

Published in :

KURBANOGLU, Serap, BOUSTANY, Joumana, ŠPIRANEC, Sonja, GRASSIAN, Esther, MIZRACHI, Diane, ROY, Lorie [eds.], 2018. *Information Literacy in the Workplace :5th European Conference, ECIL 2017, Saint Malo, France, September 18-21, 2017, Revised Selected Papers*. Cham : Springer. ECIL 2017. Communications in Computer and Information Science, vol 810. ISBN : 978-3-319-74333-2

which should be cited to refer to this work.

**The final publication is available at** : [https://link.springer.com/chapter/10.1007%2F978-3-319-74334-9\\_15](https://link.springer.com/chapter/10.1007%2F978-3-319-74334-9_15)

## **Training trainers for research data literacy: a content- and method-oriented approach**

René Schneider

Geneva School of Business Administration, Carouge, Switzerland  
[rene.schneider@hesge.ch](mailto:rene.schneider@hesge.ch)

**Abstract.** The paper describes all aspects of a train-the-trainer course that was applied to research data literacy to serve two goals: a) allowing future trainers to gather and combine content according to the needs of their students and organizational parameters, b) providing them with essential knowledge concerning didactical approaches to the matter. The course was centered around a series of practical exercises that asked the participants to start with a two dimensional representation of all possible competences and skills to acquire before reducing them to a linear sequence of the concrete course content, and finally the punctual specification of a single lesson.

**Keywords:** data management, research data literacy, vocational training, train-the-trainer competences

### **1 Introduction**

Based on a theoretical work published in 2013 by the same author [1], the paper will show in detail its transposition to practice by outlining the composition as well as the course of action of a one day train-the-trainer workshop. This workshop combined domain-specific (i.e. related to research data management) as well as didactical elements that were explained and put into hands-on exercises throughout the whole day with about one third of the time being spent for theoretical input and the remaining two third for practice.

The paper will briefly describe the basic and preceding modules and mention all components of the workshop that consisted of a) an introductory example course, b) a methodological approach to create vocational trainings, c) a specific method to build a two dimensional view on research data management course content representing all aspects of research data literacy, d) didactical knowledge to structure a course in detail, e) an overview of already existing courses being available, and finally f) an exercise to create and visualize a specific course unit.

We will also give diversified references on research data literacy in the introducing part and finally discuss the results of our own study, examining the particular interests and limits of this work.

## 2 Theoretical Background

Data, in opposition to information or knowledge, has for a long time been the step child of Library and Information Science. They (i.e. data) were generally seen as the mere alphabet or primitive subset for information, which itself (i.e. information) was often defined as “*knowledge communicated*”. Thus, per definitionem and metaphorically speaking, information always oriented its face to the bright sphere of knowledge and showed a rather cold shoulder to the brittle grounds of data. This perspective built the consensus for more than half a century (regardless the rise of computer technology, i.e. electronic *data* processing) during that time and was not put into question; maybe because the sphere of knowledge bears meaning and allows reasoning about semantics, being itself one of most challenging fields in overall science. But semantics still remains difficult to define and capture exactly and due to the problem of disambiguation a camino real has not been found so far.

Data on the other side are usually quite tangible, i.e. easy to keep, handle, describe, analyze, transform etc., thus showing many pragmatic effects that justify the complete turnover that took place in the last decade: in this new scenario that is characterized by buzz words like big data, open data and linked data, data is seen as a rather vivid structure that creates information and allows knowledge extraction via algorithmic processing which leads to the revelation of unknown structures. This shift in perspective has a twofold impact on science: 1.) under the form of data science, i.e. a scientific paradigm mainly based on a combination of (inferential) statistics and machine learning; 2.) under the form of research data management incl. data curation and thus defined as the “active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education” through activities that “enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time” [2]. In this context, Research Data Management in general or Data Curation in peculiar can be seen as minor concerning quantity (opposed to big, linked, or open data) but major or at least as important concerning quality, due to the fact that big data, open data and linked data are scientific movements that need proper data to treat and well curated data to perform better.

In the eyes of the author this shift in perspective (from knowledge to data) will also lead to major changes in the domain of library and information science and its subdomain of information literacy: the way we understand, treat, compile, and deal with data will on the one hand lead to new professional roles (such as data librarians, data archivists and data curators as well as new cooperations between the latter and data managers, data scientists and data journalists) (compare [3]) and on the other hand to new training programs in information or better said data literacy (see [4] and section 3 of this paper). This means that a) curricula have to be created, b) training materials have to be delivered, c) data-oriented teaching approaches have to be developed, and d) trainers have to be trained to form students from all disciplines and library and information science students in peculiar as well as scientists and data professionals as described above.

This article should be considered as a first trial to answer this growing demand and report on an initiative being undertaking in the context of a Swiss project named

Train2Dacar (standing for Train the Trainer for Data Curation in Advanced Research). To do so, we will shortly present the modules developed for a vocational training in research data management and describe in more detail a train-the-trainer approach for the creation of flexible and context adaptive course materials.

### **3 Research Data Management and Literacy Training**

Data literacy [5], as information literacy and data management of any kind needs training. It is no surprise that training is identified to be one the major services in the realm of research data management [6]. Training – like all forms of education – consists mainly of two components: 1.) of knowledge communicated (i.e. information) and 2.) the methods used to transfer this knowledge (i.e. formation), in other words and in its uttermost simplification the “What” and the “How” of a matter to be taught [7]. A simple training program is usually focused on a certain content to be mediated. Train-the-trainer programs should cover in contrast both aspects. Interestingly, the number of train-the-trainer programs in research data management are not very large in number and, if they exist, generally limited to the “What”, i.e. classical training.

Those classical training are already quite numerous [8], with some of them being well known and of high quality (e.g. the ANDS 23 Things on RDM program [9], MANTRA [10], or Essentials for Data Support [11], to name only a few.). As mentioned, train-the-trainer programs are quite smaller in number, starting with the Library of Congress Program on Digital Curation [12] where the train-the-trainer aspect is understood as a training program focused on content and given to a limited number of participants whose role is seen as that of multipliers. These multipliers will then train others after having done the initial training. Hence, no real distinction between content and didactical methods and their combination is done.

Since the success of a training always depends on the methods used for training and due to the fact that adult training (or andragogy) differs considerably from classical pedagogy [13], it was decided to add a specific train-the-trainer module for (research) data literacy that combined both aspects to classical modules for research data management. This approach will be presented in the following sections of this paper.

### **4 Training trainers for research data management**

For a proper understanding of the train-the-trainer workshop described in detail in this chapter, it is important to know that the workshop itself was the last part of a modular vocational training composed of nine modules, with four basic and four intermediate modules (and with a duration of four hours for each module) whereas the latter were specifically dedicated to research data management in digital humanities and thus play a minor role in the context of this paper. The essential thing to understand is the modular composition and division in three types of training modules: basic, immersive, and didactical. The idea behind was to start with the mediation of factual knowledge and to

finish with andragogic (i.e. didactics for adults) skills to further enable the participants to pass over the knowledge acquired and the experiences made to their own clientele, becoming thus multipliers in this new domain. The course was mainly dedicated to librarians and archivists but open to researchers, managers, and computer scientists too.

#### **4.1 Content-oriented modules**

We will focus on a detailed description of the train-the-trainer module (see 4.2) and only give a short factual description of the four basic modules. A French and German e-learning module of all modules can be found under [www.researchdatamanagement.ch](http://www.researchdatamanagement.ch) [14].

##### **Module 1: Basics**

The training started with some basic definitions and several illustrations explaining what data, scientific data, and research data management mean to all sciences as well as their specific disciplines. This introduction was followed by a presentation of the two major theoretical models used in research data management, i.e. the data curation cycle and the data continuum model. The second part of the introductory module presented the three basic components that are needed to create the framework in which research data management can take place: policies, data management plans, and guidelines provided by the archive that will ingest the data created during the scientific process. The presentation of the latter document was also seen as a bridge to the subject of long term preservation presented in the second module.

##### **Module 2: Data throughout its Lifecycle**

The second module was divided into four sections that addressed different stages of the data lifecycle and their implications for the management of research data with a special focus on long term preservation. Section one was dedicated to the subject of active data management, i.e. the handling of data during the research process. The second part aimed to give an understanding of data storage and discovery infrastructures with a special emphasis on data catalogues, registries and repositories. The subsequent part focused on the description of data (i.e. metadata) and its impact on the findability and (re)usability of research data. The fourth and final section concentrated on basic principles and procedures of long term preservation that ought to ensure long term access and (re)usage of research data. All four sections were accompanied by group work with practical examples that encouraged discussion.

##### **Module 3: Use of Data**

The third module demonstrated tools and services to support researchers to discover, use and reuse data to develop further their research. In the first part of the module, different datasets were uploaded into a visualization tool to show the benefits of data visualization for different research needs. The second part of the module looked into the concepts of Linked Data and the Resource Description Framework (RDF) as a set of techniques that allow to integrate and combine heterogeneous data sources and to

address new and challenging research questions.

In a hands-on session, it was illustrated how data can be explored, queried and expanded by means of several external RDF resources already present in the Linked Open Data cloud using Semantic Web technologies. A RDF dataset was explored to demonstrate the power of linked data and the range of questions that can be asked and answered using the query language SPARQL.

#### **Module 4: Publication and Services**

The last module started with a presentation of the major issues concerning the publication and citation of data. To illustrate the added value of data publication, a special focus was set on data papers and data journals. Afterwards, a further focus was set on the problem of dynamic and persistent identifiers for research data assuring the link between short term and long term repositories.

The module ended with a presentation of the major services that evolve around research data management. After studying the activities related to each service thoroughly, the participants then worked in groups to develop proposals for roadmaps or describe in more detail the services they are planning to install at their home institutions. The different proposals were then presented and discussed by the participants.

### **4.2 Method-oriented modules**

The one day (i.e. eight hours) train-the-trainer workshop was designed as an alternation of theoretical inputs and practical exercises whereas the theoretical parts did never last longer than 20 minutes and emphasis was always put on the practical parts including the explanation and discussion of the results obtained during the hands-on sessions.

The workshop started with (1) a short ice-breaking exercise that was immediately followed by (2) a 45 minutes example course simulating a lunch time lecture given as an introduction to research data management by a librarian to researchers. This example course included two exercises of minor size (in the middle and towards the end) and followed a didactical model known under the name of e-class reference model [15], a model that was also used for the elaboration of the e-learning modules mentioned earlier [14].

This model was explained in detail in the following part of the training that consisted of (3) a twenty minute general introduction to teaching with a special focus on adult learning, i.e. a subfield of teaching also known under the term of andragogy.

After a first break, the training continued with (4) another 20 minutes theoretical introduction, this time with respect to research data literacy, namely a short lecture based on the paper mentioned earlier and immediately linked with (5) a one hour exercise that transformed the theoretical explanations on research data literacy into practical operations similar to those of a non-deterministic algorithm. The exercise was derived from the matrix presented in Table 1 that consists of an attribution of specific skills in the field of research data management to eight major competences in the domain of research data literacy as derived in [1].

The exercise consisted of the elaboration of a metro line based on this attribution, because – metaphorically speaking – the competences in research data literacy (column

1 of Table 1) can firstly be seen as lines of a metro network and secondly the according skills in research data management (column 2 of Table 1) as corresponding stops along these lines. Skills that are difficult to allocate to one single line and that might be stops in another or even more than two lines build intersections between those lines. Thirdly, metro maps are usually divided into zones, generally with respect to their proximity or distance to the center. Hence, the last analogy proposed was the division of the metro map into several zones, whereas the central zone should gather all skills that were to be taught to all participants in a minimal amount of time, the second zone those skills that needed more time and a distinct audience etc.

**Table 1.** Competences and Skills

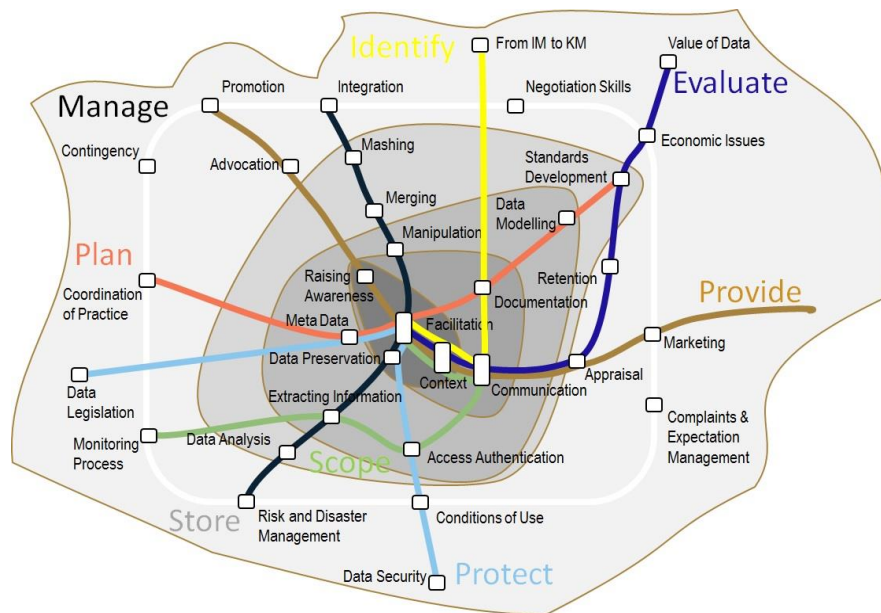
<b>Competences in Research Data Literacy</b>	<b>Specific Skills in Research Data Management</b>
Identify	Documentation (research environmental, temporal) / Context / From Information Management to Knowledge Management
Scope	Monitoring Process / Extracting Information from Data Models (and People)
Plan	Data Modeling / Meta Data / Standards Development
Store	Data Analysis and Manipulation / Merging, Mashing, Integration
Protect	Data Preservation / Data Security / Access Authentication / Conditions of Use / Data Legislation
Evaluate	Data Appraisal and Retention / Value of Data / Economic Issues
Manage	Data Appraisal and Retention / Value of Data / Economic Issues
Provide	Facilitation, Communication / Raising Awareness

For the sake of simplicity and according to the distinction done in the DigCurV [16] project that distinguished between three lenses, namely that of practitioner, manager and executive, special attention was given to the distinction of abilities within these lenses, namely the three capacities “is aware of”, “understands”, and “is able to”. Following these distinctions, it was recommended to the participants to divide the map into three zones, either a map with three overlapping zones for practitioners, managers, and executives or concentric zones following the distinction within each lens, i.e. a central zone for “Awareness Building”, an intermediate zone for “Creating deeper understanding” and a peripheral zone “Enabling for Practice” (without necessarily using these labels), or similarly, according to the “23 Things on research data management” distinction between “Getting Started”, “Learn More”, and “Challenge Me” [8]. Figure 1 shows an idealistic view of the exercise established by the author of this paper. This map was also used during the training for the sake of illustration.

It should be stated in this context that the purpose of this exercise was not to achieve a “first yield” for all aspects of a training program during the one hour given to come to a result but rather to bring a group of four to five future trainers with different backgrounds together and to make them exchange their different points of view via the relay of a two dimensional map that might be used as a tangible sketch that allows flexible modifications on the one hand and continuous evolution on the other.

This meant practically that several groups of four to five participants were created and

several card sorts [17] were handed out that represented different points of view of the common field-of-interest (in our case research data management and the eight major competences). Further material was given to allow the building of intersections with different colors and the distinction of zones via numbers. After one hour, each group was asked to shortly present the maps, report on the major outcomes and difficulties that occurred, before switching to the second half of the workshop which was oriented towards the essential didactical and methodological foundations for the sequential elaboration of a course.



**Fig. 1.** Metro networked derived from Table 1

After giving (6) a 20 minutes general introduction concerning the rough planning of a course, the participants were asked (7) to pick out those contents of the map that they would like to integrate in the course to create and thus to build a linear subset of the two-dimensional map. To ease the decision, (8) a short overview of existing onsite and online training courses was presented (similar to the one given in chapter 3 of this paper).

In the final steps of the workshop, the participants (9) selected one peculiar class of the course (i.e. one punctual stop in the linear subsection of the two dimensional map) and worked on the concrete elaboration of a specific lesson. The instructors provided the participants with corresponding sheets that enabled them to plan a course in detail by defining the framework and goals of the course, the theoretical input, practical exercises and didactical methods used including the duration of each of the parts. Each group was then invited to illustrate on a flip chart the content and course of action.

The workshop concluded with short presentations of the results of this last exercise and (10) an oral feedback and evaluation given by all participants.

All course materials as well as the results of all group works were gathered on a web

site during the two weeks that followed the course.

## 5 Final remarks

In this paper we presented a train-the-trainer approach for research data literacy combining methodological, didactical, and content-oriented aspects including a flexible method that allows trainers to determine the content by putting it into context according to the needs of their public and the organizational circumstances of the training. This method of combining competences (as metro lines) and skills (as stops) in levels (zones) in a two dimensional metro map, reducing it then to a linear subset of a course to be given and the punctual extrapolation of a singular lesson was tested beside two one-day-training workshops in another context with students in Library and Information Science at Geneva Business School of Administration. The students were asked to create courses of different length for the following topics: “Digital Data Management”, “Trusting Results from Search Engines”, “Plagiarism”, “Use of e-readers”, and finally “Risks of Social Media”. For each topic the core competences, i.e. the categories for the card sorting to build the metro map were adapted with respect to the context.

Experience showed that the overall majority of the workshop participants and students appreciated the balance of theoretical input and practical experience as well as the idea of combining content- and method-based aspects in a setting that allowed to create mutually training scenarios. However, some participants seemed to have problems with the two dimensional representation, probably because they prefer an immediate linear or punctual approach to create a limited number of sequential or simply punctual teaching units.

A German and French online-version of this course can be found under [www.researchdatamanagement.ch](http://www.researchdatamanagement.ch) [14].

**Acknowledgements.** The author would like to thank Eliane Blumer for her contributions during the preparation and execution of the train-the-trainer workshop described in this paper.

## References

1. Schneider, R.: Research Data Literacy. In: Kurbanoglu, S., Grassian, E., Mizrachi, D., Catts, R., Špiranec, S. (eds.), *Worldwide Commonalities and Challenges in Information Literacy Research and Practice Communications in Computer and Information Science*, Vol. 397. Springer International Publishing, Cham (2013)  
<http://dx.doi.org/10.1007/978-3-319-03919-0>
2. Cragin, M.H., Heidorn, P.B., Palmer, C.L., Smith, L.C.: An Educational Program on Data Curation. In: Science and Technology Section of the annual American Library Association conference. Washington, DC. (2007)
3. Lyon, L., Mattern, E.: Education for Real-World Data Science Roles (Part 2): A Translational Approach to Curriculum Development. In: *International Journal of Digital Curation*, 11(2), 13-26. (2016)



4. Tibbo, H. R.: Digital Curation Education and Training: from Digitization to Graduate Curricula to MOOCs. *International Journal of Digital Curation*, 10(1), 144-153 (2015)
5. Koltay, T.: Data literacy: in search of a name and identity. In: *Journal of Documentation*, Vol. 71 Issue: 2, pp.401-415. (2015) doi: 10.1108/JD-02-2014-0026
6. Jones, S., Pryor, G., Whyte, A.: 'How to Develop Research Data Management Services - a guide for HEIs'. DCC How-to Guides, Edinburgh: Digital Curation Centre. (2013) <http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>
7. Prongué, N., Schneider, R.: The What and the How of Research Data Management. Towards a Unified View of Train-the-trainer Competencies. *Proceedings of the 12<sup>th</sup> International Digital Curation Conference (IDCC) Edinburgh, Scotland, 20-23 February.* (2017) <https://hesso.tind.io/record/1974/files/Poster.pdf>
8. Molloy, L.: Research Data Management (RDM) Open Training Materials. (2017) <https://zenodo.org/communities/dcc-rdm-training-materials>
9. Australian National Data Service (ANDS). 23 (Research Data) Things. (2016) <http://www.ands.org.au/partners-and-communities/23-research-data-things>
10. The University of Edinburgh: Mantra – Research Data Management Training. 2010-2011
11. Grootveld, M., Verbakel, E.: Essentials for Data Support: Training the Front Office. *International Journal of Digital Curation*, 10(1), 240-248. (2015)
12. Library of Congress: Digital Preservation Outreach and Communication – Baseline Digital Preservation Curriculum. (2017) <http://www.digitalpreservation.gov/education/curriculum.html>
13. Canadian Literacy and learning network: Principles of Adult Learning. Literacy.ca. (2016) <http://www.literacy.ca/professionals/professional-development-2/principles-of-adult-learning>
14. Mastrandrea, E., Prongué, N., Schneider, R., Stettler, N.: Kursbuch Forschungsdatenmanagement. HTW Chur / HEG Genf. (in press) [www.researchdatamanagement.ch](http://www.researchdatamanagement.ch)
15. Gerson, E.: E-CLASS: Creating a Guide to Online Course Development for Distance Learning Faculty. (2000) <http://www.westga.edu/~distance/ojdla/winter34/gerson34.html>
16. Moles, N. DigCurV Curriculum Framework: Applications in the Development of Digital Curation Vocational Training Programmes. *iConference 2014 Proceedings*. (2014)
17. Hudson, W. Card Sorting. In: Soegaard, M., Dam, R.F. (eds.). *The Encyclopedia of Human-Computer Interaction*, 2nd Ed. Aarhus, Denmark: The Interaction Design Foundation. (2013) [http://www.interaction-design.org/encyclopedia/card\\_sorting.html](http://www.interaction-design.org/encyclopedia/card_sorting.html)