Training Data Stewards in Italy: Reflection on the FAIR RDM Summer School

Anna Maria Tammaro(✉) and Stefano Caselli

University of Parma, Parma, Italy
{annamaria.tammaro,stefano.caselli}@unipr.it

Abstract. "Fair Research Data Management" Summer School in Parma focused on the skills gap in Italy for data stewards. A distinct feature of the Summer School was its aim to bring together participants from different backgrounds and from different countries. The paper is a reflection on the organization of the Summer School and the evaluation received by the participants.

Keywords: Research data management · Data stewards · Open Science

1 Introduction: Research Data Management Skills Gap

1.1 A Profile for Data Stewards

An understanding of the role defined by "data stewards", and the formal titles with which they typically operate (such as data librarians, data scientists, curators of research data, etc.), can vary considerably over time, between different institutional contexts and across international borders.

Research data must be managed and preserved for the long term. The support service for the management of research data, however, is not a simple "restyling" of preservation for new types of dynamic documents, but implies a role of the data steward that extends to all phases of the research cycle, starting with design and implementation of the research project. The most important feature of "data curation" is that it represents a socio-technological system: new partnerships of librarians and support staff with teachers and researchers are needed along with the technological infrastructure and institutional organization. The necessary approach is the systemic one defined by Borgman (2015) as a socio-technological system. We must not start from the granularity of the collection, but from the needs of users (scholars and re-users) and from the political and organizational context of the institution (Corrall 2012).

The vision of the academic libraries' service was clearly designed by ARL: *"In 2033, the research library will have shifted from its role as a knowledge service provider within the university to become a collaborative partner within a rich and diverse learning and research ecosystem"* ARL (2015).

A definition of data librarian proposed by the Project of the Library Theory and Research Section of IFLA is the following (Tammaro et al. 2019): "*Data librarians are supporters of data re-users, in the process of production, storage and access of data"*.

M. Ceci et al. (Eds.): IRCDL 2020, CCIS 1177, pp. 163–172, 2020.
https://doi.org/10.1007/978-3-030-39905-4_16

The activities characterizing the role and cited by the interviewees from the IFLA LTR Project, in order of decreasing importance, are in the first place activities in close contact with researchers such as training (70%), orientation (66%), communication (61%), followed by access (58%) preservation (58%), policy preparation (54%), data management plan (51%), information systems (44%), statistics (21%), copyright (18%), knowledge of the subject (26%). Most IFLA Project respondents work in universities (78%) and academic libraries (10%), a minority work in research centers (6%), public offices (5%), public libraries (1%) and other organizations. The majority of data librarian positions are for qualified roles and require a Master (35%) or even a PhD (6%).

Data stewardship is defined as *"the process and attitudes that make one deal responsibly with one's own and other people data throughout and after the initial scientific creation and discovery cycle"* (Mons 2018). Initially, the idea was that a domain expert had the responsibility of qualifying and documenting the data from their professional point of view. In fact, Data stewards share some responsibilities with data curator and data librarian.

1.2 Research Data Management in Italy: The Skills Gap

The mandate of the European Commission for Open Science states: "As open as possible, as closed as necessary". In practice, the European Commission asks the professors and researchers involved in projects for which public funding is requested and obtained to deposit research results, including scientific publications, in open access repositories and to produce a Data Management Plan for the data.

In Italy, after Law No. 112/2013 which launched a first political initiative for Open Access to scientific publications, a bill (called "Gallo" from the name of the speaker) is now being discussed in Parliament, which includes the research data among its topics of interest, as the first legal basis of a national policy for the research data infrastructure, starting from the existing network of institutional repositories.

Even in Italy, some research data management pioneering experiences are born, many linked to the IOSSG group. The characteristic that distinguishes the service of these Italian universities has been defined as: "single point of entry". This model of open data support services, developed by the IOSSG Group, highlights the social role of the data librarian, who must work with the other University offices involved in the research cycle. However, adequate skills are lacking in Italy, and moreover new roles are needed for support services. Only two training experiences have been carried out at the Universities of Bologna and Padua, where the Library Systems have trained staff involved in support services.

Most of the professional literature believes that the librarian's background includes fundamental skills for the data librarian, but often does not describe them. During the Stelline Conferences in 2018 and 2019 two Seminars were held on the Data Librarian profile to understand the needs of Italian universities that pioneered the research data management service. The organizers of the two Seminars were DILL (International Master in Digital Library Learning) together with IOSSG, OpenAire and RDA. The good practices of some of the major Italian universities were illustrated during the seminar by Elena Giglia (University of Torino), Paola Galimberti (University of Milan), Elena Bianchi (University of Padua), Marialaura Vignocchi (University of Bologna).

The questions on which the organizers wanted to start the discussion were: What are the activities necessary for research data management? and what skills are consequently necessary for the training and updating of the data librarian? Before the first Seminar in 2018, a questionnaire was distributed on the skills of the data librarian. A conversation before and after the Seminar took place online, and helped to clarify the different opinions of librarians.

The questionnaire that preceded the Stelline Seminar allowed to highlight the opinions of 165 librarians who responded to the survey (Tammaro 2018). The competencies listed in the questionnaire had to be classified by the respondents as: fundamental (core identity of the librarian), specialized (emerging competencies of the data librarian), important (common to librarians in different roles) and not important. The objectives that had been set were:

- Understand the skills that require specialized training for the data librarian;
- Highlight the core skills that each librarian curriculum should propose.

The results of the questionnaire for the first macro-area of the support activities highlight as core skills in order of importance: cataloging (59%) including classifications and ontologies (28%), evaluation and selection of resources (28%) and preservation (28%). It is interesting to note that the management of research data (35%) is considered both a core competence and an emergent competence: librarians believe that the management of research data is part of the professional profile.

Technological skills are also considered important for all librarians. In particular, in order of importance we highlight: institutional repositories (35%), user experience (33%), interoperability (41%), semantic web (46%), information architecture (40%).

All management skills are considered important: research methods for user studies and service evaluation (50%), support for institutional services policies (49%), resource management (47%), legal aspects and ethical (42%), costs and business plans (39%). It is interesting to note that many librarians consider the business plan a specialist competence of the data librarian (44%). Even the organizational culture is a competence of common importance and considered fundamental: as an aptitude to work in a group (72%) and ability to network (68%), orientation to problem solving (68%), multiculturalism (55%).

Other skills that have been added by respondents concern aspects that had been neglected in the questionnaire such as: statistics, humanistic skills to understand the social value of research results, knowledge of international good practices, information literacy, empathy, ability to delegate and to ask questions if something is not known.

2 FAIR Research Data Management Summer School

The planning of the "Fair Research Data Management" Summer School in Parma began during a management meeting of the ROMOR Project in Brighton (16–18 January 2019) with an initial discussion of the learning objectives from the partner institutions. Since the Summer School would take place in the last year of the ROMOR project (http://romor.iugaza.edu.ps), the partners agreed on the dual objective of consolidating the knowledge of Open Access Institutional Repositories (OAIR) topics and developing the data stewards skills on more advanced topics. From the Stel-line

Seminars survey on learning needs and from documentary research, it was highlighted that currently no training offer is available in Italy on the management of research data. In agreement with the ROMOR Project coordinator, it was decided to open the Summer School also to external participants.

The purpose of this paper is to evaluate what has been done for the Italian participants and discuss how to improve the curriculum of the Summer School in the next experience.

2.1 Learning Objectives

In developing the Summer School program, there were some actions that had a prominent role, both as a process of self-learning of the Parma team and as an outcome of the ROMOR project (Awadallah et al. 2019), where the Parma team had the opportunity to collaborate with the most experienced European and Palestinian partners in the management of research outputs.

One of the first activities was an applied documentary research (benchmarking) on the teaching of "data management" and "FAIR principles", together with conversation with experts involved in teaching in ROMOR (Peter Burnhill, Janet and David Anderson, Susanna Mornati, Joseph Torn, Ilaria Fava, Marialaura Vignocchi, Marisol Occioni, Emma Lazzeri) and together with Paola Gargiulo, Elena Giglia: all of them allowed us to identify different education and training proposals on the management of research data, which was a key point for analysis of the needs and the planning phase of the program.

The broad context of the Summer School at the University of Parma covered the FAIR principles relating to interoperability and re-use of research data.

FAIR data are data which meet standards of findability, accessibility, interoperability, and reusability. The expected learning outcomes were:

- Planning a management campaign of research results
- Re-use of research data for the entire research life cycle.

2.2 Learning Material

The materials and the bibliography of the Summer School were selected from various sources: the reference materials of the previous training courses of the ROMOR Project and the presentations used for the lessons that were regularly taught. A variety of educational material has been used for the study schools, such as videos, links, documents to adhere to the learning objectives and the expected outcomes.

Copyright was taken into consideration, providing all the material that the students of the Summer School could be used in open access with a CC-By license. All teaching material was available from the FAIR RDM Study School website (http://fair-rdm.unipr.it).

2.3 Course Content

A combination of frontal and laboratory lessons has been planned. The first day was an introduction to the FAIR principles for the quality of research data together with a

focus on the user of the data and the needs for re-use and interoperability. The second day was dedicated to the new generation of institutional repositories and OpenAire with practical exercises for evaluating different types of needs. The third day focused on the legal issues of copyright and privacy together with policy impact on the organization of support services. A "Data Carpentry" laboratory was planned and managed in collaboration with Data Carpentry team in the last two days (http://datacarpentry.org).

Other mobility activities included visits to the libraries of Italian Universities of Bologna and Venice (the latter visited virtually) to learn about the management of research data.

The final assignment (group activity) was based on the planning of an information website for accessing from a single point the support service for the management of research data.

The course structure is reported in the Appendix.

2.4 Selection of Participants

Data stewards are support staff and subject-specific experts and often have different skills and knowledge that they bring to the research data management team. However, this also means that data stewards may have varying degrees of understanding of general needs and expectations in the field of research and Open Science data management. However, the Summer School focused on an intensive training program on topics of common interest for different backgrounds.

The Summer School was planned for a total of 40 Palestinian participants and a maximum of 20 external participants. Meeting rooms and facilities have been planned based on the expected size of the class. Three tutors (Computer Engineering students with personal interest in the School topics) were also recruited to support laboratory activities, as recommended by the Data Carpentry laboratory methodology.

The criteria for the selection of the participants in the School of Study were the following. Staff and researchers from the Palestinian institutions affiliated with ROMOR were selected by the ROMOR partners based on their previous involvement in the project, the role within the university and their knowledge of OAIR issues.

The criteria for Italian candidates were based on their background and role (librarians, researchers, IT staff, etc.) declared in their online application, together with their motivation to participate. The applications were accepted according to the order of the applications received. The window for the application of participants remained open for 10 days.

A total of 63 qualified applications were received (40 Palestinian participants and 23 Italian participants). All candidates were admitted with an official invitation, as the total number was close to the maximum expected and all candidatures showed adequate motivation and background.

2.5 Attendance

The duration of the Summer School was 5 days, from 1 to 5 July 2019. Participation in the lessons was very constant, despite the hot temperature of the Italian summer and the

very dense Summer school program. Participants enthusiastically participated in all the lessons, engaged in discussions, mixed in working groups, were stimulated to active and proactive interactions between them and with the teachers.

2.6 Institutional and Media Coverage

The Summer School was opened with a welcome speech by the Vice-Chancellor for Research of the University of Parma. In the coming days, it included brief statements by the Vice-Chancellor for International Relations and the Director of the Center for International Cooperation at the University of Parma.

The Summer School was advertised on the University of Parma website and in specialized mailing lists. The Summer School has also been marked with interviews, photos and articles from local newspapers and TV.

3 Summer School Final Evaluation

At the end of the Summer School, two types of assessment were carried out with the 63 participants with differentiated questionnaires for Palestinian participants and Italian participants.

All participants were asked to evaluate the general programme structure and organization of the contents on the various aspects of the Summer School to gather their opinions, as well as their advice for future editions of the School.

3.1 Evaluation of Summer School Programme

Many Italian participants declared that they were very satisfied with the open format of the School. As for the relevance of the contents, a great degree of satisfaction has been reported, which allows us to affirm that the topics have been well selected (Figs. 1 and 2).

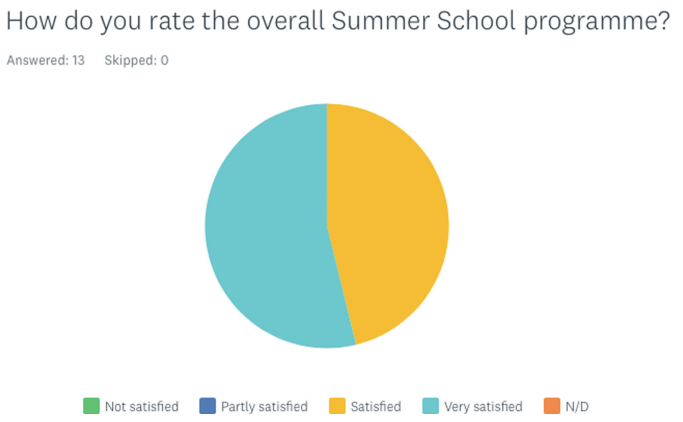

Fig. 1. Satisfaction for the Summer School programme.

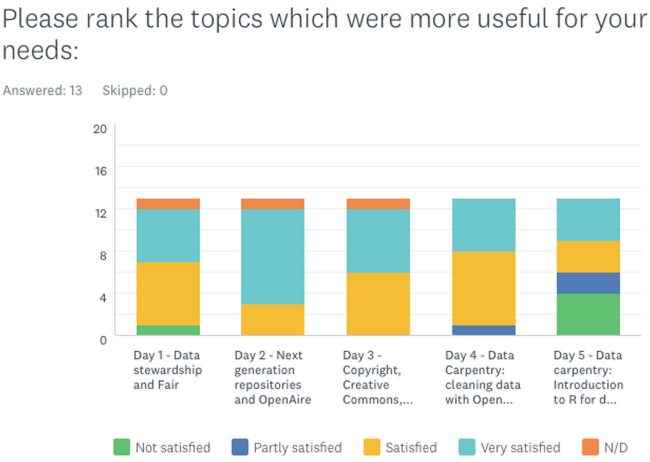

Fig. 2. Satisfaction for the different topics of the Summer School.

The contents (Fig. 2) that received the highest satisfaction (69% very satisfied) were in the second day in the Next Generation repositories and OpenAire, together with the fourth day (60% satisfied) in Data Carpentry using Open Refine. Both of these sessions were based on practical exercises following a theoretical introduction.

The contents proposed in all other days of the Summer School satisfied on the average about 50% of the participants, including the Introduction to FAIR and Copyright and privacy.

Some problems can be noticed on the last day, devoted to the application of R, in which satisfaction fell to 24% and very satisfied to 31%, with 31% partially satisfied. This session was somehow more technical and perhaps not fully suitable for all attendees, in the short time available.

The Summer School organization received the highest percentage (70%) of satisfaction, together with classroom and accommodation (61%).

About the mastering of the subject by the speakers, 60% of the participants stated that they were very satisfied. The pedagogical approach of all the teachers and tutors was highly appreciated. Also the learning material obtained the following evaluation: very satisfied by 46% and satisfied by 36% of respondents, partly satisfied by 15% of respondents.

In general, it can be concluded that the School had many positive elements and that it can be repeated making some adjustments, mainly in the distinction between theory and practice.

Open comments and suggestions from participants include:

- *A more practical focus on the long term preservation of digital objects, with example of software and formats to be used.*
- *Other platforms and tools (open and free if possible) for data repository.*
- *More practical applications of RDM on a technical point of view, for instance through case studies, working groups and labs.*
- *Keep technical sessions separated from the rest of the training for those really interested in hands-on training.*

– *Software and tools for data management as lecture on a handful programmes focusing on the concept and tasks or assignment that those tools can accomplish.*

4 Conclusions

Courses with classes of this size generate many expectations in the participants, who can interact and share experiences and establish contacts with colleagues from different institutions at national and international level. The development of the FAIR RDM Summer School provides the following conclusions, which should be taken into account in other education and training projects of this style:

– The role of the data steward is truly *international*, with FAIR principles and homogeneous organizational criteria that can be applied to various contexts. The multicultural aspects of the Summer school were very gratifying, for both Palestinian and Italian participants.
– Another important aspect is the need to include *flexibility* in the programme. In a course open to participants with different backgrounds and from different institutions, the times, beyond their programming, should allow and give attendees the opportunity to develop assignments and do exercises at different times, although such an organization involves a greater demand for teachers and tutors.
– In a study school developed with different teachers and tutors there must be a permanent *communication and coordination* effort in order to guarantee homogeneity in delivering content and provide full support to participants.

Appendix

Data Stewardship and FAIR principles - Monday 1st July

9.30-11.00	*Welcome, Introduction to the Study School, Data stewardship core* Stefano Caselli, Anna Maria Tammaro, Janet and David Anderson
11.30–13.00	*Data Stewardship (1): Introduction and Overview of FAIR* Peter Burnhill
14.00–15.30	*Data Stewardship (2): Focus on the Data User. Provision for Access. Relationship to Articles, Books and Web resources* Peter Burnhill
15:30-16:30	*Group work* Group task and to be presented in class by end of the School

Next Generation repositories - Tuesday 2 July

9.30–11.00	*Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group - Part 1 - Behaviours* *Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group - Part 2 - Technologies and Implementations* Susanna Mornati
11.30–13.00	*OpenAIRE and RDA: community driven tools and support for research data access and reuse* Emma Lazzeri
14:00-16:30	*Library system organisation for RDM: Visit to the University of Bologna* Marialaura Vignocchi

Copyright, CC and Privacy - Wednesday 3 July

9.30–11.00	*Copyright, creative commons, privacy issues for research output management* Janet Anderson and David Anderson
11.30-13.00	*Data stewardship (3). Use, preservation and citation of 'web resources': a look into the future* Peter Burnhill
14.00–15:00	*Single point of entry: RDM management at University of Venice* Marisol Occioni (Videoconference)
15:00-16.30	Group work

Data carpentry - Thursday-Friday 4–5 July

Data carpentry Laboratory and Group work Marianne Corvellec, Nilani Ganeshwaran Day 1: 9:30 am-1 pm : *Introduction to data, and best practices in using spreadsheets* 2 pm-4:30 pm : C*leaning data with Openrefine* Day 2: 9:30 am - 1 pm : *Introduction to R for data analysis* 2 pm - 4:30 pm : *Using R for data visualization and generating reports*

References

ARL: Strategic Thinking + Design Initiative report (2015). https://www.arl.org/wp-content/uploads/2015/02/strategic-thinking-design-full-report-aug2014.pdf

Awadallah, R., et al.: Setting Up Open Access Repositories: Challenges and Lessons from Palestine, iPRES, Amsterdam (2019)

Borgman, C.: Big Data, Little Data, No Data: Scholarship in the Networked World. MIT Press, Cambridge (2015)

Corrall, S.: Roles and responsibilities: libraries, librarians and data. In: Pryor, G. (ed.) Managing Research Data, pp. 105–133. Facet, London (2012)

Mons, B.: Data Stewardship for Open Science: Implementing FAIR Principles. CRC Press Taylor and Francis Group, London, New York (2018)

Tammaro, A.M.: Una proposta non sovversiva. Biblioteche oggi **36**, 65–71 (2018)

Tammaro, A.M., Matusiak, K., Sposito, A., Casarosa, V.: Data curator's roles and responsibilities: an international perspective. Libri **69**(2), 89–104 (2019)