'Essentials 4 Data Support': five years' experience with data management training

Grootveld, Marjan; Verbakel, Ellen

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

'Essentials 4 Data Support': five years' experience with data management training

Ellen Verbakel
3TU.Datacentrum, NL

Marjan Grootveld
Data Archiving and Networked Services (DANS), NL

Corresponding author:
Ellen Verbakel, TU Delft Library, PO Box 98, NL-2600 MG Delft
+31 15 2787296
Email: p.m.verbakel@tudelft.nl

Abstract
This article describes a research data management course for support staff such as librarians and IT staff. The authors, who coach the participants, introduce the three course formats and describe the training in more detail. In the last years over 170 persons have participated in this training. It combines a wealth of online information with face-to-face meetings. The aim of the course is to support the participants in strengthening various skills and acquiring knowledge so they feel confident to support, advise and train researchers. Interaction among the students is embedded in the structure of the training, because we regard it as a valuable instrument to develop a professional network. Recently the course has taken on a new challenge: in addition to the regular courses, a couple of in house trainings has been delivered on request. The paper ends with a description of the key group assignments for such compact trainings.

Introduction to 'Essentials 4 Data Support'

At the end of 2011 a *Data Intelligence 4 Librarians* course was developed by

3TU.Datacentrum to provide online resources and training for digital preservation

practitioners, specifically for library staff. The preparation, design and mission of this

training is described in De Smaele et al. (2013). 'The course objectives are to transfer

and exchange knowledge about data management, and to provide participants with

the skills required to advise researchers or research groups on efficient and effective

ways of adding value to their data.' (De Smaele et al., 2013: 218).

Lessons learned during these trainings and developments in the Research Data

Management landscape have led to a revision in 2014. By then, the training had

become the flagship service of Research Data Netherlands (RDNL). RDNL is a coalition

in the field of long-term research data archiving, consisting of 4TU.Centre for Research

Data, Data Archiving and Networked Services (DANS) and SURFsara. The goal of the

training was redefined as 'The Essentials 4 Data Support course aims to contribute to

professionalization of data supporters and coordination between them. Data

Supporters are people who support researchers in storing, managing, archiving and

sharing their research data.' (Essentials 4 Data Support, 2014). The training setup as

well as all training content were thoroughly revised, as described in (Grootveld and

Verbakel, 2015).

As of May 2016, more than 170 data supporters from Dutch universities, Higher Vocational Education organisations, academic hospitals and other knowledge institutes have participated in the face-to-face training that Research Data Netherlands provides. We are aware that the term "data supporter" is unusual, but exactly this makes it inclusive: The course appeals to (aspiring) information professionals, data stewards, data librarians, research support officers and others. The intentionally unusual term saves a lot of discussion about job titles and definitions in a highly dynamic field.

In this paper the coaches present 'Essentials 4 Data Support', one of the few courses on data management that explicitly focus on data supporters. Early 2016 Knowledge Exchange organised a survey and workshop to collect and share information on current practices around RDM training. One of the conclusions in the survey report is: 'It is striking that a near totality of respondents provides training for PhD students (…) librarians, data curators and IT departments/computer centres are catered for by a third or less of respondents.' (Goldstein, 2016: 7).

In the next section the three training variants are introduced. This is followed by sections on the setup of the blended learning course and on the intended competencies and learning goals. After a sketch of the course website the authors describe how the regular blended learning course enables them to also accept invitations for tailor-made in house trainings.

Training variants

Since the first training in 2011 the course materials are publicly available at the website, both in English and in Dutch. The current offering consists of three variants:

- The full – blended – course consists of two face-to-face days, with fellow students, experts and coaches. In the six weeks between the first and the second day students familiarise themselves with the content of the online learning environment, do assignments,
and give feedback on the assignments of their fellow students. For this they use the group's private forum. Students pay a fee for the full course.

- In the Online+ course, registered students have access to the online course materials. They are not entitled to support from coaches, but are welcome to contribute to discussions in the open forum.

- Online-Only students can take the course at their own pace, without access to social features.

The next section describes the set-up of the blended course. Recently, we have accepted a few requests for in house training. Such trainings are bespoke and reduced variations on the full course and will be described in Section 'Recent developments'.

Set-up of the blended *Essentials* course

During the six weeks of a full course the coaches and the students have a very intensive contact. The official start is at the first face-to-face day. Here the students meet their fellow students and the coaches. The coaches inform the students about the course, what will be asked from the students and what the coaches will deliver. In a game setting the students discuss several aspects of data support – differences in the maturity of data management in their respective organisations start to show. On the agenda are also the first expert presentations. The coaches have a broad network of specialists working in the field of research data management. These experts are invited to give a presentation on their daily work in the field: What problems do occur, what worked well, what didn't? What experiences can they share with the students? In general, the students appreciate these presentations very much; they ask many questions which feed the discussion and the subsequent assignments. Data management planning is a standard topic on the first day, to prepare the participants for their first assignment, viz. to write a data management plan.

A very important part of the course is the communication among the students and with the coaches. A private forum is provided for the group at the course website.

Students have to upload their assignments and are required to give feedback on the work of their fellow students. This helps them to explore different angles and approaches. Furthermore, questions can be posted at the forum about the course and/or about the content. Because the forum is restricted to the group the students usually feel safe to do so. This can lead to very vivid discussions on various aspects of data management. The coaches use the group forum too for their comments on the weekly assignments and to answer questions.

The second face-to-face day again contains one or two guest lectures, but a substantial part of this day is reserved for student presentations. One of the assignments, spanning the full six weeks, is to follow recent developments in one of the fields of research data management close to their personal learning goals and to share some findings with the group. That gives an overview of the subjects they learned about and how they approached the task – varying from using general search engines to subscribing to data and research blogs, mailing lists, and twitter streams. From their colleagues' presentations on the information gathering process and the achieved results every student learns a lot on recent topics.

The final task at the second day, called 'good intentions', aims at the immediate future: The students must formulate data-related activities that they will carry out once they return home, e.g. design a data support workflow, discuss institutional data policies with their manager, or localise data teaching materials for young researchers. The activities are ideally formulated as small steps towards a concrete goal that can be reached in the next six months. At that time the coaches contact the former students to ask what has been achieved. The students appreciate the interest, as they recognise the purpose to keep the flow of the course going during daily work, and to really put into practice what they found inspiring.

Competencies

In the process of developing as well as during revising the course the following competencies were defined to be essential for the future data supporters. A data supporter:

- Skillfully handles ICT:

  Efficiently uses available information technology. In the last decades research data became digitally born, and ICT knowledge becomes even more important.

- Shows entrepreneurship:

Proactive attitude to improve data services in response to changing needs in the field. Keeps an eye on trends which emerge in the profession, knows where knowledge is available (networks) and disseminates important information to key people in the organisation.

- Sees from the whole:

  Acknowledges that data are only part of the scientific lifecycle and is aware of the significance research data have or carrying out scientific research.

  Sees data and information services as part of larger whole in which decisions are made.

- Has consulting skills:

  Can handle questions skillfully and knows when to address a dedicated expert. Can empathise with customer perceptions.

- Has co-operative skills:

  Examines how collaboration with others (employees, researchers, institutions) may enhance service provision.

The DigCurV project has developed the DigCurV Curriculum Framework (DigCurV, 2013) which offers a means to identify, evaluate, and plan training to meet the skill requirements of staff engaged in digital curation, both now and in the future. This framework has inspired us when we redesigned the training, and the DigCurv game is

still part of the face-to-face training (see Section 'Setup of the blended *Essentials* course'). A more recent initiative, the Interest Group on Education and Training from the Research Data Alliance, (Interest group on Education and Training handling research data, 2013) is also defining the competencies needed for handling research data. For our purposes it is important to keep in mind that RDNL intends to offer *basic* knowledge and skills to support staff; a project like EDISON (EDISON, 2015), for example, which is developing a curriculum for data scientists, has clearly higher ambitions.

Learning goals

The name for the course – *Essentials 4 Data Support* – refers to the main goal of the course: teaching the basic knowledge and skills (essentials) to enable a data supporter to take the first steps towards supporting researchers in storing, managing, archiving and sharing their research data. Some learning goals are:

- To get an overview of the data supporter's field of work, including the research life cycle and the organisation of data support, e.g. in the form of the RDNL front office – back office model (Dillo and Doorn, 2014);

- To understand the different parts of a data management plan;

- To know about the various ways to store, backup, organize and document research data;

- To know types of archives, data publication and data citation;

- To advise researchers in balancing legislation and practice;

- To be able to engage in a discussion with researchers.

The complete set of learning objectives of the six chapters can be found in (Grootveld and Verbakel, 2015).

An implicit goal, which for us has been dominant from the start, however, is that participants gain self-confidence in discussing research data with researchers. The training has been developed to encourage them in their professional dealings, by providing the knowledge to be a partner to the researcher. Strictly speaking this rather concerns the participants' attitude than their knowledge or skill set.

The *Essentials* website

In 2014 the website has been redesigned by an external developer. The platform for the course is T3Elearning, a Learning Content Management Framework based on Typo3 CMS. The content has been designed and developed with valuable input from subject matter experts within and outside RDNL, of whom several also give a guest lecture at one of the training days.

Figure 1. User interface of Essentials 4 Data Support, Chapter II

The website changed into a more user-friendly design. This was achieved by including more images and some videos with brief talks by researchers and 'how to' videos about data management planning, data sharing concerns, reproducible research, data citation and persistent identifiers. New in the revised design are also the end of chapter quizzes, which allow for a quick recapitulation of the chapter's contents.

Whereas the initial training featured chapters like 'Technical skills' and 'Advisory skills' of the data supporters, in the current version the research cycle is leading, rather than the activities of the data supporters. This can be seen at the top of Figure 1, where the planning phase precedes the research phase, as well as the phase in which researchers make their data available for reuse. Each chapter starts with a short outline of the sections and learning objectives of that particular chapter. In addition to RDNL's own content (available under a CC-BY-SA licence) all chapters contain a large amount of external links to good practices, definitions, and reading materials provided by other organisations. This makes the site a valuable work of reference – the video clip about data management planning has been viewed 3,500 times – and the students rate the website highly in their evaluations. The assignments and quizzes in the left-hand menu in Figure 1 are only accessible to participants who have enrolled in the full course. The private forum, which these students and the coaches use between the face-to-face meetings for assignments and discussion, is part of this website and remains accessible afterwards for future knowledge sharing.

Recent developments

In the previous year we initiated two new developments, which we describe in this section. The first initiative served to strengthen the contacts with and among former course participants, while the second one can be seen as a spin-off of the regular blended training.

Last autumn the coaches organised a reunion for all participants from 'Data Intelligence 4 Librarians' and 'Essentials 4 Data Support'. The response was quite good: About 20% of the students, from both course iterations, accepted the invitation to discuss their daily work in RDM. They were surprised they had so much in common, although they represented very different institutions. The key speakers of the meeting were former students, who now have become specialists, very able and willing to share their experiences. With regard to our goal to increase the data supporters' professional self-confidence, this meeting convinced us of the longer-term value of our training offer, in a way that even positive evaluations at the end of a training day cannot do.

The second development concerns in house trainings, on request. In earlier days RDNL had received some requests about in house trainings for (prospective) front office staff, but decided against it because we value the networking possibilities of a mixed audience even higher than the intended team building effect. However, two recent requests did concern mixed audiences from the start. First, a Dutch institute for Higher Vocational Education was looking for an RDM training for support staff and researchers, and eventually the institute's policy department was involved as well. Second, the Danish Forum for Data Management, which consists of all Danish universities, two national libraries and the national archive, was looking for a train-the-trainer course. The process in these cases was similar: the customer provided information about the training needs of the group and any wishes regarding the duration and intensity of the training, e.g. with or without assignments ahead of the training day or days. On that basis the coaches drafted a proposal in which they included the relevant parts from the regular face-to-face training, or suggested to develop new modules. Clearly, there cannot be a standard fee for a tailor-made training.

In comparison to the off-the-shelf training, the coaches now became trainers: without guest presenters, it was up to them to convey the contents of the course, in addition

to coaching the participants in the practical exercises. Furthermore, for these exercises relatively more time was planned than in the regular training, because all networking interaction and knowledge transfer between the participants has to take place during the meeting. The combination of the customer's requirements, the broad but not always specialist knowledge of the coaches/trainers and the time needed for interaction has led, in both cases, to an abridged and more focused version of the regular course.

In both in house trainings we have included exercises with writing a data management plan (DMP) and with identifying the stakeholders in data management. With regards to the former, participants have to draft a DMP for a given, fictitious project. This is an assignment that students do pairwise (in the regular training) or in small groups (in abridged trainings), because it is relatively demanding. However, it is usually the most appreciated assignment: All participants know that research funders and research institutes increasingly require DMPs, but they seldom have any practical experience with them at the start of the training. Having written a DMP themselves, they feel more comfortable at the prospect of advising researchers. In the stakeholder exercise students have to chart all staff roles, departments and organisations which also play a role in data management, and to reflect on the quality of all communication and

collaboration: 'Do I know who to contact when… and how do I make sure that department XYZ informs me about…?'. We regard both exercises as crucial stepping stones for exploring the diversity of research data management and this has worked out well in both in house trainings. The two in house trainings also featured an item on storing and archiving research data, and for the Danish train-the-trainer course we designed an exercise to analyse and prepare an institutional implementation of the recent Danish Code of Conduct for Research Integrity.

Conclusion

RDNL's training 'Essentials 4 Data Support' results from a clear need for professional data management support training, a need that was identified five years ago. Since then Research Data Netherlands has trained about 170 so-called data supporters, working in research libraries, IT departments, medical centres and elsewhere. As front office staff they play an essential role in research data advocacy and support for gradually realising the international Open Science ambitions. We are very pleased to see a vibrant network of data supporters in The Netherlands and abroad, and to be recognised as a party that contributes to their professional development.

References:

De Smaele M et al (2013) Data intelligence training for library staff. *International Journal of Digital Curation* 8(1): 218-228. DOI: 10.2218/ijdc.v8i1.255.

DigCurV (2013) A curriculum framework for digital curation. Available at: http://www.digcurv.gla.ac.uk/ (accessed 22 September 2016).

Dillo I and Doorn P (2014) The front office–back office model: supporting research data management in the Netherlands. *International Journal of Digital Curation* 9(2): 39-46. DOI: 10.2218/ijdc.v9i2.333.

EDISON (2015) EDISON – Building the data science profession. Available at: http://edison-project.eu/ (accessed 22 September 2016).

Essentials 4 Data Support (2014) Essentials 4 Data Support course. Available at: http://datasupport.researchdata.nl/en/ (accessed 22 September 2016)

Goldstein S (2016) Training for research data management: comparative European approaches. Report, Knowledge Exchange, May 2016. DOI: 10.5281/zenodo.50068.

Grootveld M and Verbakel E (2015) Essentials for Data Support: training the front office. *International Journal of Digital Curation* 10(1): 240-248. DOI: 10.2218/ijdc.v10i1.364.

Interest group Education and Training on handling of research data (2013). Research Data Alliance Working Group Education and Training on handling of research data. Available at: https://rd-alliance.org/groups/education-and-training-handling-research-data.html (accessed 22 September 2016).

Author biographies
Ellen Verbakel is a trained librarian, but moved to 3TU.Datacentrum where she is now working as a datalibrarian. She is situated in the TU Delft Library. She co-developed the course 'Data Intelligence 4 Librarians' and was involved in redesigning the course into 'Essentials 4 Data Support'. Being a coach at the course is a very important part of her daily work.

A computational linguist by training, Dr **Marjan Grootveld** is senior policy consultant at Data Archiving and Networked Services (DANS). She advises knowledge institutes and research funders on data management policy and practice, both in international research projects and by coaching participants of the Research Data Netherlands training 'Essentials 4 Data Support'.