Qualitative and Quantitative Methods in Libraries (QQML) 8,3: 337-343 2019

Open Research Data in Biomedicine: A Step Forward in Defining an Italian National Policy

Moreno Curti[1], Paola De Castro[2], Rosalia Ferrara[3], Pietro La Placa[4], Cristina Mancini[3], Elisabetta Poltronieri[2], Franco Toni[3], Angela Vullo[5]

(BISA, Bibliosan for Open Science Working Group)
[1] IRCCS Fondazione Policlinico San Matteo, Pavia (Italy), Scientific Direction
[2] Istituto Superiore di Sanità, Rome (Italy), Scientific Knowledge and Communication Service/Publishing Unit
[3] Istituto Superiore di Sanità, Rome (Italy), Scientific Knowledge and Communication Service/Library
[4] Istituto Zooprofilattico Sperimentale della Sicilia, Palermo (Italy), Editorial Service and Library
[5] Istituto Zooprofilattico Sperimentale della Sicilia, Palermo (Italy), Local Environmental Health Unit

Abstract: The concept of openness to research output has been strengthened by the current EU funding research programme Horizon 2020. In early 2017 the BISA (Bibliosan for Open Science) inter-institutional Working Group, including scientific information experts mainly from Istituto Superiore di Sanità (ISS, Italian National Institute of Health), carried out a survey on the practical handling of research data produced by Bibliosan affiliated institutions, the Italian Biomedical Research Libraries Network promoted by the Italian Ministry of Health. The BISA online questionnaire collected over 2,400 responses obtained from 58 Bibliosan institutions. Among the various aspects of data management covered by the BISA questionnaire, that one referring to the need of a policy and guidelines, to be adopted by Bibliosan research institutions, for dealing with research data was deeply investigated.

Keywords: Biomedical research; Open data; Information dissemination; Access information; Italy

1. Introduction

Open science paradigm is gaining increasing impact among stakeholders of scientific communication system by fostering visibility, free access and sharing of research results (publications and data) in the biomedical field. Acquiring knowledge of the evolving landscape of research data is therefore fundamental in order to plan their optimal managing for the progress of science. According to the UK Engineered and Physical Sciences Research Council (EPSRC), "research data is defined as recorded factual material commonly retained by and

---

Received: 1.5.2019 Accepted: 13.9.2019 ISSN 2241-1925

accepted in the scientific community as necessary to validate research findings" (https://epsrc.ukri.org/about/standards/researchdata/scope/).

BISA Working Group was set up in 2016 by information professionals affiliated with Bibliosan institutions (the Italian network of health libraries) and coordinated by Istituto Superiore di Sanità (ISS, Italian National Institute of Health). Its members are playing the role of facilitators of a bottom-up initiative, in close collaboration with researchers to develop mutually satisfactory concepts and reach a final agreement with all stakeholders involved. The Working Group was committed from the beginning to fulfilling two main goals: firstly, that of creating a survey addressed to Bibliosan network to reveal the habits of researchers in managing their primary data, and secondly that of outlining a shared policy able to regulated research data management.

2. BISA Survey Results

BISA survey was intended to measuring the level of openness of Bibliosan research institutions by exploring data practices and expectations of scientists about data handling, to further adopting sustainable solutions for data sharing. Detailed results and comments originated from the survey were presented in a one-day conference entitled *Open data as science bricks: results of the survey by Bibliosan per la scienza aperta – BISA*, organized by ISS in May 2017 (http://www.bibliosan.it/ftp/bisa_atti_15052017/bisa_15_05_2017.html). A brief report of the event was timely published by BISA Working Group (2017a) on the ISS monthly newsletter, while conference proceedings alongside contributions from advocates of open science principles within the umbrella initiatives of the European Union, were then collected in an *ad-hoc* report edited by BISA Working Group (2017b) and published by ISS.

This BISA initiative in its early stage was inspired from a report by Bauer (2015) relating to a survey carried out in 2015 within the *Project E-infrastructures among Austrian academies*. The questionnaire on which the Austrian report is based was translated into Italian by the BISA Working Group in 2016 and then adapted to the context of Bibliosan institutions. In early 2017, BISA online questionnaire was sent to the 60 institutions affiliated with Bibliosan network. It gathered a 2,433 responses from 58 institutions, reaching a target population of about 15,000 scientists, thus gaining a survey response rate of 15,5%.

In BISA survey the main focus was given to the following aspects of handling data: data type and format, access, storage, re-use, licensing, attitude to sharing, preservation, data services and infrastructures.

As tools used for saving data, the variety of responses reveal that there are no set procedures for a given institutional *ad hoc* device as a memory location; surprisingly, the majority of respondents (33% to a multiple-choice question) declared they save data on their own computer.

As roles and responsibilities for preserving research data, 52% of respondents answered they are the only responsible for data archiving, namely the prevalence is for single researchers rather than for project or group managers. Almost the half of respondents sometimes handle/generate sensitive data. The issue of confidential research data affects accessibility, sharing and re-use of data and researchers should be able to rely on guidelines, policies from legally educated personnel or from a supervisor to avoid legal consequences. As far as which audience should be given access to research data, three quarters of respondents indicated they grant access only to interested persons upon request or to people affiliated with their own institution. So, access to data seems to be not so open, but rather restrictive or, at least, selective.

Another question referred to where data are made available, as the premise is that research data are valuable material not only to those who produce them, but also to all who re-use them. Here, the most frequently selected options (42%), e-mail and pen drive, lead to think that researchers rely on personal and non-institutional devices to circulate their data. In facts, data repositories or cloud applications, as certified platform to deposit data, are still less common. As reasons declared by respondents, the most significant obstacles to the sharing of research data appeared to be the fear of misinterpretation and misuse of data (34%), the fear of privacy violation (26%) and the risk of commercial exploitation of data by third parties (16%).

In conclusion, BISA survey's respondents revealed the expectation of being involved in an institutional process of re-organising research data workflow as far as storage, preservation and dissemination of research data. This implies active coworking among all institutional research supporting components as scientific information professionals, IT personnel, statisticians, administrative and legal staff, in order to offer services for an effective internal managing of data. Criticalities are still represented by lack of organizational measures expected from researchers of the institutions surveyed, in terms of roles and responsibilities to be defined for managing data. In this regard, the implementation of structural and strategic measures should be based primarily on guidelines and policies as well as on trained qualified staff for dealing with research data. Concrete actions from the top managers of single Bibliosan institutions are then expected under the synergy of governmental authorities as the Italian Ministries of Health and Education. In this respect, ISS member of BISA are playing a leading role to foster the definition of a draft policy to be circulated for feedback from Bibliosan stakeholders (Scientific Directorates and researchers' representatives). The challenge is, definitely, to create an organizational culture favouring effective regulation of research data at national level.

BISA survey results were illustrated during some events addressed to the scientific community at international and level, in order to raise the debate on crucial aspects of data management and regulation. A conference presentation by BISA Working Group (2017c) was given in June 2017 at the Annual meeting of the European Association for Health Information and Libraries (EAHIL) traditionally involved in discussing innovative issues faced by information professionals in the biomedical field. It is also worth mentioning the participation of some BISA members (2017d) in the annual conference of the Italian Research and Education Network (GARR), the organism who brings together users, technical operators and managers of the Italian public and private research institutions and academies.

3. Development of a Model Policy for Research Data Management at ISS

As a further action in view of defining a policy to be adopted by Bibliosan research institutions, BISA Working Group, coordinated by BISA ISS members, is currently circulating, for internal feedback, a draft roadmap concerning research data definition and typology, management, sharing modalities, roles and responsibilities of data producers and administrators, data protection and technical aspects relating to infrastructures and services. The main goal is firstly raising awareness in ISS research community, in order to gather comments and different views on this delicate matter. Once carried out this preliminary round reflecting ISS perspective on the management of research data and according to the Italian Ministry of Health guidelines, the policy will be circulated among Bibliosan institutions to receive feedback from a large number of bodies involved in biomedical research as well as in health care. Along this agenda of initiatives, the BISA Working Group is keeping close contacts with a similar Study Group operating within the Conference of Italian University Vice-Chancellors (CRUI, Conferenza dei Rettori delle Università Italiane) the main Academic Authority grouping all Italian universities.

Scope and Benefits of a Data Repository

For the transparency of science to scholars and general public, data accessibility and sharing are strongly recommended or mandatory for certain types of data (except reasonable restrictions) according to:

- the general principle that publicly funded research should be freely available
- the terms of data availability set by the *Data Management Plans (DMP)* required by research funders, firstly the European Commission;
- the data access conditions required by scientific publishers for datasets associated to submitted articles reporting original research (eg. Writing a data availability statement http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf and ICMJE *Data sharing statement* envisaged for clinical trials results (by 01/07/2018) http://www.icmje.org/icmje-recommendations.pdf

In order to build up a certified online public platform to host and preserve the datasets generated or managed by ISS internal staff, a roadmap of research data was outlined with the aim of including all aspect and needs relating to a fruitful management of data. The outlining of such an infrastructure is undoubtedly an essential institutional tool for supporting the definition of a research data policy.

According to suggestions relating data handling experienced by ISS internal staff and acquired by BISA Working Group members through informal meetings in early 2018, a tentative index of resources and tools of an hypothetical data repository has been sketched as a basis for further implementation. The envisaged layout should be as follows:

Nature, type and structure of data

- Subject - specif data, relating to the multiple subject domains covered by ISS institutional mission (eg. environmental pollutants biomonitoring, epidemiological surveillance, food safety)
- Primary data (observed or collected directly from first-hand experience) *versus* aggregated data (grouped and summarised in tables)
- Internally or externally generated datasets
- Structured data available from registries and databases
- Sensitive data (in the light of EU Regulation 2016/679 of the European Parliament on the Protection of Personal Data https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=IT
- Publicly available data (eg. supplementary data provided alongside published papers)
- Not-publicly available data
- Data (published or not published) available on request from the authors (not publicly available due to privacy or third party restrictions)

Actions and services available in a data repository

- Institutional storage
- Data format definition
- Uniform and persistent data identification (eg. inclusion of accession codes or DOI, Digital Object Identifier)
- Metadata assignation
- Data search and retrieval functions
- Access to data (free access for publicly available data or restricted access set by project agreements, patents, patient’s consent in case of clinical data) and data set access authorization levels
- Provision of web links to publicly available datasets
- Statement of conditions for access to non-publicly available datasets
- Use of internal data (generated while performing institutional activities)
- Use of external data (with possible restrictions of use)

- Data long-term conservation
- Data sharing and re-use (whenever possible)

Online applications and tools (utilities)

- Support and guidance for developing a Data Management Plan (eg. *MANTRA Research data management training* https://mantra.edina.ac.uk/)
- Qualified online training (eg. through webinars)
- Help-desk services on request

4. Conclusions

The current scholarly communication system seems to have truly entered the era of data economy, where data is considered the new gold or the new oil…Actually, the actors of scientific publishing as research funding organisms, decision makers, scientists and publishers are deeply embedded in the methodology of research data collection, storage and preservation for free provision, sharing and re-use under the paradigm of open science. Within this scenery, professionals in searching and managing biomedical information are called once again to exert their expertise in supporting scientists in their daily publishing practises for evaluating scientific quality of publications. Thanks to this constant support, authors may gradually regain partial control of the publishing industry, as claimed by supporters of open science principles against a dominant logic of profit among science publishers. In such a context, BISA Working Group efforts are on the side of Italian researchers in the biomedical field, trying to help them with a tool kit of regulating policies, legal advice, training support and appropriate technology for optimal circulation of scientific literature (publications and data). In view of increasing transparency in scientific research, all stakeholders of global exchange of health information are challenged by the era of "data capitalism", as argued by Hagner (2018) in an interesting essay on effects (in terms of digital revolution and free flow of information) of the open access business model on the traditional scholarly publication system.

References

Gruppo di lavoro Bibliosan per la Scienza Aperta (BISA) (2017a). Convegno I dati aperti cemento della scienza: risultati dell'indagine Bibliosan per la Scienza Aperta (BISA). *Notiziario dell'Istituto Superiore di Sanità*, Vol. 30, No. 5, 19-21.

Gruppo di lavoro BISA (2017b). Risultati dell'indagine BISA (Bibliosan per la Scienza Aperta) sui dati aperti per la ricerca. Istituto Superiore di Sanità, Rome. *Rapporti ISTISAN* 17/32.

Bauer, B., Ferus, A., Gorraiz, J., Gründhammer, V., Gumpenberger, C., Maly, N., Mühlegger, J. M., Preza J. L., Sánchez Solís, B., Schmidt, N., and Steineder, C. (2015). *Researchers and their data. Results of an Austria survey. Report 2015*. University of Vienna, Vienna. Available at:

https://services.phaidra.univie.ac.at/api/object/o:409318/diss/Content/download. Last visited: 27/04/2018.

Poltronieri, E., Cognetti, G., La Placa, P., Vullo, A., De Castro, P., Ferrara, R., Mancini, C. and Toni, F. (2017c). Let your open data smoothly settle in your publishing habits. *12. International Congress on Medical Librarianship (ICML) and 2017 EAHIL Workshop. Diversity in Practice: Integrating, Inspiring and Innovative. Book of abstracts.* June 12-16, 2017; Dublin. p.20-21.

Curti, M., De Castro, P., Di Benedetto, C., Ferrara, R., La Placa, P., Mancini, C., Minghetti, L., Poltronieri, E., Santoro, F., Toni, F. and Vullo, A. (2017d): I dati della ricerca biomedica in Italia: verso la definizione di una policy nazionale? *Conferenza GARR 2017. The data way to science.* November 15-17, 2017; Venice. Available at: https://www.eventi.garr.it/it/conf17/programma/conferenza/blog-2/172-elisabetta-poltronieri. Last visited: 27/04/2018.

Hagner M. (2018). Open access, data capitalism and academic publishing. *Swiss Medical Weekly*, 148, w14600.