Essentials for Data Support: Training the Front Office

Marjan Grootveld
Data Archiving and Networked Services

Ellen Verbakel
3TU.Datacentrum

Abstract

At the end of 2011 a *Data Intelligence 4 Librarians* course was developed to provide online resources and training for digital preservation practitioners, specifically library staff. Lessons learned during the first rounds of the course and developments in the Research Data Management landscape have led to a revision of the positioning, the structure and the content of the course. This paper describes both the three main drivers for the revision, the changes themselves and the lessons that can be drawn from them, after three training rounds in 2014 in the revised format under the new programmatic title of *Essentials 4 Data Support*.

240

Setting the Stage

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 course objectives were '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). In a sense, the training has always aimed to contribute to the data literacy that the recent RDA Europe report *The Data Harvest* so strongly advocates (RDA Europe, 2014) The preparation, design and mission of this training is described in de Smaele et al. (2013). In 2012 and 2013 more than sixty data supporters from Dutch universities, Higher Vocational Education organisations and other knowledge institutes have participated in the face-to-face training that Research Data Netherlands provides. Lessons learned during these trainings, together with developments in the Research Data Management landscape, have led to a revision of both the structure of the course and of all content, which is available online.

New Collaboration

Since this initiative there have been three major changes. First, a collaboration with Data Archiving and Networked Services (DANS) was initiated to complement the scientific disciplines of the three Dutch Universities of Technology with humanities and social sciences. In 2013, 3TU.Datacentrum and DANS launched Research Data Netherlands (RDNL), a coalition in the field of long-term research data archiving. SURFsara, the Dutch national ICT infrastructure provider for research, joined the alliance in 2014. The course is now RDNL's flagship.

Support for the Front Office

A second major change concerns the target group. In repositioning the course we have moved away from librarians to data supporters, which is made explicit in the title *Essentials 4 Data Support*. A reason for this is the actual variety in participants' backgrounds: while many do work in libraries, their job titles differ widely, and other students represent the IT department or are data specialists and data managers working on data curation in research groups.

It is the kind of variety that we want to stimulate: we welcome the trend that support staff from different units in a research institution collaborate as a "one-stop shop" for researchers. "Front office" is the name RDNL applies to this collaboration. The emerging Dutch research data infrastructure consists of four layers, representing the various stakeholders in Research Data Management with their tasks and responsibilities, as shown in Figure 1.
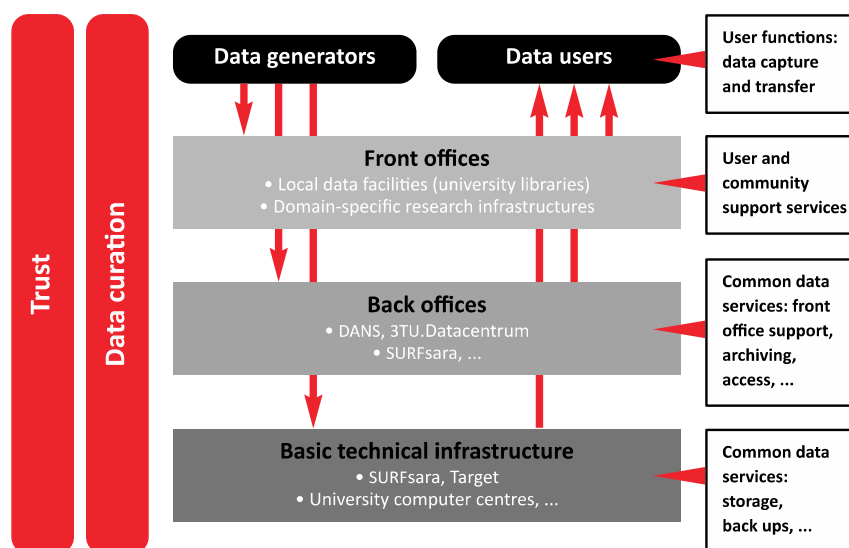
Figure 1**.** Emerging Dutch federated research data infrastructure.

Figure 1 is modelled along the lines of the *Riding the Wave* report (European Union, 2010), but with a focus on the complementary roles of front offices and back offices. The foundation is a basic technical infrastructure, which facilitates data storage and back up. Above that is a layer of back office data services, providing facilities and support for long-term archiving and accessibility. The next level includes the front office services. Front offices provide for the first-line contacts, supporting, advising and training researchers and students – the top level – in responsible data management.

When the responsibilities are made clear in such an infrastructure, each layer can rely on services and/or expertise of the others; there is no need to know or to organise everything oneself. For instance, there is a worldwide trend that research funding organisations (which are not explicit in Figure 1) demand that research data be made available for re-use and the advancement of science. Examples are the European Commission with the Open Research Data Pilot in the Horizon 2020[1] programme, the National Science Foundation[2], and the Dutch funding organisations NWO[3] and ZonMw[4]. Depositing data in a trusted digital repository is typically part of such a scenario, but researchers in their role of data producers are not always familiar with relevant repositories and the conditions that they impose on deposition, such as particular data or metadata formats. However, in the federated landscape researchers can benefit from increased data curation knowledge at their front office. Front office staff, in turn, can rely on the expertise and the pertaining services of the back office.

Furthermore, researchers gain time when the front office helps them to engage in Research Data Management at an early stage, e.g. by training, by promoting the local data policy, by advice about a data management plan that funders ask for and by offering secure storage facilities for research projects. The respective advantages of the model for researchers, data supporters in the front office and service providers in the

---

1   Horizon 2020: http://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access

2   National Science Foundation: http://www.nsf.gov/bfa/dias/policy/dmp.jsp

3   Netherlands Organisation for Scientific Research (NWO): http://www.nwo.nl/en/news-and-events/dossiers/datamanagement

4   The Netherlands Organisation for Health Research and Development (ZonMw): http://www.zonmw.nl/en/programmes/data-management/programme/

back office are described in more detail in (Dillo et al., 2014; RDNL, 2014; Schoots et al., 2013).

The training *Essentials 4 Data Support* in particular conveys expertise from back office organisations to data supporters in the front office, and vice versa. The data supporters inform the back office about researchers' needs and other developments with respect to research data in their organisations. The training is one of the services provided by RDNL to front offices, along with services provided by individual RDNL members. It should be noted that RDNL is not a legal entity, but a partnership, in which differences exist, for instance in the service fee structures of the three partners. In a similar vein, front offices across the various research organisations may differ in ambition and capabilities with respect to Research Data Management. This diversity explains why the balance in tasks and responsibilities between front offices and back offices is not fixed and most likely will never be fixed at all. Grootveld (2014) describes different scenarios that are currently being explored with individual front offices.

Redesigning the Course

So far two drivers for changing the course were mentioned: the collaboration within Research Data Netherlands and the federation of data responsibilities that RDNL is promoting. Our experiences with the initial course forms the third driver. In 2013, after three rounds, 3TU.Datacentrum and DANS invited Marina Noordegraaf (Verbeeldingskr8), who had designed the web content for the initial course, to help to revise the course. In general the learning goals and envisioned skills (see Grim et al., 2011; Smaele et al., 2013) were considered to be still relevant, but it was time to adapt the course to the feedback from students and coaches, to reflect the changes in the professional setting of the target group and to increase the brand awareness of Research Data Netherlands.

Important elements in the participants' feedback were (see also Verbakel, 2014):

- Participants enjoyed and appreciated their discussions resulting from the homework assignments, which were seen as the most valuable element of the course;

- Participants were interested to hear from researchers themselves in terms of how they deal with data management issues, and about differences between disciplines;

- Participants missed the opportunity to practice writing an actual data management plan;

- The participants appreciated the images included in the course material on the website, which they thought were a memorable way to clarify concepts.

The participants and the coaches agreed that four days of face-to-face tuition is a very substantial time investment. It was decided to reduce face-to-face contact to two – quite intensive – days, which mark the beginning and the end of the course. In the meantime participants should study the online content and make the corresponding assignments individually or with fellow students. Given that all content had to be updated and extended anyway, RDNL decided to provide an online course alongside the blended training, to accommodate a larger audience. The course website was to be shifted from the 3TU domain to RDNL's website.

Furthermore, 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.' New in this mission are the notion of coordination – an important skill – and the definition of data supporter. The latter should make it clear that, although the core service of DANS and 3TU.Datacentrum is to provide long-term archiving services (i.e. after the end of a research project), the course also helps participants to support researchers before and during their projects.

The Current Version

With Noordegraaf as project manager and main developer, the website[5] has been redesigned. The platform for the course is T3Elearning[6], a learning content management framework based on the Typo3 CMS. The content has been revised, with valuable input from subject matter experts within and outside RDNL, several of whom also give a guest lecture at one of the meetings.

Figure 2. User interface of *Essentials 4 Data Support*, Chapter II.

Whereas the old training featured chapters like "Technical skills" and "Advisory skills", 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 2, 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. The complete set of learning objectives can be found in Figure 3 at the end of this paper.

Because of the positive feedback on the images, it was decided to give them an even larger role in a new design. Completely new are audio visual elements, for instance brief talks by researchers and how-to videos about data management planning, data sharing concerns, reproducible research, data citation and persistent identifiers. Also new are the end-of-chapter quizzes, which allow for a quick recapitulation of the chapter's contents. The weekly assignments have been updated as well. Most importantly and meeting earlier participants' wishes, the students now have to write a data management plan for a fictitious research project. To stimulate the exchange of knowledge and ideas between them, they must work in pairs. Furthermore, in the middle of the course the students must reflect on their learning goals and what (or who!) they need to achieve these. Other assignments deal with organising data and with advising a researcher in a fictitious IPR dispute. The assignments and quizzes, as well as the progress bar, are only available to course participants who have enrolled in the full course. Other users won't have these items in the left-hand menu in Figure 2.

Other interactive elements that are now integrated in the website are, first, a forum that the students and the coaches must use between the face-to-face meetings for assignments and commenting on each other's assignments, and which remains accessible afterwards for future knowledge sharing; and second, the option to comment on and make suggestions regarding the course content.

Already in 2011 the course materials where publicly available at the website, both in English and Dutch. On top of that, 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, create a data management plan, and reflect and comment on their own work and that of others. 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.

Conclusions and Outlook

Three groups of students (33 people in total) have currently participated in the revised blended training and their ratings are good: the average is nearly eight, on a scale from one (very bad) to ten (excellent). The decision to add an assignment about data management planning is indeed an improvement: choosing a template and writing such a plan, together with a fellow student, and subsequently commenting on the alternatives provided by the other pairs, is very positively regarded. The networking possibilities were highly appreciated; especially for those participants who only recently became a data supporter, the fellow students form a substantial part of their external network. Furthermore, the participants appreciate the guest lectures, the videos and the online self-tests. On the negative side, several participants reported that it was cumbersome to use the forum for making and commenting on the online assignments. However, overall

introducing the participants to the online forum was much easier and took less time than introducing the combination of Google+ and Google Docs used to be in earlier trainings.

With respect to the goals and the content of the training it remains important to manage expectations, even during a course. As the name entails *Essentials 4 Data Support* has been designed as a generic course at an introductory level and it cannot answer all questions about research data that participants may have or that arise during the course. When asked "What else do you need now to reach your learning goals?" many of them answer "practical experience". Although a two-day course cannot provide this, the practical data management planning assignment as well as the guest lecturers, who represent different disciplines, serve to meet this need. We conclude from the reflections and feedback that the course helps the students to articulate their questions better, which should enable them to engage researchers in "Tell me how I could help you" conversations.

Some universities in The Netherlands have inquired about in house trainings for their (prospective) front office staff. Although this is certainly appealing, we have decided against this because we value the networking possibilities of a mixed group of participants even higher than the intended team building effect.

Furthermore, both in the early stages of the initial course and recently, a few training requests from abroad were received. However, this interest is not substantial enough to offer the full course in English. It is assumed that the two online versions will meet the demands of foreign data supporters. Since early 2014 nearly 40 people from various countries have registered for the Online+ course.

Two aftercare and networking interventions are now being introduced. First, we will contact the students three months after the second face-to-face meeting and ask them to tell us how their personal learning goal has progressed. This "friendly reminder" and ensuing comments at the group's private forum might help them to stay on track. Both the private and the public forum remain accessible anyway. Second, a survey among previous participants will collect information about their current Research Data Management support activities and solicit suggestions that help to keep the training relevant for new generations of data supporters.

Learning objectives of *Essentials 4 Data Support*: "After Chapter X, you will know/ be able to…"

1. Chapter "Definitions":
    *   what information is considered research data;
    *   what open data is, and the various stages of openness;
    *   explain that the meaning of research data varies per target group and per phase of the research lifecycle;
    *   the commonly used definitions in the data supporter's field of work;
    *   sources of information to help you become up-to-date in the field.

2. Chapter "Planning phase":
    *   identify the elements discussed in a research proposal;
    *   how the principles of scientific integrity encourage reproducible research;
    *   the advantages of a systematic approach to research data in view of reuse;
    *   the different parts of a data management plan.

3. Chapter "Research phase":
    *   various ways to store, backup, organise and document research data;
    *   indicate which data formats are suitable for long-term storage;
    *   define the pros and cons of the various storage strategies;
    *   familiarity with a number of datalabs; define datalab characteristics.

4. Chapter "User phase":
    *   what a data archive is; how to archive data and find data sets;
    *   selection criteria that determine whether a data set qualifies to be included in a data archive;
    *   identify the significance of certification for data archives;
    *   identify various types of data publication;
    *   data citation and its advantages; Digital Object Identifiers.

5. Chapter "Data, legislation and policy":
    *   the laws involved in handling research data;
    *   advise researchers in balancing legislation and practice;
    *   different licence types for sharing research data.

6. Chapter "Data support":
    *   awareness of the influence spheres of a data supporter;
    *   engage as a data supporter in a discussion with your client;
    *   explain how the RDNL front office - back office model may serve as one of the ways to organise data support.

Figure 3. Learning objectives of *Essentials 4 Data Support.*

References

de Smaele, M., Verbakel, E., Potters, N., & Noordegraaf, M. (2013). Data intelligence training for library staff. *International Journal of Digital Curation, 8*(1), 218–228. doi:10.2218/ijdc.v8i1.255

Dillo, I., & 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

European Union. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data. Retrieved from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

Grim, R., Van der Heijden, M., De Smaele, M., & Verbakel, E. (2011). *Witboek dataprofessionals in Nederland. White Paper Data Professionals in The Netherlands*. Retrieved from https://www.surf.nl/kennis-en-innovatie/kennisbank/2011/whitepaper-witboek-dataprofessionals-in-nederland.html

Grootveld, M. (2014). *Data management planning in a federated landscape*. Presentation given at the 2nd Data Management Workshop, Cologne. Retrieved from http://www.tr32db.uni-koeln.de/workshops/program.php?wsID=4

RDA Europe. (2014). *The data harvest: How sharing research data can yield knowledge, jobs and growth*. Retrieved from http://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf

Research Data Netherlands. (2014). *A federated data infrastructure for the Netherlands: The front-office–back-office model*. Retrieved from http://www.researchdata.nl/fileadmin/content/RDNL_algemeen/Documenten/RDNL_FOBOmodel-UK-web.pdf

Schoots, F., Verhaar, P., Dillo, I., & de Smaele, M. (2013). *Building a bridge together: Supporting data curation through the front office and back office model*. Poster presented at the 8th International Digital Curation Conference, Amsterdam. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/idcc13posters/Poster200.pdf

Verbakel, E. (2014). *LIBER case study: Essentials 4 Data Support*. Retrieved from http://libereurope.eu/wp-content/uploads/2014/06/LIBER-Case-Study-Essential-4-Data-Support.pdf