



# Research and Innovation Action

## CESSDA Strengthening and Widening

Project Number: 674939      Start Date of Project: 01/08/2015      Duration: 27 months

### Deliverable 5.3: Report on the establishment of an international curriculum for professional development in digital data services for the social sciences

Dissemination Level	PU
Due Date of Deliverable	31/10/17
Actual Submission Date	24/10/2017
Work Package	WP5 - Strengthening and widening through global engagement
Task	5.3
Type	Report
EC Approval Status	16 November 2017
Version	V1.0
Number of Pages	p.1 – p.26
<p><b>Abstract:</b> The current report describes activities taken out in CESSDA-SaW Task 5.3. Its primary aim is the identification of skills and competencies required of staff in data curation. Furthermore, the report gives recommendations for CESSDA on how to strengthen its member organisations by training their staff for current and future challenges of data curation.</p>	
<p>The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.</p>	



## History

Version	Date	Reason	Revised by
0.0	25/09/2017	Review (SND)	Iris Alfredsson
0.1	04/10/2017	Final Version	Sebastian Netscher

## Author List

Organisation	Name	Contact Information
GESIS	Sebastian Netscher	<a href="mailto:sebastian.netscher@gesis.org">sebastian.netscher@gesis.org</a>
GESIS	Anna Schwickerath	<a href="mailto:anna.schwickerath@gesis.org">anna.schwickerath@gesis.org</a>
SND	Birger Jerlehag	<a href="mailto:birger.jerlehag@snd.gu.se">birger.jerlehag@snd.gu.se</a>
SND	Iris Alfredsson	<a href="mailto:iris.alfredsson@snd.gu.se">iris.alfredsson@snd.gu.se</a>
SND	Eira Brandby	<a href="mailto:eira.brandby@snd.gu.se">eira.brandby@snd.gu.se</a>

## Time Schedule before Delivery

Next Action	Deadline	Care of
Review by task partners	25/09/2017	SND
Review by the WP leader	04/10/2017	GESIS
Review by the Chair of the Delivery Committee	23/10/2017	CSDA
Review by the Project Coordinator	23/10/2017	CESSDA
Approval and Submission by the Project Coordinator to the European Commission	24/10/2017	CESSDA

## EXECUTIVE SUMMARY

As described in the Grant Agreement of CESSDA-SaW, task 5.3 set out “to explore the value of, and appetite for, an internationally recognised and moderated professional qualification in digital data services for the social sciences” (GA, ANNEX, pp35). Task members therefore aimed at identifying most relevant elements of professions in data curation. The specific objective of the task was to contribute to a deeper understanding of all aspects to be recognised in the training of current and future employees of CESSDA organisations.

As data curation activities span from more generic tasks performed in any data repository to highly specialized ones tailored for social science data repositories this report mainly deals with the latter. This is an area where CESSDA, with assistance from its Service Providers, can provide support and training for staff both at aspiring and existing member organizations.

## ABBREVIATIONS AND ACRONYMS

ACM	Association for Computing Machinery
CDM	Capability Development Model
CESSDA	Consortium of European Social Science Data Archives
CF-DS	The Data Science Competence Framework
CLARIN	Common Language Resources and Technology Infrastructure
CPA	Capability Process Area
CRA	Capability Requirement Area
CWA	CEN Workshop Agreement
DDI	Data Documentation Initiative
DigCCurr	Digital Curation Curriculum
DigCurV	Curriculum Framework for Digital Curation
DMP	Data Management Plan
e-CF	e-Competence Framework
EDISON	Education for Data Intensive Science to Open New science frontiers
EOSC	European Open Science Cloud
ESCO	European Skills, Competences, Qualifications and Occupations
ESFRI	European Strategy Forum on Research Infrastructures
ETHRD	The Education & Training on Handling of Research Data Interest Group
GA	Grant Agreement
GESIS	GESIS - Leibniz Institute for the Social Sciences
HLEG	High Level Expert Group
IASSIST	International Association for Social Science Information Services & Technology
ICT	Information and Communications Technology
ICPSR	Inter-University Consortium for Political and Social Research
IPSDS	International Program in Survey and Data Science
NBDIF	National Big Data Interoperability Framework
NIST	National Institute of Standards and Technologies of USA
RDA	Research Data Alliance
SIMS	SND's Internal Management System
SND	Swedish National Data Service

## TABLE OF CONTENTS

1.	INTRODUCTION	6
2.	SKILLS AND COMPETENCIES	7
2.1.	EDISON – Data Scientist Competences and Skills Framework	7
2.2.	DigCurV - CURRICULUM FRAMEWORK FOR DIGITAL CURATION	9
2.3.	RDA – Research Data Alliance	10
2.4.	CESSDA-SaW WP3 - Heuristic Development Model	11
3.	JOB DESCRIPTIONS AS AN INDICATOR FOR SKILLS REQUIRED IN DATA CURATION	13
4.	STUDY PROGRAMMES	17
5.	BEST-PRACTICE	19
6.	CONCLUSION AND RECOMMENDATIONS	23
	REFERENCES	25

## FIGURES, TABLES, ETC.

Figure 1:	Data Science definition by NIST	8
Figure 2:	The Curriculum Framework by DigCurV	9
Figure 3:	Illustration of the multi-skilled	10
Table 1:	Selected study programmes	18
Table 2:	Example of introductory schedule for data manager	21

# 1. INTRODUCTION

Traditionally most social science data archives have been founded by scientists and their institutions, working with as well as creating research data. Being aware of the relevance to store and share data, such data archives were often hosted by universities and research institutions, staffed with researchers and students somewhat proficient in handling research data. Working methods and routines were learned and implemented by adopting best practices and guidelines from already existing Data Archives, often by participating in CESSDA Expert Seminars.

The situation today is quite different, driven by e.g. more complex data types and formats, demands from a multitude of stakeholders such as funders, publishers, and of course scientists. Such demands include trustworthiness/certification, quality/FAIR data, protection of integrity/legal framework and so on. To meet these demands new skills and competences are needed at the data archives and therefore required from their staff.

Task 5.3 of the CESSDA-SaW project (*Strengthening and Widening the European Infrastructure for Social Science*) sets out “to explore the value of, and appetite for, an internationally recognised and moderated professional qualification in digital data services for the social sciences” (GA, ANNEX, pp35). In order to identify the most important elements of data curation, task members examined different fields with the aim of providing a complete picture of all aspects relevant for the development of a curriculum for the training of employees in long-term preservation:

- At a theoretical level, so-called competency models with regard to data curation were examined. The idea behind this was to identify different forms of professional activities relevant in this context.
- At a practical level, partners of task 5.3 explored descriptions of skills necessary in data curation by analysing existing study programmes as well as job descriptions published by social science data archives, universities, and their libraries. Furthermore, best-practise examples from CESSDA member archives were explored.

In sum, task 5.3 aimed at collecting all information necessary to identify the most relevant fields to be included in a curriculum for social science data archive employees. The partners furthermore prepared a list of recommendations with the aim of fostering the efforts made by CESSDA and its member archives in the process of training their current and future employees.

## 2. SKILLS AND COMPETENCIES

The challenge of supplying data repositories with skilled staff is mainly due to that there are no designated occupations or educations in this area. One of the implications this has led to is that repositories have to use resources on in-house training of new staff, resources supposed to be used for data curation. To remedy this situation there are quite a few initiatives taken that will produce a set of desired skills and competences for staff working with digital curation. These skills and competences are of a generic type and will fit any type of digital repository regardless of disciplinary content. The outcomes of these projects are nevertheless highly interesting for CESSDA as they give the basic foundation where more domain-specific social science qualifications can be added on top.

This chapter will take a closer look at three of these initiatives that are of particular interest for CESSDA: the EDISON - Data Scientist Competences and Skills Framework; the DigCurV - Curriculum Framework for Digital Curation; and the Research Data Alliance – The Education & Training on Handling of Research Data Interest Group project. In addition, it will give an introduction to the CESSDA Heuristic Development Model, developed by WP3 of the CESSDA-SaW project, and thereby paint a broad picture of recent approaches in the area made with the aim of identifying the most relevant skills and competencies for data curation.

- **EDISON (Education for Data Intensive Science to Open New science frontiers) – Data Scientist Competences and Skills Framework**

EDISON aims to establish the so-called data scientist as a profession. To do this they will match available career paths in industry and science with existing curricula with respect to required expertise and professional certification.

- **DigCurV - Curriculum Framework for Digital Curation**

DigCurV addressed the availability of vocational training for digital curators in the library, archive, museum and cultural heritage sectors needed to develop new skills that are essential for the long-term management of digital collections.

- **Research Data Alliance (RDA) – The Education & Training on Handling of Research Data Interest Group (ETHRD)**

The ETHRD has been active since mid-2014 and consists, among others, of several individuals active in the EDISON and DigCurV projects. This makes it an excellent forum for exchange of news and info between all stakeholders in the education and training field.

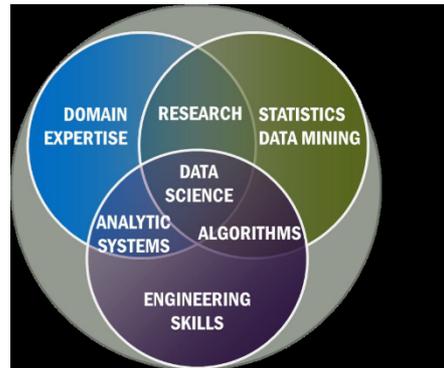
### 2.1. EDISON – DATA SCIENTIST COMPETENCES AND SKILLS FRAMEWORK

EDISON is a 2-year project (2015-2017) with the purpose of accelerating the creation of the Data Science profession suitable for European research and industry. “A data scientist is a practitioner who has sufficient knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills, and software and systems engineering to manage the end-to-end data processes in the data life cycle.”<sup>1</sup> To do this they have created the Data Science Competence Framework (CF-DS) as a cornerstone of the EDISON Data Science Framework.

---

1 [\(EDISON Framework. P8\).](#)

Figure 1: Data Science definition by NIST (National Institute of Standards and Technologies of USA)



EDISON names five so-called competence areas and thereby tasks in this field<sup>2</sup>

- PLAN: business strategy, service management, architecture & application design, technology monitoring, innovation
- BUILD: application development, component integration, testing, solution deployment, documentation production, systems engineering
- RUN: user & change support, service delivery, problem management
- ENABLE: information security strategy, ict quality strategy development, education & training provision, purchasing, sales proposal development, channel, sales, contract, and information & knowledge management, personnel development, needs identification, digital marketing
- MANAGE: forecast development, process improvement, is governance, management: project & portfolio, risks, relationships, ict quality, business change, and information security

The Data Science Competence Framework (CF-DS) builds on the work of several existing frameworks and studies from a wide range of bodies both private and official. These are:

- NIST NBDIF Data Science and Data Scientist definition<sup>3</sup>
- European e-Competence Framework (e-CFv3.0)<sup>4</sup>
- European ICT profiles CWA 16458 (2012)<sup>5</sup>
- European Skills, Competences, Qualifications and Occupations (ESCO)<sup>6</sup>
- ACM Computing Classification System (ACM CCS2012)<sup>7</sup>
- Analysing the Analysers, O'Reilly Strata Survey (2013)<sup>8</sup>
- EC Report on the Consultation Workshop (May 2012) "Skills and Human Resources for e-Infrastructures within Horizon 2020"<sup>9</sup>.

2 [http://edison-project.eu/sites/edison-project.eu/files/filefield\\_paths/edison\\_cf-ds-release1-v07.pdf](http://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_cf-ds-release1-v07.pdf), see page 10, accessed 28 Sept 2017.

3 <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf>, accessed 28 Sept 2017.

4 <http://www.ecompetences.eu/get-the-e-cf/>, accessed 28 Sept 2017.

5 <http://www.ecompetences.eu/ict-professional-profiles/>, accessed 28 Sept 2017.

6 <https://ec.europa.eu/esco/portal/home>, accessed 28 Sept 2017.

7 <http://www.acm.org/about/class/class/2012>, accessed 28 Sept 2017.

8 [http://cdn.oreillystatic.com/oreilly/radarreport/0636920029014/Analyzing\\_the\\_Analyzers.pdf](http://cdn.oreillystatic.com/oreilly/radarreport/0636920029014/Analyzing_the_Analyzers.pdf), accessed 28 Sept 2017.

By analysing these frameworks and studies, EDISON defines five different data science competence groups (p21):

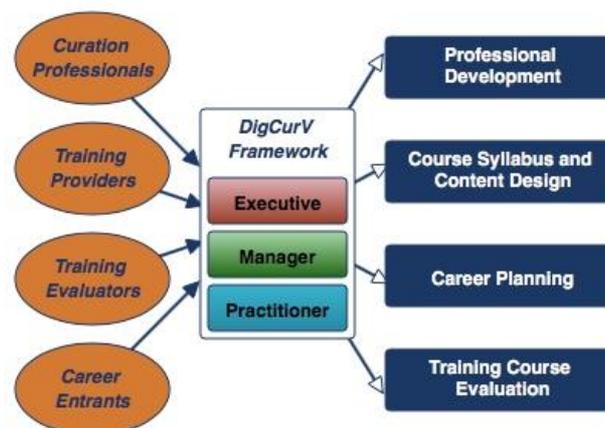
- Data Analytics including statistical methods, Machine Learning and Business Analytics
- Data Management, Curation, Preservation
- Engineering: software and infrastructure
- Scientific or Research Methods for research related professions and Business Process Management for business related professions
- Subject/Scientific Domain competences and knowledge.

EDISON has also compiled an extensive list of Universities worldwide where programmes relating to Data Science are arranged<sup>10</sup>.

## 2.2. DIGCURV - CURRICULUM FRAMEWORK FOR DIGITAL CURATION

The DigCurV project (2011-2013) framework is research based. “In particular, the DigCurV team has undertaken multinational research to understand both the skills currently used by those working in digital curation in the Cultural Heritage sector, and the skills sought by employers in this sector”.

Figure 2: The Curriculum Framework by DigCurV



The framework differentiates three views: executive, manager, and practitioner plus a comparing one. Furthermore, it defines four skill categories where each can be allocated to different levels, namely basic, intermediate and advanced. The four skill categories are

- management & quality assurance
- knowledge & intellectual abilities
- professional conduct
- personal qualities

9 [https://www.innovationpolicyplatform.org/system/files/European Commission 2012-Skills%20and%20Human%20Resources%20for%20e-Infrastructures%20within%20Horizon%202020\\_5.pdf](https://www.innovationpolicyplatform.org/system/files/European_Commission_2012-Skills%20and%20Human%20Resources%20for%20e-Infrastructures%20within%20Horizon%202020_5.pdf), accessed 28 Sept 2017.

10 The list of study programmes is available online: <http://edison-project.eu/university-programs-list>.

Figure 3: Illustration of the multi-skilled digital curator according to DigCurV.



These categories include a number of sub-skills<sup>11</sup>, which will help users of the framework to navigate across a wide range of skills.

The idea behind the framework is the recognition that digital curation is a complex profession. Staff must demonstrate both subject-specific skills such as understanding digital preservation technologies, legal frameworks and metadata standards and generic professional/project skills, for example, being able to plan tasks, communicate with diverse groups, and perform risk assessments.

The most relevant 'lens' to use for this task is the 'Practitioner Lens'<sup>12</sup> where it is easy to recognize a Data Archive employee.

### 2.3. RDA – RESEARCH DATA ALLIANCE

The Education & Training on Handling of Research Data interest group at RDA has compiled a list of essential and desired skills and competences for four different groups of professions<sup>13</sup>:

- research librarians;
- research administrators;
- research infrastructure managers / operators;
- researchers.

Besides the list, there are a multitude of references and links to relevant documents and sites where further information can be obtained.

<sup>11</sup> <http://www.digcurv.gla.ac.uk/skills.html>, accessed 28 Sept 2017.

<sup>12</sup> [http://www.digcurv.gla.ac.uk/assets/downloads/dcsf\\_practitioner\\_lens.pdf](http://www.digcurv.gla.ac.uk/assets/downloads/dcsf_practitioner_lens.pdf), accessed 28 Sept 2017.

<sup>13</sup> <https://www.rd-alliance.org/group/education-and-training-handling-research-data-ig/wiki/task-force-defining-data-handling>, accessed 28 Sept 2017.

## 2.4. CESSDA-SAW WP3 - HEURISTIC DEVELOPMENT MODEL

The CESSDA-SaW Heuristic Development Model (CESSDA-CDM)<sup>14</sup> names three so called Capability Requirement Areas: *organizational infrastructure*, *digital object management*, and *technical infrastructure*, which are subdivided into the following *Capability Process Areas*.

- **CRA1 Organisational Infrastructure:**
  1. Mission and Scope
  2. Contracts, Licenses and Liabilities
  3. Funding, Staff, Resources
  4. Outreach and Communication
  5. Confidentiality, Ethics and Disclosure risk
  6. Documentation
  7. Management Oversight
- **CRA2 Digital Object Management:**
  1. Data Acquisition and Ingest
  2. Data Preservation: storage, curation and planning
  3. Access / Provision
- **CRA3 Technical Infrastructure:**
  1. Risk Assessment
  2. Technical Planning and Management
  3. Technical Resilience Infrastructure
  4. Technical Resilience Security
  5. Technical Resilience Disaster planning

At first glance, the requirements listed under *Digital Object Management* appear to be genuinely associated to digital preservation, whereas other tasks require skills not necessarily linked to data archiving. This aspect marks two general categories of requirements/skills. Secondly, the skills mentioned are not easily to differentiate; e.g. skills necessary to work in “outreach and communication” will be needed as well in the area “data acquisition and ingest”. The model looks at the requirements/skills as functions at the repository level, so they can be facilitated by automated workflows for instance in the form of web forms for authentication and ingest.

The tasks/requirements listed in the model was used to measure the maturity level of the different Service Providers, so they are considered important parts of repository activities. This means that they can serve as pointers to what necessary skills and competences are needed amongst the staff of a CESSDA Data Archive. In this part, different criteria from Capability Process Area 2 (see above) will be listed to show which necessary tasks and functions can serve as a guideline to create a curriculum for training of CESSDA staff.

### **Data Acquisition and Ingest**

Purpose - Decide what to preserve; Plan and execute the selection, acquisition and transfer of information products to the archive.

- Acquisition and selection of data

---

<sup>14</sup> See <https://www.cessda.eu/eng/Projects/All-projects/CESSDA-SaW/WP3/CESSDA-CDM> for further information on the model. Accessed 28 Sept 2017.

- Documentation / Metadata requirements
- Collection policy
- Systems for submission/ingest
- Authentication and authorisation
- Provenance information
- Citations
- Conditions placed on content, deposit licenses
- Legal transfer of custody; agreements on rights/responsibilities
- Completeness and correctness
- Authenticity checks of deposited material
- Quality control standards and reporting mechanisms

### **Data Preservation: storage, curation and planning**

Purpose - Provide the services and functions for the storage, maintenance and retrieval of data; Provide services, functions, recommendations and preservation plans to ensure that the information stored in the archive remains accessible to, and understandable by, the Designated Community over the long term.

- Transfer to permanent storage
- Persistent identifiers (PIDs) / locators
- Backup and version control/change procedures
- Authentication measures
- Fixity checks
- Error detection / unwanted changes
- Metadata management
- Preservation policy
- Evaluation of content and preservation environment
- Monitor technology
- Preservation strategies

### **Access /Provision**

Purpose - Provide the services and functions that support the users in determining the existence, description, location and availability of information stored in the archive, and allowing users to request and receive data.

- Access interfaces
- Searchable and indexed content
- Downloadable data holdings
- Metadata harvesting
- Data formats
- Metadata standards
- Access policy
- Failures, anomalies, system failures
- Persistent identifiers (PIDs) / locators
- Authentication and authorization

The aim of this section is to introduce existing models with regard to the identification of skills and competencies relevant for data curation in general and the social sciences in particular. The

tasks and requirements listed in the model can loosely be mapped to the skills listed in the DigCurV Practitioner Lens<sup>15</sup> which can serve as a guideline to educational activities.

### 3. JOB DESCRIPTIONS AS AN INDICATOR FOR SKILLS REQUIRED IN DATA CURATION

In order to identify skills and competencies necessary for working in data curation, it is important to consider the archives' view, i.e. finding out which skills and competences they require from their staff. One way of assessing these requirements is to evaluate job advertisements archives and research institutions publish in the process of hiring new staff, as they “provide a window to the marketplace” (Kennan et al., 2006:190).

The idea behind this approach is to consider “job ads as relatively accessible indicators of the knowledge, skills and competencies required [...] by employers” (Kennan et al., 2006:179). As “[t]he third task [was] to explore the value of, and appetite for, an internationally recognised and moderated professional qualification in digital data services for the social sciences” (GA, pp35), it seemed appropriate to consult job descriptions in order to gain knowledge on skills and competences needed in data curation. This section therefore describes the skills and competences found in the respective literature for data curation in general, as well as for digital preservation in the field of social sciences. After having described several competency models above, Task 5.3 partners decided to include this research with the aim of bringing together theoretical and practical approaches in order to reach a complete picture of requirements for staff in digital curation.

In recent years, this method has been applied several times resulting in a body of literature on the subject. For example, Choi and Rasmussen examined

“qualifications and skills required of professional positions involved in digital resources, services, and technologies as changing aspects in academic libraries. Data was collected from job advertisements for digital library positions posted in College and Research Libraries News during the nine years from 1999 to 2007” (Choi/Rasmussen, 2009:457).

They conclude “that current awareness and appropriate technological skills and experience in the digital library environment, knowledge and experience in creation and management of digital information, and metadata are the most required qualifications for digital librarian positions with high emphasis on management skills” (Choi/Rasmussen, 2009: 465). This approach is in line with the CESSDA Heuristic Development Model, as it defines the *Organisational Infrastructure*, *Digital Object Management*, and *Technical Infrastructure* as its main so-called Capability Process Areas. Furthermore, it is notable that the literature distinguishes a number of job titles for employees working in data curation. Swan and Brown name *data creator*, *data scientist*, *data manager*, and *data librarian* as the most relevant professions in this

---

15 <http://www.digcurv.gla.ac.uk/practitionerLens.html>, accessed 28 Sept 2017.

[http://www.digcurv.gla.ac.uk/assets/downloads/dcsf\\_practitioner\\_lens.pdf](http://www.digcurv.gla.ac.uk/assets/downloads/dcsf_practitioner_lens.pdf), accessed 28 Sept 2017.

field. They state, however, that “there is not yet an exact use of such terms in the data community, and the demarcation between roles may be blurred. It will take time for a clear terminology to become general currency” (Swan/Brown, 2008:1). As mentioned above, for example the RDA uses terms such as *research librarian* or *research infrastructure managers/operators*.

Analysing job advertisements was also a part of the DigCCurr project<sup>16</sup>, during which a number of skills required in data preservation were identified. Lee et al. categorise them in “first-level functions and skills” and “meta-level functions and skills”. The first category contains a large number of different tasks, mostly with regard to the OAIS model, such as “administration” or “preservation planning”. According to the authors, on the meta-level, the following functions are required: “Analysis & Documentation of Curation Functions; Evaluation & Audit of Curation Functions; Research & Development to Support Curation Functions” (Lee et al., 2007:50).

A distinction between technical and non-technical skills was found in most of the respective literature; in addition, several other skills not necessarily specific for data curation seem to play an important role in the field. For example, Choi and Rasmussen state, “that the demand for management skills, interpersonal skills, project management skills, and teamwork skills is emerging for digital librarianship” (Choi/Rasmussen, 2009: 465). For example, as funders and publishers increasingly demand access to data from researchers, service aspects of professions in digital preservation are becoming more important as well.

Kim et al. emphasise the need for staff in digital curation “to stay updated on new technologies, equipment and processes” (Kim et al., 2013:76). This must be considered as a challenge, especially as it requires financial support from the archives, enabling their staff to participate in trainings on a regular basis, which might not be easily possible for each organisation. Technologies and processes do furthermore vary with regard to the customs of the respective research disciplines. “The appearance of required domain knowledge in specialized academic disciplines, including science, engineering and history, indicates that digital curation professionals in the field may require certain domain knowledge dependent on the types of resources or data being managed” (Kim et al., 2013:77). This finding is especially relevant for the current report, as Task 5.3 focussed on identifying skills and competences particularly important in the social science context.

With regard to social sciences, Xia and Wang (2014) conducted a content analysis “of 167 job postings in 2005-2012 from the International Association for Social Science Information Services & Technology website”, discovering “a steady increase of demands for professionals” in the field (Xia/Wang, 2014:362). “These individuals must possess necessary attitudes toward data and be able to handle the intricate process of digital scientific data, along with the increasingly hot discussion of a big-data challenge both within and outside of academic settings” (Xia/Wang, 2014:362). Interestingly, one result of the study shows that employers consider non-technical skills as equally important as technical skills. (Xia/Wang, 2014:385) The

---

<sup>16</sup> “The DigCCurr (Digital Curation Curriculum) project is developing a graduate-level curricular framework, course modules, and experiential components to prepare students for digital curation in various environments” (Lee et al., 2007:49). For further information please visit: <https://ils.unc.edu/digccurr/>.

authors furthermore stress the importance of considering the data life cycle as relevant for the identification of the employees' tasks:

“As a service provider, the social science librarian needs to work closely with different units of a library or data centre, various departments of an institution, and diverse groups of organizations at local, national, and international levels. S/he will cooperate with scholars to support their research and instructional endeavours; with data providers, internal as well as external, of an institution, to acquire and prepare data sets; with technology personnel to convert, preserve, and curate digital data; with team members to ensure high standards of raw data discovery, access, and retrieval; and with experts to assist in scientific data manipulation, analysis, and visualization” (Xia/Wang, 2014:383).

In conclusion, job advertisements analysed in order to identify most relevant areas of activities in digital curation included a number of categories that were found by most researchers in this field, such as the following example from the DigCurV project<sup>17</sup>. “The majority of the tasks listed in the job descriptions collected fell into the following categories:

- Communications, outreach and liaison;
- Digital collection and data management;
- Digital curation and preservation;
- Project management;
- Trusted repository and archive;
- Resource, teaching and training;
- Technical;
- Trends, best practice and documentation;
- Supervisory [...]” (Engelhardt et al., 2012:49).

The literature generally distinguishes these main fields of expertise: organizational or administrative tasks, data management tasks, and technical tasks. In addition, personal competencies, such as communication skills, are needed. These fields are reflected in the CESSDA Heuristic Development Model and its three capability requirement areas: organizational infrastructure, digital object management, and technical infrastructure; they could also be identified in the skill categories mentioned in the DigCurV project (see description for both models above).

“Given this context, articulating the specific competencies, such as the knowledge, skills and abilities required to perform a broad range of digital curation functions, is an excellent basis for designing an educational and professional framework to train digital curators. (Kim et al., 2013:68)

From December 2016 to March 2017, Task 5.3 partners collected 37 job advertisements<sup>18</sup> that fitted the above-mentioned criteria (job descriptions). These job advertisements were published by universities, university libraries, other research institutions as well as data archives. By analysing them, the findings discussed above could be confirmed. More than half of

---

17 “From February 2011 to January 2012, DigCurV collected forty-eight job advertisements of posts related to the area of digital preservation and curation” (Engelhardt et al., 2012:49). For further information on the project please visit: <http://www.digcur-education.org/>, accessed 28 Sept 2017.

18 The job advertisements were collected from the JISC Digital Preservation Mailing-list: <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=DIGITAL-PRESERVATION>, accessed 28 Sept 2017.

the job advertisements emphasized the need for future employees to have strong communication skills. Also, methodological skills were required, such as experiences with statistical software. Contrary to expectations, only one of the job advertisements mentioned knowledge of the *data life cycle* as important. Besides technological skills, experiences in data curation and management were named frequently as key competences.

However, a greater variety with regard to different job titles could be observed. In addition to terms commonly used, such as *digital librarian*, *curator*, or *data scientist*, the following names were used: *research data coordinator/manager*, *digital preservation specialist*, *research outputs manager*, or *digital preservation process administrator*; some even included terms used in the OAIS model (*data ingest officer*). It can be expected that the process of hiring new staff might be simplified by using those terms, as they are more focused on particular aspects of the jobs' tasks and thereby contribute to a better understanding of its requirements.

In general, the above-mentioned fields of expertise could be identified in job postings collected for this report. As mentioned above, task 5.3 aimed at exploring the demand for a professional qualification in data curation, which in general could be identified in the analysis of job descriptions. However, skills and competences named in this context came from a broad range of different professions. It is therefore questionable, whether they could be reflected in a single curriculum. For example, staff whose main task is to provide support for researchers in the process of data collection are required to have different skills from those who are working on finding technical solutions for long-term preservation.

By asking colleagues from CESSDA member organisations about common practices of hiring staff in their institutions, Task 5.3 members aimed to gain a better understanding of the challenges archives face in the process. Most of the answers indicated that human resources departments use templates consisting of a number of different skills and competences, which are complemented by specific requirements for the respective position<sup>19</sup>. Archives, not using such templates, showed interest in receiving information on them.

Colleagues furthermore stated that new employees receive additional in-house training in the beginning. This is of particular importance because these in-house trainings might be transferrable to a broader audience, e.g. by sharing experiences between CESSDA members, training materials could benefit new employees in other archives. Chapter 5 therefore includes a description of an example from SND, highlighting key topics for the training of new employees. Knowledge transfer between CESSDA archives could thus be fostered by establishing exchange programmes between employees of member organisations with the aim of getting them acquainted with different approaches of data curation.

---

19 As these templates are mostly provided in the language of the respective country, they could not be used for this report.

## 4. STUDY PROGRAMMES

Task 5.3 furthermore looked at existing study programmes with regard to digital curation. Table 1 shows selected study programmes relevant for this report as they either include aspects of digital curation in general, or offering courses in more specific areas, such as methods of data analysis. This is not an exhaustive list. However, it provides insights into the landscape of study programmes relevant in this field.

Most of these programmes target students who already have a degree, e.g. “Archives and Records Management” at the University of Liverpool. Very few examples could be found for programmes focussing on archiving or preservation in general, such as “Information Management & Preservation (Digital)/(Archives & Records Management)” at the University of Glasgow.

As mentioned earlier in this report, employees in digital curation usually need to have methodological knowledge in the particular research area they are working in. Therefore, most of the Master’s programmes require a Bachelor’s degree in a related field. The “Master Digital Curation”, jointly offered by the King’s College in London and the Humboldt University Berlin, includes methodological aspects as well as a number of so-called optional modules, giving students the opportunity to focus on different topics, such as *open research* or *digital media*.

For the CESSDA context, these study programmes are most important, as they are offering different opportunities for collaboration with regard to potential target groups: employees of CESSDA member organisations and future staff. Employees of CESSDA member organisations could benefit from the programmes as many of the modules could be used as advanced training for their professional development. These programmes furthermore provide the opportunity to recruit new staff, for example by establishing internship programmes between universities and archives. As many CESSDA member organisations offer trainings as well, students could benefit from collaborations. Ensuring “excellent education for members of staff of existing and future data archives” (GA) requires enhancing co-operations between CESSDA and the respective universities. Furthermore, it seems important for CESSDA to foster exchange with universities in order to communicate the (constantly changing) requirements for staff in digital curation to those, who are responsible for the development of study programmes.

Table 1 - selected study programmes

Study program and link to further information
Edison Project: list of study programmes: <a href="http://edison-project.eu/university-programs-list">http://edison-project.eu/university-programs-list</a> .
Cologne Center for eHumanities: list of study programmes in digital humanities at German Universities: <a href="http://www.cceh.uni-koeln.de/dh-degrees-2011">http://www.cceh.uni-koeln.de/dh-degrees-2011</a> .
IPSDS: Survey Data Study programme: <a href="http://survey-data-science.net/program/admission">http://survey-data-science.net/program/admission</a> .
Humboldt University, Berlin & King's College London: Master Digital Curation: <a href="https://www.kcl.ac.uk/study/postgraduate/taught-courses/digital-curation-ma.aspx">https://www.kcl.ac.uk/study/postgraduate/taught-courses/digital-curation-ma.aspx</a> .
University Düsseldorf: Information Science & Linguistic Technology (description only in German): <a href="http://www.uni-duesseldorf.de/home/nc/studium-und-lehre-an-der-hhu/studium/alle-studiengaenge-von-a-z/studiengang-informationen/studiengaenge/informationswissenschaft-und-sprachtechnologie.html">http://www.uni-duesseldorf.de/home/nc/studium-und-lehre-an-der-hhu/studium/alle-studiengaenge-von-a-z/studiengang-informationen/studiengaenge/informationswissenschaft-und-sprachtechnologie.html</a> .
IPSDS: <a href="http://survey-data-science.net/">http://survey-data-science.net/</a> .
Society of American Archivists: <a href="http://www2.archivists.org/prof-education/a-d/exams">http://www2.archivists.org/prof-education/a-d/exams</a> .
University Mannheim: <a href="http://bewerbung.uni-mannheim.de/common/steckbriefIndex.php?abschlfach=88278&amp;submit=Steckbrief+anzeigen">http://bewerbung.uni-mannheim.de/common/steckbriefIndex.php?abschlfach=88278&amp;submit=Steckbrief+anzeigen</a> .
University Stuttgart: <a href="https://www.uni-stuttgart.de/en/study/study-programs/program/Data_Science_B.Sc-00001/">https://www.uni-stuttgart.de/en/study/study-programs/program/Data_Science_B.Sc-00001/</a> .
University Aberystwyth: <a href="https://courses.aber.ac.uk/postgraduate/archive-administration-masters/#key-facts">https://courses.aber.ac.uk/postgraduate/archive-administration-masters/#key-facts</a> .
Northumbria University: <a href="https://www.northumbria.ac.uk/about-us/academic-departments/computer-and-information-sciences/">https://www.northumbria.ac.uk/about-us/academic-departments/computer-and-information-sciences/</a> .
University College Dublin: <a href="https://sisweb.ucd.ie/uisis/!W_HU_MENU.P_PUBLISH?p_tag=PROG&amp;MAJR=T306">https://sisweb.ucd.ie/uisis/!W_HU_MENU.P_PUBLISH?p_tag=PROG&amp;MAJR=T306</a> .
University of Dundee: <a href="http://www.dundee.ac.uk/cais/programmes/postgraduatecourses/archivesrecordsmanagement/">http://www.dundee.ac.uk/cais/programmes/postgraduatecourses/archivesrecordsmanagement/</a> .
Glasgow University: <a href="http://www.gla.ac.uk/postgraduate/taught/informationmanagementpreservationdigitalarchivesrecordsmanagement/#/programstructure">http://www.gla.ac.uk/postgraduate/taught/informationmanagementpreservationdigitalarchivesrecordsmanagement/#/programstructure</a> .
University of Liverpool: <a href="https://www.liverpool.ac.uk/study/postgraduate/taught/archives-and-records-management-ma/overview/">https://www.liverpool.ac.uk/study/postgraduate/taught/archives-and-records-management-ma/overview/</a> .
University College London: <a href="http://www.ucl.ac.uk/prospective-students/graduate/taught/degrees/archives-records-management-ma">http://www.ucl.ac.uk/prospective-students/graduate/taught/degrees/archives-records-management-ma</a> . <a href="http://www.ucl.ac.uk/dis/study/pg/TMAARMSING01B">http://www.ucl.ac.uk/dis/study/pg/TMAARMSING01B</a> .
Charles Sturt University: <a href="http://www.csu.edu.au/courses/information-studies">http://www.csu.edu.au/courses/information-studies</a> .

## 5. BEST-PRACTICE

As stated in chapter 2, the challenge of supplying data repositories with skilled staff is mainly due to the fact that there is presently no designated occupations or educations in this area. New employees at CESSDA member organisations need to be prepared for their functions; therefore, a number of archives provide trainings for their staff. As these trainings might be helpful for other (and specifically new) CESSDA Service Providers, this section describes an example from SND.

The skills and competencies asked for when advertising a position as social science data manager at SND are: academic degree in relevant social science discipline; experience of using data; knowledge of data formats, methods and metadata standards used in social sciences; experience in CMS; working experience from research environments; and knowledge in using statistical tools such as SAS, SPSS, STATA

During the first four weeks of employment a new data manager at SND undergoes an in-house introductory training (see example of an introductory schedule). The aim of the training is to make the new employee acquainted with different approaches of data curation, but also to introduce the employee to all activities at SND.

Organizationally the data manager belongs to the Training and Repository Division, led by a division leader. The division leader will appoint a senior data manager to act as supervisor during the introductory period. Before the start of the employment the division leader and the supervisor prepares the first two months' work tasks, including the processing of one or several studies. A list of recommended literature, web pages and other study objects is put together. The first day of work the new data manager is introduced to the introduction plan by the division leader and supervisor.

The following sessions are recommended to be included in the introductory training for a data manager:

- **Overview of how SND manages data – ingest, processing, dissemination**

This session introduces the OAIS model and provides a brief overview of the OAIS-process at SND. Beginning with data and metadata being submitted via the SND web form, then following how data and metadata are processed to fulfil the requirements for a SND version 1.0 and how all files are stored in a sustainable way. Then an introduction to how data are made searchable in the web catalogue, and finally how data are disseminated.

- **Introduction to the internal data management system SIMS**

SND's internal management system, SIMS, is an important tool for data management at SND. This session provides an introduction to several of the functionalities of SIMS.

- **FAIR data – study level description**

An introduction to the concept FAIR data. How do we work to make data findable, accessible, interoperable and reusable? The benefit of using PIDs and the use of DOI at SND. A review of the metadata submitted with the example study. Mandatory and recommended metadata

elements in SIMS are presented, followed by a discussion on the minimum level of metadata needed for presenting a study in the online catalogue.

- **Why do we need metadata – introduction to metadata standards**

Metadata can be categorized in different types: Descriptive, structural, administrative and paradata. This session will give an overview of the four categories and examples of metadata elements belonging to each category. The benefits of using metadata standards and controlled vocabularies are discussed. A brief introduction of metadata standards and controlled vocabularies used at SND.

- **Kind of data: Numeric**

Most data at SND are numeric, originating from surveys or registers. This session will address different file formats and the most common analysis and documentation programs (for example SPSS and Nesstar, excel and Colectica for Excel), as well as metadata standards suitable for numeric data.

- **Kind of data: Text, audio & video**

This session will address different file formats and the most common analysis and documentation programs, as well as metadata standards suitable for text, audio and video.

- **Kind of data: Spatial & image**

This session will address different file formats and the most common analysis and documentation programs, as well as metadata standards suitable for spatial data and images.

- **SND in a bigger perspective – why data repositories in Social Science**

This session gives an overview of how SND emerged and how the organization can be placed both in a national and a global context. It includes a thorough introduction of how CESSDA has played a central role in the development of the organisation since it started.

- **Introduction to the ethical & legal framework for research data**

An introduction to legal and ethical aspects on data sharing and to national jurisdiction such as data protection, IPR, ownership of research data and legislation on archives.

- **Confidentiality and disclosure risks**

This session deals with confidentiality issues and how to detect disclosure risks.

- **Data Management and Data Management Plans**

How can SND introduce researchers to data management and data management plans? A review of the contents of SND's web guide for data management and checklist for DMP. A review of different concepts, such as data management protocols and active data management plans. An overview of DMP existing online tools.

- **Internal and external training at SND**

This session gives an overview of the training activities, both internal and external.

- **Overview of technical support and development work at SND**

This session gives an overview of the activities at the Division for Technical Support and Development and how this division cooperates with the other divisions.

- **What is a Trusted Digital Repository**

An introduction to the DSA certification process and the different levels of certification as a TDR.

- **National cooperation and projects**

In 2018 SND will become a consortium in which seven universities will cooperate. The consortium will work together with a network consisting of 20+ universities and university colleges. A presentation of the work within the consortium and network, as well as with other national actors.

- **International cooperation and projects**

SND participate in many international organizations and international projects. This session gives a brief presentation of CESSDA, CLARIN, ICPSR, DataCite, DDI Alliance, RDA and IASSIST and an overview of ongoing international projects.

- **Presentation of SND's data collections**

An overview of the contents of SND's collections. The most frequently disseminated studies, etc.

- **Introduction to how SND communicates**

This session gives an overview of the communication work, both internal and external, at SND.

Additional training, both internal and external, continues throughout the employment. Friday mornings are dedicated to training sessions, in which all staff are encouraged to participate. Each year there will be one or two series of training sessions, covering for example DDI or the legal framework.

External training include training offered at the University of Gothenburg, but also training at summer schools, workshops in connection with conferences, auscultations at other data archives, etc.

Table 2. Example of introductory schedule for data manager

Week	Day	Morning session	Afternoon session
1	Mo	Introduction to the workplace	Review of introductory schedule, work tasks and literature list
	Tu	Overview of how SND manages data – ingest, processing, dissemination	Introduction to the internal data management system SIMS
	We	FAIR data – study level description	Practical work / Reading
	Th	<i>Introduction to the Data Management Forum followed by attending the meeting</i>	<i>Division Meeting</i> Practical work / Reading
	Fr	<i>Friday internal training</i>	Why do we need metadata – introduction to metadata standards
2	Mo	Kind of data: Numeric	Practical work / Reading
	Tu	Kind of data: Text, audio & video	Practical work / Reading
	We	Kind of data: Spatial & image	SND in a bigger perspective – why data repositories in Social Sciences

	Th	<i>Introduction to the Training Forum followed by attending the meeting</i>	<i>Staff Meeting</i> Practical work / Reading
	Fr	<i>Friday internal training</i>	Introduction to the ethical & legal framework for research data
3	Mo	Confidentiality and disclosure risks	Practical work / Reading
	Tu	Data Management and Data Management Plans	Internal and external training at SND
	We	Overview of technical support and development work at SND	Practical work / Reading
	Th	<i>Data Management Forum</i>	<i>Division Meeting</i> Practical work / Reading
	Fr	<i>Friday internal training</i>	What is a Trusted Digital Repository
4	Mo	National cooperation and projects	International cooperation and projects
	Tu	Presentation of SND's data collections	Practical work / Reading
	We	Introduction to how SND communicates	Practical work / Reading
	Th	<i>Training Forum</i>	<i>Staff Meeting</i> Practical work / Reading
	Fr	<i>Friday internal training</i>	Evaluation of Introductory Internal Training

## 6. CONCLUSION AND RECOMMENDATIONS

Task 5.3 of the CESSDA-SaW project set out to explore skills and competences required for staff in (CESSDA) data archives. The current report therefore outlines theoretical approaches, i.e. different competency models to illustrate the entire field of functions needed in data curation. Furthermore, it included a description of current research on the topic, as well as a best-practice example from a CESSDA member organisation, namely SND. This investigation aimed at developing a full picture of skills necessary in this field. One important result of the investigation is that skills and competencies required in digital curation are numerous, and are, not least because of the continuing digitalisation, expected to increase even further. It is therefore questionable if the development of an overall curriculum for a professional qualification in digital data services could meet the needs of both data archives and their (future) employees.

The first report from the Commission High Level Expert Group (HLEG) on the European Open Science Cloud (EOSC)<sup>20</sup> stated that there is ‘an alarming shortage of data expertise in the EU’. They estimate there is need for more than 500.000 ‘Core Data Experts’ during the next 10 years, and therefore recommend<sup>21</sup> a very substantial training initiative in Europe’ with the assistance of among other stakeholders, the ESFRIs and e-Infrastructures. They also recommend that ‘[b]y 2020, to have in each Member State and for each discipline at least one certified institute to support implementation of Data Stewardship per discipline’.

CESSDA has recently endorsed the principles of EOSC and committed to join the ‘coalition of doers.’<sup>22</sup> It is furthermore possible, that some educational efforts are made in other ERICs and infrastructures that could be suitable for CESSDA. Interdisciplinary exchange might therefore be helpful in order to prepare staff for future challenges.

### Recommendations

There are a number of possible ways to enhance the capabilities of current employees as well as those of future staff with the overall aim of fostering an increasing collaboration between CESSDA member organisations.

Using existing CESSDA bodies for the development of hands-on trainings for (future) member employees:

- CESSDA Training Group: new additional programme focussing on training to enhance skills in data curation (e.g. webinar ‘Introduction to data curation in CESSDA for new employees, following the example from SND).
- CESSDA Expert Seminar: bringing together staff from different CESSDA members to exchange ideas and experiences with a special focus on the future of data curation.

---

20 Realising the European Open Science Cloud,  
[https://ec.europa.eu/research/openscience/pdf/realising\\_the\\_european\\_open\\_science\\_cloud\\_2016.pdf](https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf), accessed 28 Sept 2017.

21 Recommendation I3.

22 <https://www.cessda.eu/Consortium/Communication/News/CESSDA/CESSDA-ERIC-commits-to-European-open-science-by-joining-the-coalition-of-the-doers>, accessed 28 Sept 2017.

- CESSDA Knowledge Sharing Platform: collecting training material to create collections particularly for new employees and career development of employees.

Developing new CESSDA groups/programmes

- Bringing employees of CESSDA member organisations together to discuss their jobs' tasks in order to complete the picture suggested by this report.
- Admin exchange: bringing together staff from human resources (and other administrative) departments to jointly develop ideas on how to find new ways of hiring and supporting staff.
- Exchange programme for employees between CESSDA member organisations: different CESSDA Service Providers could be appointed for specific topics and host regular trainings (e.g. workshops) regarding these subjects.
- Sponsorship programme for new employees: finding sponsors for new employees in other CESSDA member organisation and thereby helping them to get acquainted with CESSDA and its aims.
- Collaboration programme with European universities: As described in chapter 4 of this report, a number of European universities offer study programmes preparing their students for jobs in data curation. It is of vital importance for CESSDA to increase its efforts of collaborating with the same institutions.

## REFERENCES

CESSDA-SaW – Deliverable 3.1, Heuristic Maturity Development Model, available online:

<https://www.cessda.eu/Projects/All-projects/CESSDA-SaW/WP3/CESSDA-CDM>, accessed 28 Sept 2017.

Choi, Youngok; Rasmussen, Edie (2009): What qualifications and skills are important for digital librarian positions in academic libraries? A job advertisement analysis. In: *The journal of academic librarianship* (35), S. 457–467, available online:

<http://www.sciencedirect.com/science/article/pii/S0099133309001074>, accessed 28 Sept 2017.

DigCurV - Curriculum Framework for Digital Curation, available online:

<http://www.digcurv.gla.ac.uk/>, accessed 28 Sept 2017.

EDISON (Education for Data Intensive Science to Open New science frontiers) – Data Scientist Competences and Skills Framework; available online: [http://edison-project.eu/sites/edison-project.eu/files/filefield\\_paths/edison\\_cf-ds-draft-cc-v07\\_0.pdf](http://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_cf-ds-draft-cc-v07_0.pdf), accessed 28 Sept 2017.

Engelhardt, Claudia; Strathmann, Stefan; McCadden, Katie (2012): Report and analysis of the survey of Training Needs, available online:

[https://scholar.google.de/scholar?q=Report+and+analysis+of+the+survey+of+Training+Needs&btnG=&hl=de&as\\_sdt=0%2C5](https://scholar.google.de/scholar?q=Report+and+analysis+of+the+survey+of+Training+Needs&btnG=&hl=de&as_sdt=0%2C5), accessed 28 Sept 2017.

Kennan, Mary Anne; Cole, Fletcher; Willard, Patricia; Wilson, Concepción; Marion, Linda (2006): Changing workplace demands: what job ads tell us. In: *Aslib Proceedings: New Information* (58, No. 3), S. 179–196. available online: [www.emeraldinsight.com/0001-253X.htm](http://www.emeraldinsight.com/0001-253X.htm), accessed 28 Sept 2017.

Kim, Jeonghyun; Warga, Edward; Moen, William E. (2013): Competencies Required for Digital Curation: An Analysis of Job Advertisements. In: *The International Journal of Digital Curation* (8), S. 66–83, available online: <http://www.ijdc.net/index.php/ijdc/article/view/8.1.66>, accessed 28 Sept 2017.

Lee, Christopher A.; Tibbo, Helen R.; Schaefer, John C. (2007): Defining What Digital Curators Do and What they Need to Know: The DigCCurr Project. School of Information and Library Science, University of North Carolina at Chapel Hill. 7th ACM/IEEE-CS joint conference on Digital libraries. Vancouver, 18.06.2007, available online:

<http://dl.acm.org/citation.cfm?id=1255183>, accessed 28 Sept 2017.

RDA – Education & Training on Handling of Research Data (wiki index): Task Force on Defining data handling related competences and skills for different groups of professions – working area, available online: <https://www.rd-alliance.org/groups/education-and-training-handling-research-data/wiki/education-and-skills-data-science-wiki>, accessed 28 Sept 2017.

Swan, Alma; Brown, Sheridan (2008), *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs*. Report to the JISC, available online: <https://eprints.soton.ac.uk/266675/>, accessed 4 Oct 2017.

Job descriptions UKDA, available online: [http://www.data-archive.ac.uk/media/422036/ukda\\_job\\_descriptions.pdf](http://www.data-archive.ac.uk/media/422036/ukda_job_descriptions.pdf), accessed 28 Sept 2017.

Xia, Jingfeng; Wang, Minglu (2014), Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions, available online: <https://doi.org/10.5860/crl13-435>, accessed 28 Sept 2017.