

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

Session date: 18th of May 2021

Chair: Elizabeth Newbold

Rapporteur: Hugh Shanahan

All recommendations and action plan on pp. 59-75 in [Turning FAIR into Reality](#)

Event Host: Marialetizia Mari, Serenella Muradore (Trust-it)

Meeting Attendance:

	Name	Organisation and project
1	Elizabeth Newbold	STFC/FAIRsFAIR
2	Hugh Shanahan	Royal Holloway, University of London, FAIRsFAIR
3	Marjan Grootveld	DANS, FAIRsFAIR
4	Maggie Hellström	ICOS ERIC and ENVRI-FAIR
5	Katrin Seemeyer	Forschungszentrum Juelich, ENVRI-FAIR
6	Ruta Petrauskaitė	FAIRsFAIR HLAC
7	Stan Gielen	NWO, FAIRsFAIR HLAC
8	Oliver Knodel	Helmholtz-Zentrum Dresden-Rossendorf, ExPaNDS
9	Iryna Kuchma	EIFL, OpenAIRE
10	Angus Whyte	Digital Curation Centre, FAIRsFAIR
11	Maria Johnsson	Lund University, ENVRI-FAIR, ICOS ERIC
12	Mark Allen	ESCAPE, FAIR champion
13	Marialuisa Lavitrano	University Milano-Bicocca / BBMRI.it / EOSC-Association
14	Susanna-A Sansone	FAIRsFAIR Data Champion; standing in for EOSC-LIFE; also FAIRsharing, and ELIXIR-UK; University of Oxford.
15	Andreas O Jaunsen	EOSC-Nordic WP4, Nordforsk
16	Aoife Coffey	UCC/NORF
17	Andreas Athenodorou	The Cyprus Institute, NI4OS-Europe

18	Elli Papadopoulou	ATHENA RC / OpenAIRE and NI4OS-Europe
19	Simon Hodson	CODATA, FAIRsFAIR
20	Timea Biro	DRI / NORF
21	Barbara Magagna	Environment Agency Austria, ENVRI-FAIR, eLTER, GO FAIR
22	Carlos Casorrán	EC, EOSC Executive board
23	Isabel Bernal	CSIC, EOSC Synergy, FAIR Champion
24	Erzsébet Tóth-Czifra	DARIAH-EU, SSHOC
25	Karsten Peters-von Gehlen	German Climate Computing Center (DKRZ), Hamburg, Germany
26	Carthage Smith	OECD, FAIRsFAIR HLAC

This session is about recommendations 10 and 11.

There are 2 questions per recommendation, followed by 2 questions about the pillar.

Rec. 10: professionalise data science & stewardship roles

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.*

Andreas Jaunsen (EOSC NORDIC) - Not much (funded) activity in EOSC-NORDIC on FAIR training, but NeIC (EOSC-Nordic coordinator) is running a series of FAIR data stewardship courses (5 full day equivalent) - so far *160 Data Stewards* trained. The goal is to build competence on a common curriculum. This is formal training. In EOSC-Nordic we run a series of (informal training) FAIRification webinars that are somewhat more technical and more “steps” are coming after summer.

Andreas Athenodorou (NI4OS) - strong training agenda on skills for FAIR for service providers (thematic and repository).

Oliver Knodel (Expands) - working with PANOSC project and completed internal workshop series 'Train the Trainers' on creating courses on the PaN training portal.

Marialuisa Lavitrano (EOSC Association/EOSC Synergy) - **EOSC Synergy** are providing a resource/e-learning platform (Moodle) to allow people to set up their own online courses. Not much available yet but includes examples such as resources from FOSTER. Second part - putting together training material from EOSC Synergy to promote usage in universities, particularly Data Science programmes.

Maria Johnson (ENVRI-FAIR) - the focus has been on raising researcher awareness of data skills rather than on data stewards. During the past year have run many webinars and

training events, both technical and introductory. Common training platform. Early deliverables discussed in the spreadsheet.

Maggie Hellström (ENVRI-FAIR) - within the ENVRI-FAIR collaboration it is a very mixed bag of partners, some very advanced, some just starting, some not even thinking about setting up repositories.

Susanna Assunta Sansone (EOSCLife) - Set of training activities: <https://www.eosc-life.eu/services/training> incl. (i) knowledge exchange workshops; (ii) teaching researchers and clinicians how to share COVID-19-related data and software (planned); (iii) mentoring and Training opportunities for EOSC Research Infrastructures => Training Open Calls: <https://www.eosc-life.eu/news/4-projects-awarded-funding-within-first-training-open-call> terms4FAIRskills: <https://terms4fairskills.github.io>

Mark Allen (ESCAPE) - training for data providers will run later in the year; across all the projects everything will be put in the context of FAIR. Two approaches: researchers on the one hand and data providers on the other. Schools for researchers to use tools. More focus for data providers in the future.

Barbara Magagna (GO-FAIR) - working with GO-FAIR training with Data Stewards - trained 140 DSs for COVID-19 ZonMw (Dutch funder) projects and creating FAIR vocabularies and creating handbook for Metadata4Machines. Preliminary work done but not yet published.

Erzsébet Tóth-Czifra (SSHOC) - SSHOC has run train-the-trainer bootcamps. There is a wide variety of different SSHOC disciplines that need this type of training. To aid this three data communities listed below have been set up. Three kicked off recently which represent a diverse range of areas.
<https://sshopencloud.eu/training> ; <https://www.sshopencloud.eu/ssh-data-communities>
<https://sshopencloud.eu/training/training-events>

Iryna Kuchma (OpenAIRE) - The activities take place at a national or institutional level. OpenAIRE provides a Community of Practice for training coordinators:
<https://www.openaire.eu/cop-training>

Isabel Bernal (EOSC Synergy) - the emphasis has been on organising the first edition of training courses, one on reproducibility which has been a big success. Researchers are from different communities. Different modules being developed on different issues of data science and EOSC. Organising FAIRsFAIR roadshow in Spain for policy makers.

Karsten Peters-von Gehlen (FAIR Champion) as an older institution (almost 30 yrs) DKRZ is getting the hang of FAIR and addressing user demands in that direction. E.g. a researcher wants to "archive their data in a FAIR manner".

Marialuise Lavitrano (EOSC Association)- (via chat) EOSC-Association just launched *Research careers and curricula Advisory Group* also to connect with the different projects already ongoing. The Task Forces within this Advisory Group focus on the training and skills agenda. Multiple stakeholder groups are addressed, specifically research communities, data stewards and national bodies.

This TF should build on activities to define data stewardship curricula to ensure these are recognised and aligned across Europe. Attention should also be paid to career paths to ensure appropriate recognition and rewards for data management activities. Recommendations from the Digital Skills for FAIR and Open Science report (<https://doi.org/10.2777/59065>, EOSC WG Skills & Training) should be taken forward by this group. Research careers, recognition and credit.

In parallel with the *data stewardship TF*, this group should address incentives for researchers to manage and share their data, code and other research outputs. This could include criteria on Open Science and FAIR being part of selection and promotion criteria, funders and governments requiring examples of data/code sharing as part of assessment exercises, or data prizes.

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*

Rec. 11: implement curriculum frameworks and training

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.*

Marjan Grootveld (FAIRsFAIR) - development of practical handbook based on D7.3. There is some overlap with a Dutch report Professionalising data stewardship: <https://doi.org/10.5281/zenodo.4623713> (data steward competences and job descriptions/career perspective, partly in relation to research software engineers).

Stan Gielen (FSF HLAC) - Dutch Government provides support for Open Science. They considered different approaches whether it should be top down or bottom up. Bottom up favoured and hence funding provided to Universities. The Dutch government has provided money to organise Data Competence Centers, which fund and coordinate the training of data stewards. This will lead to sharing best practices within two years. The idea is that different Universities have different approaches.

Andreas Athenodorou (NI4OS) - have train- the- trainers events.

Oliver Knodel (Expands) - we are still developing a training portal to submit materials. Also completed a series 'Train the Trainers' on creating courses on the PaN training portal (moodle courses).

Isabel Bernal EOSC Synergy - The focus on giving guidance and good practice on how to set up good training material, preparing modules, promote usage by Universities in Data Science programmes.

Maria Johnson (ENVRI-FAIR) ENVRI-FAIR Summer School (1 online, 1 physical) raising awareness of skills. More train the trainer events so that they can develop their pedagogy and work with trainers.

Maggie Hellström [MH] (ENVRI-FAIR) (chat) The ENVRI-FAIR schools are a continuation of an earlier summer school series that were developed in a collaboration between LifeWatch ERIC and ENVRIplus (the "precursor" of ENVRI-FAIR). Progression from theoretical FAIR to practical implementation.

Susanna Assunta Sansone [SAS] (EOSCLife FAIRplus - FAIR Cookbook, as specific FAIR 'training' online resource (via the FAIRplus, Elixir and pharma driven project):
<https://fairplus.github.io/the-fair-cookbook> => Webinar next week, May 26th: <https://elixir-europe.org/events/fairplus-webinar-discovering-fair-cookbook>

Mark Allen [MA] (ESCAPE) - Mainly Scientific training activities and behind that is Data Stewardship. Astronomy is small and hence people are brought in with either a specific research background or from a repository. There isn't a curriculum of data stewardship. Tend to think that will have to train them on the job,

Marialuisa Lavitrano (EOSC Association) - EOSC Association - number of advisory groups, one on career development. Several task groups, one on training and skills. Attention will be paid to career path. Another on research career.

--

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*

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.

Isabel Bernal EOSC Synergy- with respect to training and promoting skills, there is an increasing challenge to train people who are doing interdisciplinary work in the same problem. This can be very disparate (humanities and imaging) and trainers need to consider how to deliver that type of training.

Stan Gielen (FSF HLAC): related to this point then is the question then should we distinguish criteria and skills for data scientists/stewards between various disciplines?

Maria Johnson (ENVRI-FAIR) this is also true in ENVRI-FAIR. It is a challenge but also an opportunity. There needs to be a greater focus on train-the-trainer (instructor training) support.

Simon Hodson (FAIRsFAIR) - Three topics were raised to consider as recommendations and/or for the white paper:

- train the trainer (instructor training), it is important development and key for scaling up the number of data stewards. There is a significant amount of activity in this area and there is a need for formalised skill sets and to teach or revisit the role of pedagogy in such training.
- Cross-domain or interdisciplinary work, in particular what are the requirements for this, understand skills and competences, provide examples of projects and growing research areas and the skill sets needed to understand 'other peoples' data';

- 'Learning on the job' - namely learning during the process of being a data steward, researcher etc. There is a need to have a clear grasp of and hence formalise the relevant skills and competencies. There is a duty to ensure that people come out of post graduate studies with clearly defined data skills both as domain researcher and as appropriate, if they go down the data steward route, with the skills required for that role.

Angus Whyte (FAIRsFAIR) - Is there a need to distinguish between data stewardship that data centres do, and the data stewardship that Research Performing Organisations do?
(chat)

Marialuisa Lavitrano (EOSC Association) - The process of learning elements of data stewardship in an ad hoc fashion on the job is not enough. There is a need for formal training and a need to define a competency profile. This needs to be integrated with University programmes at all levels from Bachelor to Post Doc.

Elizabeth Newbold (FAIRsFAIR) re Marualuisa comment above who will formalise this training? Is it within the domain or somewhere else?

Marialuisa Lavitrano (EOSC Association) Community developed standards need to be developed for career paths and for competences.

Hugh Shanahan (FAIRsFAIR)] both generalist and domain expertise are required. There will be a tension between specific domains, where there is a tendency to getting people up to speed quickly and a domain agnostic approach to data stewardship. The argument that you need some element of a domain agnostic approach needs to be won.

Susanna Assunta Sansone [SAS] (EOSCLife) At the University of Oxford, this discussion has started. It is clear that only a certain level of things can be done in a centralised and domain agnostic matter. Researchers ask for domain/data-specific help and guidance and those can only be delivered by domain/disciplines experts, in that same organisation or outside (in a digested manner).

Carlos Casorran European Commission- very useful to hear what has been said.

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?