oxymoron*. MIT press.

Grand, A., Wilkinson, C., Bultitude, K., & Winfield, A.F.T. (2012). Open Science: A new "trust technology"? *Science Communication*, *34*(5), 679-689.

Grigas, V., Fedosejevaitė, R., & Mierzecka, A. (2016, October). Librarians as educators: Affective dimensions experienced in teaching. In *European Conference on Information Literacy*, Springer, Cham, pp. 619-633.

Quentin, G., Weatherdon, L., & Geijzendorffer, I.R. (2017). Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology*, *54*(2), 612-617.

Hart, E.M., Barmby, P., LeBauer, D., Michonneau, F., Mount, S., Mulrooney, P., & Hollister, J.W. (2016). Ten simple rules for digital data storage. *PLoS Computational Biology*, *12*(10), e1005097.

Heidorn, P.B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, *51*(7-8), 662-672.

Hesse, B.W. (2018). Can psychology walk the walk of open science? *American Psychologist*, *73*(2), 126-137.

Howse, D.K., Bracke, P.J., & Keim, S.M. (2006). Technology mediator: A new role for the reference librarian? *Biomedical Digital Libraries*, *3*(1), 10.

Hoy, M.B. (2014). Big data: An introduction for librarians. *Medical Reference Services Quarterly*, *33*(3), 320-326.

Hunter, L., & Leahey, E. (2008). Collaborative research in sociology: Trends and contributing factors. *The American Sociologist*, *39*(4), 290-306.

Jensen, K., & West, Q. (2015). Open educational resources and the higher education environment: A leadership opportunity for libraries. *College & Research Libraries News*, *76*(4), 215-218.

Kindling, M., van de Sandt, S., Rucknagel, J., & Schirmbacher, P. (2017). The landscape of research data repositories in 2015: A re3data analysis. *D-Lib Magazine*, *23*(3/4).

Knapp, A.K., Collins, S.L., Turkington, R., Long, R., White, S., Cahill, J.F., & Lind, E. (2013). Coordinated distributed experiments: An emerging tool for testing global hypotheses in ecology and environmental science. *Frontiers in Ecology and the Environment*, *11*(3), 147-155.

Kraker, P., Leony, D., Reinhardt, W., & Beham, G. (2011). The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning*, *3*(6), 643-654.

Kouper, I., Akers, K.G., Nicholls, N.H., & Sferdean, F.C. (2013, July). A roadmap for data services. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, ACM, pp. 375-376.

Kulczycki, E. (2016). Rethinking Open Science: The role of communication. *Analele Universităţii din Craiova. Seria Filosofie*, *37*(1), 82-97.

Lukoianova, T., & Rubin, V.L. (2014). Veracity roadmap: Is Big Data objective, truthful and credible? *Advances in Classification Research Online*, *24*(1), 4-15.

Marcial, L.H., & Hemminger, B.M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology*, *61*(10), 2029-2048.

Max-Planck-Gesellschaft. (2003). Berlin Declaration to Open Access to Knowledge in the Sciences and Humanities. Retrieved from https://openaccess.mpg.de/Berlin-Declaration.

McNutt, M., Lehnert, K., Hanson, B., Nosek, B.A., Ellison, A.M., & King, J.L. (2016). Liberating field science samples and data. *Science*, *351*(6277), 1024-1026.

Medaille, A. (2011). Librarians view instruction as integral to their professional identities. *Evidence Based Library and Information Practice*, *6*(4), 120-123.

Molloy, J.C. (2011). The open knowledge foundation: Open data means better science. *PLoS Biol*, *9*(12), e1001195.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L.O.B., & Wilkinson, M.D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the european open science cloud. *Information Services & Use*, *37*(1), 49-56.

Nature. (2016). Recommended Data Repositories. Retrieved from https://www.nature.com/sdata/policies/repositories.

Office of Science and Technology Policy. (2013). Memorandum for the Heads of Executive Departments and Agencies. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Okamoto, K. (2013). Making higher education more affordable, one course reading at a time: Academic libraries as key advocates for open access textbooks and educational resources. *Public Services Quarterly*, *9*(4), 267-283.

Open Society Foundations. Retrieved from https://www.opensocietyfoundations.org/explainers/what-open-access.

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., & Dierolf, U. (2013). Making research data repositories visible: The re3data. org registry. *PloS One*, *8*(11), e78080.

Potter, W.G., Cook, C., & Kyrillidou, M. (2010). ARL profiles: Qualitative descriptions of research libraries in the early 21st century. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, (271), 25-32.

Pountain, D. (2003). *The Penguin Dictionary of Computing*. New York: Penguin Putnam.

Roe, D.M., & Moody, D. (1999). *The librarian as mediator: a significant change in the educational role of librarians*. Association of College and Research Libraries.

Saez, R.C.M.-F. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428-436.

Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., & Ungar, L.H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One*, *8*(9), e73791.

Shen, Y., & Varvel, V.E. (2013). Developing data management services at the Johns Hopkins University. *The Journal of Academic Librarianship*, *39*(6), 552-557.

Soehner, C., Steeves, C., & Ward, J. (2010). *E-Science and Data Support Services: A Study of ARL Member Institutions*. Association of Research Libraries, Washington, DC.

Stodden, V., McNutt, M., Bailey, D.H., Deelman, E., Gil, Y., Hanson, B., & Taufer, M. (2016). Enhancing reproducibility for computational methods. *Science*, *354*(6317), 1240-1241.

Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., & Astell, M. (2018). Practical challenges for researchers in data sharing. *Springer Nature. https://doi.org/10.6084/m9. figsh are*, *59750*(11), v1.

Sumbal, M.S., Tsui, E., & See-to, E.W. (2017). Interrelationship between big data and knowledge management: An exploratory study in the oil and gas sector. *Journal of Knowledge Management*, *21*(1), 180-196.

Tammaro, A.M., & Casarosa, V. (2014). Research data management in the curriculum: An interdisciplinary approach. *Procedia Computer Science*, *38*, 138-142.

Tennant, J.P., Waldner, F., Jacques, D.C., Masuzzo, P., Collister, L.B., & Hartgerink, C.H. (2016). The academic, economic and societal impacts of Open Access: An evidence-based review. *F1000Res*, *5*, 632.

Van Tuyl, S., & Whitmire, A.L. (2016). Water, water, everywhere: Defining and assessing data sharing in academia. *PLoS One*, *11*(2): e0147942.

Wallis, J.C., Rolando, E., & Borgman, C.L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS One*, *8*(7), e67332.

White, E.P., Baldridge, E., Brym, Z.T., Locey, K.J., McGlinn, D.J., & Supp, S.R. (2013). Nine simple ways to make it easier to (re) use your data. *Ideas in Ecology and Evolution*, *6*(2), 1-10.

Zuccala, A. (2010). Open access and civic scientific information literacy. *Information Research: An International Electronic Journal*, *15*(1), 1-27.