



Skills for FAIR: data science and data stewardship, curriculum frameworks - Pillar 4

Session date: 13th of May 2020

Chair: Lennart Stoy (FAIRsFAIR)

Rapporteur: Sara Pittonet (FAIRsFAIR)

Information on the participants, the projects and working groups that they represent, and the spreadsheet used during the workshop can be found in the workshop report: <https://doi.org/10.5281/zenodo.3953979>

All recommendations and the action plan can be found on pp. 59-75 in *Turning FAIR into Reality*: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

This session is about recommendations 10 and 11.

Table of contents

Skills for FAIR: data science and data stewardship, curriculum frameworks - Pillar 4	1
Rec. 10: professionalise data science & stewardship roles	2
10.1 In place	2
View from European Group of FAIR Champions	4
10.2 Planned	5
Rec. 11: implement curriculum frameworks and training	6
11.1 In place	6
11.2 Planned	6
Whole-Pillar.1 What's missing in the recommendations and actions in this pillar?	7
Whole-Pillar.2 Any recommendations not addressed?	8





Rec. 10: Professionalise data science and data stewardship roles and train researchers

Steps need to be taken to develop two cohorts of professionals to support FAIR data: data scientists embedded in research projects, and data stewards who will ensure the management and curation of FAIR data. All researchers also need a foundational level of data skills.

Action 10.1: Key data roles need to be recognised and rewarded, in particular, the data scientists who will assist research design and data analysis, visualisation and modelling; and data stewards who will inform the process of data curation and take responsibility for data management.

Stakeholders: Funders; Institutions; Research communities.

Action 10.2: Formal career pathways must be implemented to demonstrate the value of these roles and retain such professionalised roles in support of research teams.

Stakeholders: Institutions; Coordination fora.

Action 10.3: Professional bodies for these roles should be created, consolidated when they exist, and promoted. Accreditation should be developed for training and qualifications for these roles.

Stakeholders: Institutions; Data service providers; Research communities.

Action 10.4: Data skills, including an appropriate foundational level in data science and data stewardship, should be included in undergraduate and postgraduate training across disciplines, and in the provision of continuing professional development (CPD) credits for researchers.

Stakeholders: Institutions; Data service providers; Research communities.

Related recommendations: Rec. 11: Implement curriculum frameworks and training; Rec. 6: Recognise and reward FAIR data and data stewardship.

Rec. 10: professionalise data science & stewardship roles

Recommendation 10 is containing 4 sub-recommendations on

- 1 key data roles
- 2 formal career pathways
- 3 professional bodies
- 4 data skills in formal and continued professional education

10.1 In place

*What have the projects **already done** that addresses this recommendation? This should build on the information in the spreadsheet. Please check that there is a link to the concrete deliverable.*

FAIRsFAIR WP6 - WP6 is about establishing a Competence Centre for Data Stewardship, providing an help desk and collecting materials and guidance sources for different roles, trying to relate what FsF is doing with what other projects are also doing. As part fo WP6 we also co-organise RDA-CODATA-FAIRsFAIR summer schools developing curricula on data stewardship - first one in the remit of the project organised in August 2019.

With reference to rec 10 referring to professionalisation a key asset is the school and train-the-trainer material available for other projects to develop joint activities.

EOSC Skills WG. We started this WG later than the others but we defined several task forces, one is on defining minimal skills set for the EOSC. Another one is related to competence centres, another one on National strategies,. At the moment the only outcome we have comes from Task force on minimal skills set - where we are about to finalise a diagram that



FAIRsFAIR

Fostering Fair Data Practices in Europe

represent the target users for the EOSC ecosystem for the minimal viable EOSC. It's defining who we are talking to - so target users - and their needs. This should be finalised by end of this week (15 May).

FAIRsFAIR WP7: FAIRsFAIR Deliverable 7.1 also included some questions on the establishment of data support roles at various levels within HEIs - and the recommendation to institutions that lack them to develop policies, processes and structures for institutional data management.

EOSC Nordic: We don't have training as part of the project but the Project coordinator is running a regional series of data stewardship courses - 2 run so far, 1 happening in October 2020. Target: 100-150 data stewards to be trained in the Nordic region, leveraging on the GOFAIR trainers.

Then, through hackathons and dissemination events we can contribute to training-related activities.

NI4OS: Rec 10: we have training activities focusing on open and FAIR data management. we currently organise a series of train-the-trainers activities to then perform training activity in the national settings, leveraging on single partners. WP2 is about building national open science initiatives and its working with partners to influence strategic policy roadmaps and plan in each country - including training aspects.

ExpANDS: in ELETTRA Synchrotron we are working on Data Policies and namely on a glossary of common terms to implement in WP2 workflow (Abigail is also involved). We are still in the phase of data collection.

EOSC Synergy: Professionalisation & Stewardship roles: we are focusing on the quality of trainers we want for the future.

ENVRI-FAIR: in ENVRI FAIR WP6 is dedicated to Capacity building. We started with a gap analysis in 2019 about the existing training materials. We did a survey with Research infrastructures and sub-domains teams about training to understand their need. We produced a deliverable out of this survey (D6.1). What's not available we've tried to develop it. We joined the EOSC-LIFE schools in 2019.

ESCAPE: Our project focuses on Astronomy and particle physics. We share the same interest for FAIR. Training is addressed by different WPs and its more linked to training events - schools and hackathons. Often Data stewardship is connected to researchers who might not become professional data curators as librarians, for instance. Data steward profiles are different from researchers.

SSHOC: SSHOC published D6.7 on inventory of existing training materials <https://doi.org/10.5281/zenodo.3596003>. Based on this inventory SSHOC published a





[training toolkit](#) on the website guiding resources to organise webinars and design training curricula in SSHOC.

There will be SSHOC webinars with LIBER and Trust-IT on 19 May about FAIRification in the scholix framework. <https://sshopencloud.eu/training/training-events>

<https://sshopencloud.eu/training>

Training materials are important building blocks of the [SSHOC Open Marketplace](#) as well where training materials will be interlinked with tools, services and data to contextualize these assets to SSH scholars.

At DARIAH we have a Research Data Management WG (<https://www.dariah.eu/activities/working-groups/research-data-management/>) for 2 main stakeholders: 1. Humanities scholars to collect and share use cases of FAIR data workflows from different disciplines (Archaeology, Lexicography, Corpus linguistics, Geohumanities etc.) 2. To build a knowledge hub for new data support roles working in the humanities fields. There are many profiles called data stewards, open science officers or others, but they are all doing a similar job and we should have a unique definition of this role in EOSC in the humanities.

- **FAIRsFAIR WP7:** in FAIRsFAIR we have the ambition to create a competence framework that addresses both data science and data management
- **FAIRsFAIR WP3:** echoing the above: training should take place at different level of institutions. Costing RDM is important.
- **FAIRsFAIR WP6 supported by several participants:** it's also important that training is not seen as a "one off" activity and should be part of continuing professional development . We need to think of it as in any working environment, you need to keep on updating training and adapting on new knowledge

View from European Group of FAIR Champions

FAIR Champion 1: At the Research Council in Spain we started running research data training programmes in 2015. Originally the training was the same for all sectors - librarians, project managers, etc... but then we started to prepare more tailor made courses according to the audience, such as:

- Researchers → understanding licences or tools to analyses data
- Librarians → lighter courses on data stewardship

The Spanish National Research Council launched the first master in Data Science targeting different type of experts in data management in the country. We are experiencing difficulties in recruiting profiles who are NOT engineers, we are struggling in getting other experts like librarians or people from the legal sector or from the policy sector. There is a lack but an interest in getting this type of profiles.

FAIR Champion 2: the Swedish consortium for research data is engaging most universities in Sweden. We are expanding training for archiving staff and also for researchers directly - this is an expanding activity in the country.





FAIRsFAIR

Fostering Fair Data Practices in Europe

From the RDA WG on Training we have different projects ongoing and this group is growing importance.

10.2 Planned

*What are the projects represented **developing or planning** to do? Again, this should build on the information in the spreadsheet: information about a planned deliverable, i.e. title, due date, short description*

FAIRsFAIR WP7: WP7 works on a competence framework for FAIR for use the higher education. This will be supported by an implementation handbook and training events for integration of RDM & FAIR in higher education curricula. Work is currently in preparation - challenge is to narrow it down (what level, discipline or more generic, data science broadly or data stewardship).

FAIRsFAIR WP3: WP3 plans to work with the proposed RDA IG on Professionalising data stewardship, to transfer lessons from the RSE community in establishing the Society of Research Software Engineers.

EOSC Skills & Training WG: We are then going to map the existing frameworks in order not to duplicate efforts. Commonalities with hubs will also be defined in the upcoming months. We are also contributing to the Strategic Research and Innovation Agendas.

EOSC FAIR WG: In the FAIR WG in terms of Skills for FAIR we are collecting info that is feeding into other reports by organisations to see how they are developing tools. A report is going to be produced in June 2020. Potential overlap with FAIR Metrics & Certification Task Force. I am also involved in the [OECD GSF Digital Skills for Science group](#) which will be publishing a report on recommendations on “Digital capacity and skills for data-intensive science” later this year - this includes data stewards and data scientist roles.

NI4OS: WP2 is about building national open science initiatives and is working with partners to influence strategic policy roadmaps and plan in each country - including training aspects. We still don't know how this can be applied to the individual countries since they all have slightly different agendas. We are targeting service providers and researchers as the main target of this activity. A [training plan](#) is available.

DARIAH (SSHOC): Future plans: we want to establish stronger ties with the cultural heritage sector as they are important partners in humanities data workflows providing source materials for their work.

ESCAPE: With respect to Rec 10 we will run a specific event next year targeting people working in archives and data centres dealing with space data. A deliverable will gather the outcomes of this.





Rec. 11: implement curriculum frameworks and training

Rec. 11: Implement curriculum frameworks and training

A concerted effort should be made to coordinate and accelerate the pedagogy for professional data roles. To support uptake, skills transfer schemes, fellowships, staff exchanges and informal training opportunities are needed, as well as formal curricula.

Action 11.1: Curriculum frameworks for data science and data stewardship should be made available and be easily adaptable and reusable.

Stakeholders: Institutions; Coordination fora.

Action 11.2: Sharing and reuse of Open Educational Resources and reusable materials for data science and data stewardship programmes should be encouraged and facilitated.

Stakeholders: Institutions; Coordination fora; Data service providers.

Action 11.3: Practical, on-the-job methods of training such as fellowships and staff exchanges should be supported, as well as Train-the-Trainer programmes so the body of data professionals can rapidly scale.

Stakeholders: Institutions; Data service providers; Data stewards; Funders.

Action 11.4: A programme of certification and endorsement should be developed for organisations and programmes delivering train-the-trainer and/or informal data science and data stewardship training. As a first step, a lightweight peer-reviewed self-assessment would be a means of accelerating the development and implementation of quality training.

Stakeholders: Institutions; Coordination fora; Standards bodies.

Related recommendation: Rec. 10: Professionalise data science and data stewardship roles.

11.1 In place

*What have the projects **already done** that addresses this recommendation? This should build on the information in the spreadsheet. Please check that there is a link to the concrete deliverable.*

- ENVRI-FAIR has conducted an “Inventory & gap analysis of FAIR training materials” (https://envri.eu/wp-content/uploads/2019/10/ENVRI-FAIR_D_6-1.pdf) that identifies priorities for training requirements across several domains and skills groups.
- SSHOC has mapped existing training material on FAIR data (and EOSC) (https://zenodo.org/record/3596003#.XseT0cBS_Vg) that e.g. includes a gap analysis and an analysis of metadata used and needed for the training material.
- FAIRsFAIR includes training and train-the-trainer activities - it already supported the 2019 CODATA/RDA Summer School on Data Science and Data Stewardship.
- NI4OS: Training the trainers programme focuses on Open and FAIR RDM; promoting those principles at the national level by the trainers is expected.

11.2 Planned

*What are the projects represented **developing or planning** to do? Again, this should build on the information in the spreadsheet: information about a planned deliverable, i.e. title, due date, short description*

Regarding 11.1 - curriculum frameworks:

FAIRsFAIR: WP7 works on a competence framework for FAIR for use in higher education. Work is currently in preparation - challenge is to narrow it down (what level, discipline or more generic, data science broadly or data stewardship)



FAIRsFAIR

Fostering Fair Data Practices in Europe

Regarding Action 11.2 - Open Educational Resources:

FAIRsFAIR: Making materials available is one of the key aspects of the FAIRsFAIR project (WP6 + WP7).

EOSC Skills WG: Working on recommendations for a training catalogue.

FAIRsFAIR: Useful resource: Workshop report "Training in the EOSC" (<https://doi.org/10.5281/zenodo.3739055>)

EOSC Synergy: Being a very technical project we are providing the infrastructure to provide self-learning and we are investigating which is the best platform - MOOC to compile. Guidelines and recommendations are being developed for the train the trainers - so one step back to really define what the trainers should do.

NI4OS: The project supports Training of Partners and creation of training material to be translated to the partners' languages for Fairness. The material will be based on existing training material/guidelines transformed to meet stakeholders' needs according to the project's stakeholders' survey findings.

EOSC-Pillar:: For Task 5.3 in EOSC Pillar (Helpdesk and documentation for promoting FAIR practices and support to FAIR-oriented data stewardship), we are preparing guidelines and a catalog of resources aimed at data steward teams / RDM support staff. The output could support and/or complement the FAIRsFAIR Knowledge Base of the competence centers.

Whole-Pillar.1 What's missing in the recommendations and actions in this pillar?

What do projects do - related to implementing FAIR in the context of the EOSC - that is not covered by the original recommendations? Should it be included in an updated action plan and revised set of recommendations? Please focus on this pillar.

GEOMAR: What about non-Researchers working with data? How are they addressed? E.g Laboratory personnel, technicians. These should also be addressed by professional training programmes. Training too often is far too specialised for researchers. At Kiel University we've run a programme to start training people from the bachelor into basic curricula and we figured that the problem is bigger than training PhD. Librarians and laboratory people need to be trained. Training should appear in many other TFiR Pillars as well.

Example: training laboratory personnel: these people are much more motivated in documenting their methods and are willing to contribute.

EOSC Synergy: we should reconsider this recommendation as providing inputs to all the other 5 TFiR pillars. See TFiR Rec.1 - Action 2: Action 2.2: "Educational programmes are needed to raise awareness, understanding and use of relevant standards; tools are needed to facilitate the routine capture of metadata during the research process."





Whole-Pillar.2 Any recommendations not addressed?

Are the recommendations being covered enough by these activities? If not: what should be done? And by whom?

Regarding 10.3 - Establishing professional bodies:

FAIRsFAIR & FAIR Champion 1: General agreement on the need for professional bodies to be involved to acknowledge/endorse/accredit data steward roles. It may seem that data stewards are not being so high profiled as other types of data experts

FAIRsFAIR - RDA has a new Interest Group being proposed on [Professionalisation of Data Stewardship](#), which might lead to some form of professional body for data stewards. GOFAIR is also working towards the same direction through [DSCC-IN](#) (Data Stewardship Competence Centre - Implementation Network).

Regarding 11.4 - Certification:

FAIRsFAIR: about Certification & formalisation: there is the need to harmonize specs for professionalisation whose criteria might vary according to sectors and institutes

Regarding 11.3 - mentorship & staff exchanges:

FAIRsFAIR: mentorship and staff exchanges should also be considered as important activities in light of addressing stewardship. The FAIRPlus project is working on a fellowship programme. Other option: the ERASMUS plus programme?

Regarding 10.1 & 10.2 - recognition and formal career pathways:

FAIR Champion 1: what I see in many institutions in Spain is that data stewards are not getting much professional reward and this might become frustrating and affect the sustainability of the entire system.

FAIRsFAIR: Rewards have been addressed to some extent in pillar session 2, about the FAIR culture. We've been trying to focus on skills and professionalisation for all of the roles, and had some discussion about badges for quality assessment. **Lennart Stoy:** Rethinking academic careers is something under discussion at EUA for instance, though support roles in institutions are not commonly addressed in this discussion.

FAIRsFAIR - training is something that could be easily addressed by projects while projects quite often don't have the decision power to decide autonomously on actions to be taken for recognition and formal careers. This might be a reason why this group cannot assess the Pillar recommendations in detail.



FAIRsFAIR

Fostering Fair Data Practices in Europe

GEOMAR: Maybe projects can provide beyond recommendations and material, and devise workflows, which can easily be followed as a first approach on data practices and implemented in institutions. - see the planned actions of NI4OS on influencing strategic policy roadmaps and plans in each country

FAIR Champion 2: as a Champion I confirm data science is becoming a reality in astronomy, people are investing in machine learning. while data stewardship requires more institutional support and infrastructure - we need precise profiles to do that. There are communities in astronomy and particle physics who are talking about this with conferences every 3 years <http://www.eso.org/sci/libraries/lisa.html> (now happening in 2021). In France there are networks of librarians in astronomy who are bringing FAIR into their discussions. Getting funding agencies and governing bodies to acknowledge the importance of these roles is fundamental to ensure support to these new profiles.

