Perspective

Progress toward a comprehensive teaching approach to the FAIR data principles

Hugh Shanahan, 1, * Nancy Hoebelheinrich, 2 and Angus Whyte 3

1Department of Computer Science, Royal Holloway, University of London, Egham, England TW20 0EX, UK
2Knowledge Motifs LLC, San Mateo, CA 94401, USA
3Digital Curation Centre, University of Edinburgh, Edinburgh, Scotland EH3 9DR, UK

*Correspondence: hugh.shanahan@rhul.ac.uk

https://doi.org/10.1016/j.patter.2021.100324

SUMMARY

We evaluate recent efforts to further the effective teaching of FAIR data principles by examining existing and developing educational frameworks focused upon FAIR, training initiatives that have informed teaching on FAIR skills’ topics, and a number of key sources for discovering FAIR training materials and how much those sources provide descriptive information about the materials. FAIR4S, providing a coherent description of skills and competencies, is analyzed by target audience using the description of actors found in a European Open Science Cloud ecosystem report and by comparison of the coverage and extent of description of educational and training materials available from the list of sources for finding such materials. Our analysis elucidates the importance of linking resources to FAIR-related educational frameworks, providing consistent descriptions of them using a community-based metadata scheme, and developing an instructor community of practice where ideas and methods can be shared on how to teach FAIR data skills.

INTRODUCTION

The FAIR (findable, accessible, interoperable, and re-usable) data principles, published in 2016, are a set of guiding principles proposed by a consortium of scientists and organizations to support the reusability of data given their rapid and unprecedented volume, complexity, and speed of creation. 1, 2 The guidelines have been adopted by research institutions worldwide, yet implementation of the FAIR principles has been slow. Since its inception, FAIR principles have increasingly been adopted as a generic set of principles to enable data sharing as well as reusability. The implementation of the principles themselves entails a wide range of skills that will need to be employed by individuals working in many different roles (including data specialists) to ensure data meet FAIR implementation standards for discrete communities. Training will need to be delivered to individuals, often referred to as data stewards, 3 who can be defined as those involved in making data FAIR and keeping it FAIR. Data stewards may be researchers, data scientists, data curators and librarians, data and repository managers at various levels, as well as those working in policy development and who may be involved in government, academia, and industry sectors. 4 The range of audiences and skills needed presents unique challenges from a training perspective as the training needs to be delivered to individuals from many disciplines, although effective materials developed for one discipline can potentially be transferred to many other disciplines, and also adapted to localized language and cultural needs.

As it is, both formal education (institutionalized, intentional, and planned) and informal education (differentiated from formal education by its defining characteristic as an addition, alternative or complement to formal education within the process of lifelong learning of individuals), 5 and continuing professional development (CPD; intentional maintenance and development of the knowledge and skills needed to perform in a professional manner) 6 training in FAIR standards and skills is still very much in development, although this is likely to change rapidly in the near future. There have been a number of initiatives that have moved the training of FAIR forward (these are covered in the section, “training initiatives”). The initiatives described have informed an important next step in the effective training of FAIR data skills.
In order to best accommodate FAIR principles in data skills training, the skills needed should be clearly defined within an educational framework, particularly one that links skills with competencies. Such a framework provides guidance to those building curricula and courses on what learning objectives, learning outcomes, and performance assessment criteria should be envisaged and can then be included in corresponding curriculum frameworks. The educational frameworks that have already been developed cover aspects of FAIR-related skills and provide a coherent description of the skills to make data FAIR and keep data FAIR, particularly in the detailed description of skills found in the FAIR4S project.7

Training initiatives
An increasing number of domain-specific training initiatives and events with coverage of FAIR-related topics have come into play recently. These training initiatives and events have informed efforts to establish educational frameworks for FAIR.

The Carpentries1,3 (https://www.carpentries.org) is the merger of three separate but related activities: Software Carpentry, Data Carpentry and Library Carpentry. It provides introductory training in a variety of topics, such as the command line, Git, Python, and R. In the Library Carpentries strand,10 training to library and information-related roles are associated with topics such as data cleaning. A specific lesson in FAIR data is in development. The Carpentries lessons are highly structured and are available from the Carpentries site. Each course will typically be less than a day to run. Typical Carpentry events run up to 2 days. They have an excellent instructor training program with a strong emphasis on improving pedagogy.

The Top 10 FAIR Data & Software Things11 is a collection of short documents on various aspects of FAIR standards that was created by means of a global crowdsourcing event. The Research Bazaar (ResBaz)12 is a series of 2–3-day events run globally that cover a number of topics with a similar emphasis on coding skills and other topics, such as machine learning. In 2020 they introduced a specific section on training FAIR skills.

In ELIXIR, a working group in FAIR training has been active since 2018,13 and the group is working on (1) best practices in teaching the FAIR principles and (2) making training materials FAIR (https://f1000research.com/posters/8-856).

The CODATA-Research Data Alliance (RDA) schools in research data science (https://github.com/CODATA-RDA-DataScienceSchools/TaskGroup) are 10-day schools to provide foundational training in data science skills for early career researchers (ECRs) with an emphasis on supporting ECRs from low- and middle-income countries. A wide variety of topics are covered that are highly relevant to FAIR, including Carpentries skills, open science research data management (RDM), machine learning, visualization, and computational infrastructures. Versions of the schools for European ECRs and data stewards and in specific domains are currently being rolled out in collaboration with FAIR4S.

OpenAire provides a community of practice (https://www.openaire.eu/cop-training) for training coordinators that aims to map out the training activities of various pan-European and European Open Science Cloud (EOSC)-related initiatives. The goal of the community of practice is to strengthen instructors’ training capacity by improving alignment, sharing experiences and good practices, and initiating cross-infrastructure training activities.

GO FAIR (https://www.go-fair.org/) is an initiative to develop FAIR services by allowing the initiation of “implementation networks” (INs) to focus on implementation of and training on FAIR. At present, there are INs covering a large range of different domains, including data stewardship. Many of these INs will deliver their own training with coordination from GO FAIR’s training focus, in one of the GO FAIR “pillars”, called GO TRAIN14 (https://www.go-fair.org/implementation-networks/overview/).

Educational frameworks
Research data skills are currently being explicated for a number of different disciplines, including data stewardship, knowledge engineering, and data science. A variety of different educational frameworks in these areas have developed important insights that have informed the descriptions of those skills. Specific examples are the DigCurV Skills Framework15 and a competency-based curriculum framework presented by the University of North Texas,16 which focuses on digital curation specialists; the Global Open Data for Agriculture & Nutrition Network (GODAN) curriculum on open data and RDM for agriculture and nutrition researchers and data specialists (https://www.godan.info/documents/curriculum-open-data-and-research-data-management-agriculture-and-nutrition); the Belmont Forum’s Data Skills Curricula Framework,17 for researchers in global environmental change research; and the EDISON Data Science Framework for research and industry.18 A more thorough review of assorted sources can be found on the EOSCPilot Skills and Capability framework document (https://www.eoscpiLOT.eu/sites/default/files/eoscpilot-d7.3.pdf).

While these educational frameworks have proved useful, they are not frameworks specifically focused upon learning FAIR data skills per se. A more recently developed educational framework, the FAIR4S framework (Framework for FAIR Data Stewardship Skills in Science and Scholarship), is targeted toward those wishing to acquire FAIR data skills (see https://eoscpilot.eu/themes/skills). An additional competency framework is being developed by the FAIRSFAIR project, which provides a more detailed breakdown of FAIR skills and is in part informed by FAIR4S.19 These represent present efforts in educational frameworks specifically for FAIR skills that the authors have found evidence for.

FAIR4S is a FAIR data stewardship skills and capability framework for science and scholarship designed to help research communities and research institutions implement RDM and FAIR stewardship in the open science and data science context. The FAIR4S framework, co-authored by the third author of this paper, was developed in the context of the EOCSPilot EOSCPilot. The framework responded to an identified need to develop skills and education in RDM, data stewardship, and data sciences.20 Taking this cue, FAIR4S framed the implementation of FAIR principles as a shared responsibility requiring a synthesis of relevant competences associated with data management, data stewardship, and data sciences.

FAIR4S was initially drafted through a mapping exercise, drawing on existing frameworks. The starting point was a core group of 11 data handling competencies identified for undergraduate-level information professionals, and collectively
Table 1. Outline of FAIR-related skills, taken from the FAIR4S study

| Skills group | Description |
|--------------|-------------|
| PD           | plan and design for stewardship of data |
| CP           | capture and process data and workflows |
| IA           | carry out steps necessary to integrate and analyze the data |
| AP           | appraise data and set up its preservation |
| PR           | carrying out steps to publish FAIR outputs |
| ED           | discover other outputs and make outputs discoverable |
| GA           | ensuring higher levels principles are adhered to |
| SR           | ensuring facilities are within scope and resourced properly |
| AE           | contributing to the wider FAIR community |

AE, advise and enable; AP, appraise and preserve; CP, capture and process; ED, expose and discover; GA, govern and assess; IA, integrate and analyze; PD, plan and design; PR, publish and release; SR, scope and resource.

termed digital information literacy skills. These core competences were mapped to a simple data life cycle used in the RISE (research infrastructure self-evaluation) model for planning research data services, resulting in nine skills areas. This in turn was used to identify related competences from two sources: the EDISON competence framework for data science, and a crowdsourced list compiled by the Research Data Alliance interest group on Education and Training in Data Handling. A synthesized list of 79 main topics, organized in the nine skills areas, was published, then iteratively refined through workshops in which trainers in data management were asked to identify priority topics. These workshops also identified two further sources: European Comission Open Science Policy Platform (OSPP) Open Science Career Assessment Matrix (OS-CAM), and Dutch Techcentre for Life Sciences (DTL) Data Steward competence framework. The OS-CAM was used to check that the framework reflected research activities considered necessary for open science. The DTL framework, developed in parallel, was used as a validatory check. Following further internal review, a final workshop was held involving researchers and data stewards in prioritizing 10 key skills from those listed in FAIR4S.

A summary of the FAIR4S framework scope is given in Table 1. A more detailed summary is provided in the supplementary information.

Sources for training materials

For those seeking training resources on topics related to the FAIR principles, research data skills in general, and RDM, there are a number of options already available that pull together information about these kinds of resources. Ranging from catalogs and registries to educational service providers and portals, these training material sources provide the means for FAIR and RDM learners to find both formal and informal training resources aimed at various target audiences and subject disciplines, many of which exist under a variety of open source licenses. As there are a variety of different terms used to describe these options, we will use the generic term “source” for any one of them.

The sources listed here have been selected using the following criteria:

1. They cover topics relevant to FAIR and make explicit reference to FAIR according to the FAIR4S categories.
2. Are oriented toward relevant communities, e.g., research domains.
3. Attempt to collate materials from a variety of different resources.
4. Target different audiences associated with research life cycle or research production workflow; for example, researchers, data stewards (see Figure 1).

Given the rapid development of training resources, it is important to not only describe those sources developed at the present moment, but also future sources that may come into play, on each of the criteria above. Taking the target audiences first, these need to align with consensus views of the actors and stakeholders involved in implementing FAIR principles. One such view is shown in Figure 1, relating to the EOSC.19

Existing and future sources of training materials offer both FAIR data skills seekers and instructors many opportunities to discover training materials of interest, but the extent to which these sources provide consistent, searchable descriptions could be improved. In particular, the descriptions do not often include links to educational frameworks such as the skills framework presented by FAIR4S. In addition, the descriptions for the materials vary quite a bit in terms of the information provided and often do not include important information that helps those searching not only discover training materials but also judge whether the materials are appropriate for their learning goal and objectives. In this section, we provide (1) a narrative describing what these sources cover and how they approach the information they offer; (2) a comparative table that shows how the information the sources provide map to the skills and competencies framework of FAIR4S, and whether they offer descriptions of target audience, learning objectives or educational themes, and knowledge level expected for the target audience.

Sources narrative

In 2016, a collaborative group of Earth science organizations launched the Earth Sciences Information Clearinghouse (ESIP)-hosted Data Management Training Clearinghouse (DMTC) at http://dmtc.clearinghouse.esipfed.org. The DMTC provides a searchable registry of information on curated training and educational resources that covers topics related to research data skills and RDM. The DMTC uses the Learning Resource Metadata Initiative (LRMI) metadata scheme, which is maintained by the Dublin Core Metadata Initiative and built upon the extensive vocabulary provided by Schema.org for the resources included in that registry. The materials range from links to full online courses (Massive Open Online Courses [MOOCs]) that provide the opportunity for certification to short courses that make both presentation slides and verbal representations of the slides available for download and re-use. The foci of these resources include capacity building on data skills important during all phases of the research and data
life cycles, and on learning to use various tools appropriate for data curation and data stewardship tasks. The links to material found on the DMTC include a category of materials called educational curriculum guides, which have been created and submitted by a large number of different organizations. Subject disciplines covered in the DMTC lean heavily to the geosciences and other physical sciences, but also include arts and humanities and the social sciences. In addition, the DMTC provides the means to associate educational framework categories and their nodes to individual resources so that a learner can find resources in specific categories. Examples of the educational frameworks to which DMTC resources may be associated include the FAIR principles themselves, DataONE’s research life cycle, and the ICSU World Data System’s Training Resources Guide, discussed below. In terms of target audiences toward which materials are directed, the DMTC provides the means to associate educational framework categories and their nodes to individual resources so that a learner can find resources in specific categories. Examples of the educational frameworks to which DMTC resources may be associated include the FAIR principles themselves, DataONE’s research life cycle, and the ICSU World Data System’s Training Resources Guide, discussed below. In terms of target audiences toward which materials are directed, the DMTC uses a controlled vocabulary that includes gradations of researcher (e.g., early career, mid-career) and other audiences for those involved in the research life cycle, such as data managers, data librarians, funders, citizen scientists, and educators.

Another useful source of information about available training resources is the ICSU World Data System International Programme Office’s (WDS-IPO) Training Resources Guide for WDS members and non-members. This resource guide provides information on the Core Trustworthy Data Repositories requirements to which a training resource refers, the types of materials included, whether a fee is required for access, and other information. While many of the resources in the guide have also been added to the DMTC mentioned above, more information about the training resources guides and the resources identified in it can be found at https://www.icsu-wds.org/services/training-resources-guide/training-resources-guide.

The Belmont Forum Data Management Toolkit (https://bfe-inf.github.io/toolkit/bp&s-root.html) offers links to basic, practical data management training materials, including MOOCs, professional certifications, and virtual or live trainings. The materials are designed for an audience that is relatively unfamiliar with data management concepts. Users can search for materials on the basis of professional role, geographic region, organization, or subject. The DataONE Skillbuilding Hub (https://dataoneorg.github.io/Education/) provides a comprehensive set of modules that are related to all phases of the data life cycle. The emphasis for these materials is to explain and promote best practices. Teachers, trainers, and users are encouraged to use what they find or contribute what they have developed and can

Figure 1. Description of actors in the EOSC ecosystem and related skills needs
Reproduced from Barker et al.27
contribute materials to this community-developed resource, hosted on GitHub.

ELIXIR, the European data infrastructure for the life sciences, offers the Training Portal TeSS (https://tess.elixir-europe.org/). TeSS aggregates life-sciences-specific training materials and events produced by a variety of providers and makes them searchable by author, keyword, difficulty level, target audience, and tool. TeSS scrapes the information automatically so no manual input is needed. Furthermore, TeSS supports the bioschemas metadata schema (https://bioschemas.org/), and is also connected to the FAIRsharing resource on data and metadata standards, and data policies (https://fairsharing.org/).

The FOSTER (Facilitate Open Science Training for European Research) portal at https://www.fosteropenscience.eu/ provides an extensive list of materials on open science, RDM, text and data mining, and responsible research and innovation.

EIFLnet (Electronic Information for Libraries) provides materials targeted to countries in Africa, Asia Pacific, Europe, and Latin American on topics such as data licensing and open access at http://www.eifl.net/.

LIBER Europe is an organization that works toward providing materials on digital skills useful for researchers and librarians at https://libereurope.eu/strategy/digital-skills-services/digitalskills/.

Resources include a DMP (Data Management Plan) Catalogue, a Job Description Repository, and a number of webinars on topics such as copyright and RDM.

The EOSC training portal (https://eosc-portal.eu/services-resources/training) provides a wide variety of links of various EU networks, including CompBioMed, the EGI, INFRAFRONTIER, OpenMinTed, OpenAire, PRACE, BlueBRIDGE, and GÉANT.

Training in data skills will also be provided by EOSC-life (http://www.eosc-life.eu/).

EOSC-Pillar provides an RDM Training and Support Catalogue (https://www.eosc-pillar.eu/rdm-training-and-support-catalogue) that is described according to the FAIR4S framework. The catalog provides online searchable resources for data stewardship and RDM support, targeting data stewards and professional support groups in Austria, Belgium, France, Germany, and Italy.

There are number of FAIR-specific resource sources being planned for development in the near future. These include the FAIRsFAIR collaboration (https://www.fairsfair.eu), which aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. One of the many goals of FAIRsFAIR is to develop a competency center in FAIR standards to provide links to other resources. Another project, FAIRplus (https://fairplus-project.eu/), plans to provide training modules as part of its fellowship program. The EOSC Executive Board Skills and Training Working Group provides a description of professional roles necessary for FAIR,27 the need for competence centers, and the provision of metadata for training materials aggregated from providers of training sources across the landscape.

**RESULTS**

**Analysis of current training materials sources**

As noted above, FAIR4S provides a means to map out what is covered in the developed guides. Key words and phrases from the category text listed in detail in Table 1 of the appendix were used to search for relevant materials in the developed guides. For example “visualisation” as a key word provides results in the DMTC and hence is labeled here as having overlap with the Expose and Discover category. Variants of key words (e.g., “visualization,” “visualize”) were also used. It is important to note that this does not indicate how thoroughly a category is covered. For example, “ontologies,” which is another key word in that Expose and Discover category, does not return results from the DMTC. Furthermore, this analysis does not attempt to collate what fraction of the total number of materials listed in the sources match up with a specific key word or the depth to which the materials review the topic. Likewise, we take a liberal approach to including sources in the different properties; e.g., target audiences. We say yes to a source if it has some overlap with the criteria. With respect to target audiences, we performed a search of each of the actor terms described in Figure 1 and have listed them in the field.

The findings are summarized in Table 2.

**Data and code availability**

No code was developed for this paper. The data that have been gathered are in Table 2.

**DISCUSSION**

The training of FAIR-related skills has made significant progress since the term was first coined in 2016. We note that this progress has not been driven by any one centralized effort but rather through multiple, often coordinated activities. We anticipate that further steps in this area will continue to have this characteristic, particularly as more research domains engage with the FAIR principles. Further FAIR-related training development will, undoubtedly, be related to training activities for RDM, data preservation, and data science. While training will exist that is entirely focused on FAIR skills, that training is and should continue to be taught within the context of the areas mentioned above as well as from the perspective of specific research domains.

Despite the progress described above, training in FAIR-related skills will need to be scaled up considerably in the near future. While much effort has been expended by the organizations and projects already mentioned, still more needs to be done in several areas:

**Adaptation of skills frameworks into curricular frameworks**

Promising educational frameworks such as that of the FAIR4S framework, which provides a broadly described but encompassing description of the set of skills deemed necessary to make and to keep data FAIR need to be adapted. A next logical step would be to develop a FAIR-focused **curricular** framework that can be applied to both formal and informal training materials, and to training events. A framework such as this would substantially improve the ability to search the sources of training materials for FAIR-related skills in order to build a coherent training program. Efforts in this direction are being carried out by the terms4FAIRskills initiative (https://terms4fairskills.github.io). Related efforts toward harmonization of the competences for FAIR data stewardship are being undertaken in the FAIRsFAIR project, which plans a FAIR Competence Framework for Higher
Edison, and other sources described in this paper.

More consistent and complete descriptions
It would be very helpful if the description of both training materials and training events could be done more consistently by the sources of those training materials and events so that they are more discoverable but also provide information important for making decisions about which materials or events to choose. As we see from the analysis in Table 2, the DMTC, FOSTER, and EOSC Pillar address all of the categories, while each of the other sources addresses one or more of the categories. In terms of the other information for a training material searcher, the target audience to which a resource is directed, the Belmont Forum, TeSS, the EOSC Marketplace, and EOSC Pillar do note the audience. The DMTC notes the educational framework to which the resource is targeted. This is important information, such as an indication of the level of the training material, for which target audience they are intended, and the specific learning objectives of the materials. Of course, materials’ creators have a role to play here as well. Given the fact that training material creators do not always include the design context for their resource by describing learning objectives, expertise level, and target audiences, for example, it is not surprising that this information is not often available to include explicitly within resource descriptions. Efforts to improve the lack of consistent description are being made by the Research Data Alliance’s Education and Training on Handling Research Data Interest Group (ETHRD IG), which has two subgroups focusing on recommending both minimal metadata and extended documentation for learning resources (https://www.rd-alliance.org/group/education-and-training-handling-research-data-ig/wiki/ethrd-ig-focus-group-materials). The latter effort aims to identify a common set of metadata fields drawn from the disparate standards developed in the open educational resources (OER) community, and based on analysis of use cases for learners, trainers, and other stakeholders. This should enable a more harmonized adoption of common metadata by content creators/adapters and by catalogs and registries.

For those resources with a target audience, the following keys from the actors described in Figure 1 are used: R, researcher; C, citizen; PM, policy maker; DL, data librarian; DC, data curator; DS, data steward; E/T, EOSC educator/trainer; RI, data RI support professional; RSE, research software engineer; DA, data scientist data analytics; EE, EOSC enabler.

Education that will map to and synthesize FAIR4S, EDISON, and other sources described in this paper.

Table 2. Analysis of training material sources

| Resource | URL | FAIR4S skills groups addressed (number of skills groups covered) | Target audiences | Learning objectives | Levels |
|----------|-----|---------------------------------------------------------------|------------------|------------------|--------|
| DMTC     | http://dmtclearinghouse.esipfed.org | PD, CP, IA, AP, PR, ED, GA, SR, AE (9/9) | Yes: R, C, PM, DL, DC, DS, E/T | No (only in description/abstract if available) | No |
| WDS-IPO Training Resources Guide | https://www.icusu-wds.org/services/training-resources-guide/training-resources-guide | PD, CP, AP, ED (4/9) | No | No | No |
| Belmont Forum Data Management Toolkit | https://bfe-inf.github.io/toolkit/bp&s-root.html | PD, CP, IA (3/9) | Yes: R, C, PM, DL, DC, DS, E/T | No | No |
| DataONE Skillbuilding Hub | https://dataoneorg.github.io/Education/ | PD, CP, IA, AP, ED (5/9) | No | Yes | No |
| ELIXIR Training and Events Portal (TeSS) | https://tess.elixir-europe.org/ | PD, CP, IA, ED, SR, AE (6/9) | Yes: R, DC, DS, E/T, RSE, DA | No | Yes |
| FOSTER | https://www.fosteropenscience.eu/ | PD, CP, IA, AP, PR, ED, GA, SR, AE (9/9) | No | No | No |
| EIFLnet | http://www.eifl.net | PD, PR, ED, SR, AE (5/9) | No | No | No |
| EOSC Marketplace Training Catalogue | https://marketplace.eosc-portal.eu/services/c/training-support | CP, AP, IA, GA, SR (5/9) | Yes: R, PM, RI | No | No |
| EOSC Pillar RDM Training and Support Catalogue | https://www.eosc-pillar.eu/rdm-training-and-support-catalogue | PD, CP, IA, AP, PR, ED, GA, SR, AE (9/9) | Yes: R, DL, DS, RI, DA, EE | No | Yes |

More consistent and complete descriptions
It would be very helpful if the description of both training materials and training events could be done more consistently by the sources of those training materials and events so that they are more discoverable but also provide information important for making decisions about which materials or events to choose. As we see from the analysis in Table 2, the DMTC, FOSTER, and EOSC Pillar address all of the categories, while each of the other sources addresses one or more of the categories. In terms of the other information for a training material searcher, the target audience to which a resource is directed, the Belmont Forum, TeSS, the EOSC Marketplace, and EOSC Pillar do note the audience. The DMTC notes the educational framework to which the resource is targeted. This is important information, such as an indication of the level of the training material, for which target audience they are intended, and the specific learning objectives of the materials. Of course, materials’ creators have a role to play here as well. Given the fact that training material creators do not always include the design context for their resource by describing learning objectives, expertise level, and target audiences, for example, it is not surprising that this information is not often available to include explicitly within resource descriptions. Efforts to improve the lack of consistent description are being made by the Research Data Alliance’s Education and Training on Handling Research Data Interest Group (ETHRD IG), which has two subgroups focusing on recommending both minimal metadata and extended documentation for learning resources (https://www.rd-alliance.org/group/education-and-training-handling-research-data-ig/wiki/ethrd-ig-focus-group-materials). The latter effort aims to identify a common set of metadata fields drawn from the disparate standards developed in the open educational resources (OER) community, and based on analysis of use cases for learners, trainers, and other stakeholders. This should enable a more harmonized adoption of common metadata by content creators/adapters and by catalogs and registries.

Other guidance that trainers can follow are the steps described by Garcia et al. to make the training materials they create FAIR, particularly as they evolve with use or adaptation.

Expanding the target audiences
As different domains adopt FAIR principles, there needs to be a greater range of materials created that are relevant to those domains. Likewise, existing materials are generally drawn from
high-income countries and there need to be more examples drawn from low- and middle-income countries. Also at this point, most materials are created in English, and there is great need to either translate or create materials in other languages. This will need to be carried out in higher education and research institutions, but also in the context of both formal and informal CPD training for data professionals.

Given the need for robust tools and processes to make data FAIR, it also makes a great deal of sense to work more closely with business, non-governmental, and governmental organizations to adapt educational frameworks, including curriculum, competency, and skills and capabilities frameworks, hands-on training, and FAIR-enabling tools, to the workflows and data they produce. Adding a non-academic context to the FAIR training equation will increase the chances of varied audiences, but hopefully one or more of the educational frameworks will have enough success to eventually allow for similar approaches to common problems. For example, given the need for persistent identification throughout the application of FAIR principles to data, a common understanding of which entities need identification (and why) could be shared with discussion happening more frequently instead around which identifier schemes to use (per domain and/or data format). Ideally, FAIR training and education topics and approaches will become as ubiquitous as topics and methods for teaching statistics or linear algebra.

Developing a FAIR community of practice for instructors
Another facet of critical importance to facilitate the standard teaching of FAIR skills is the development of a widespread community of practice of FAIR trainers. The initial efforts supported by OpenAIRE (previously cited) are very promising, although the difficulties associated with collaboration over many time zones and geographic locations must be taken into consideration. Realistically, it may be necessary to develop communities of practice for groups of related disciplines, such as those of the life sciences, or of Earth sciences, but communication among them on core issues related to effective and inspirational capacity building for data skills and successful adaptation to localized language, cultural, and geographic regional needs that include FAIR data principles, methods, and tools would help move this field forward toward the development of community-based standards. In addition, it would be very helpful for both instructors and learners to have more training opportunities, such as training events, but also to provide hands-on opportunities to learn and demonstrate competencies for both current and aspiring data professionals. Of special importance are the needs for opportunities (and further research and tools) to better enable research data to be interoperable and re-usable, particularly using machine-actionable methods. The number of tools available to do this\textsuperscript{30,31} will continue to expand.

Conclusion
For FAIR training and education topics to become more standard, there needs to be a clear understanding of the skills required. Thanks to efforts such as FAIR4S and others mentioned above, that understanding is becoming more widespread. In addition, rich collections of educational materials are becoming available to use. Finally, multiple initiatives currently exist and will develop, hopefully, to provide training resources and learning opportunities that will inform appropriate and effective pedagogical methods in this area.

This paper provides an outline of the resources to teach FAIR-related skills. In particular, it has reviewed the educational frameworks that are relevant for FAIR-related skills, sources for educational materials, and domain-independent training being provided in this area.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.patter.2021.100324.

ACKNOWLEDGMENTS
H.S. acknowledges the support of the Research Data Alliance to attend their plenaries.

AUTHOR CONTRIBUTIONS
All authors contributed equally.

DECLARATION OF INTERESTS
The authors declare no competing interests.

REFERENCES
1. FORCE11 consortium. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing Version b1.0. https://www.force11.org/fairprinciples.
2. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.
3. OECD (2020). Building Digital Workforce Capability and Skills for Data-Intensive Science, OECD Science, Technology and Industry Policy Papers, No. 90 (OECD Publishing). https://doi.org/10.1787/e08aa3bb-en.
4. PwC EU Services (2019). Cost-benefit analysis for FAIR research data: policy recommendations. https://publications.europa.eu/en/publication-detail/-/publication/d3/7e/d4/1a9f-1b10-w101-74ebce1f/language-en.
5. UNESCO Institute for Statistics (2012). International Standard Classification of Education: ISCED 2011 (UNESCO Institute for Statistics). http://www.uis.unesco.org/education/Documents/isced-2011-en.pdf.
6. The CPD Standards Office. What is CPD?. https://www.cpdstandards.com/what-is-cpd/.
7. Whyte, A., Leenarts, E., de Vries, J., et al. (2019). Strategy for Sustainable Development of Skills and Capabilities, EOSCpilot D7.5. https://eoscpiilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities.
8. Wilson, G. (2016). Software Carpentry: getting scientists to write better code by making them more productive. Comput. Sci. Eng. 8, 66–69. https://doi.org/10.1109/MCSE.2006.122.
9. Teal, T.K., Cranston, K.A., Lapp, H., White, E., Wilson, G., Ram, K., et al. (2015). Data Carpentry: workshops to increase data literacy for researchers. Int. J. Digit. Curation 10, 135–143. https://doi.org/10.2218/jdjc.v10i1.301.
10. Baker, J., Moore, C., Priego, E., Alegre, R., Cope, J., Price, L., et al. (2016). Library Carpentry: software skills training for library professionals. Liber Q. 26, 141–162. https://doi.org/10.18352/lq.10176.
11. Simons, N., Otsuji, R., Labou, S., Johnson, R., Castelao, G., Boas, B.V., et al. (2019). Top 10 FAIR Data & Software Things (Utrecht University Repository). https://doi.org/10.5281/zenodo.2595498.
12. Jotanovic, D., and Sumner, D.T.D. (2018). The Digital Research Skills Cookbook: An Introduction to the Research Bazaar Community (Research Platforms Services), isbn: 978149868521.

13. Garcia, L., Batut, B., Burke, M.L., Kuzak, M., Psomopoulos, F., Arcila, R., et al. (2020). Ten simple rules for making training materials FAIR, PLoS Comput. Biol. 16, e1007854. https://doi.org/10.1371/journal.pcbi.1007854.

14. Hodson, S., and Shanahan, H. (2017). The objectives, scope and activities of a possible GO-TRAIN implementation network. https://doi.org/10.5281/zenodo.1200276.

15. Molloy, L., Gow, A., and Konstantelos, L. (2014). The DigCurV curriculum framework for digital curation in the cultural heritage sector. Int. J. Digit. Curation 9. https://doi.org/10.2218/ijdcl.v9i1.314.

16. Kim, J. (2015). Competency-based curriculum: an effective approach to digital curation education. J. Educ. Libr. Inf. Sci. 56. ISSN: 0748-5786 © 2015 Association for Library and Information Science Education. https://doi.org/10.12783/jis.2015.56.4-2.

17. Murphy, F. and Lucas, V. 2017. The Belmont Forum, Open Data and ‘Human Dimensions’: lessons learned, ongoing investigations and developing recommendations for the future. In EGU General Assembly, Geophysical Research Abstracts, 2017 19, EGU2017-19186–1.

18. Demchenko, Y., Belloum, A., Los, W., Wiktoris, T., Manieri, A., Brocks, H., et al. (2016). EDISON data science framework: a foundation for building data science profession for research and industry. In Proceedings of IEEE International Conference on Cloud Computing Technology and Science (CloudCom), pp. 620–626.

19. Principe, P., André, V., Stoy, L., Engelhardt, C., Saenen, B., Davidson, J., Whyte, A., Demchenko, Y., Herterich, P., and Lennarts, E. (2020). D7.2 Briefing on FAIR competences and synergies. https://doi.org/10.5281/zenodo.4009007.

20. (2020). European Commission DG research & innovation 2017: EOSC declaration. https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud.

21. Sapp Nelson, M.R. (2017). A pilot competency matrix for data management skills: a step toward the development of systematic data information literacy programs. J. EScience Librarian. 6. https://doi.org/10.7191/jeslib.2017.1096.

22. Rans, J., and Whyte, A. Using RISE, the research infrastructure self-evaluation framework’ v.1.1 Edinburgh: digital curation centre. http://www.dcc.ac.uk/resources/how-guides.

23. Demchenko, Y., Belloum, A., and Wiktoris, T. (2016). EDISON Data Science Framework: Part 1 (Data Science Competence Framework (CF-DS)), zenodo.4009007.

24. Molloy, L., Biernacka, K., Jetten, M., and Nurnberger, A. (2016). https://doi.org/10.5281/zenodo.3471707.

25. O’Carroll, C., Rentier, B., Cabello Valdes, C., Esposito, F., Kaunisimaa, E., Maas, K., Metcalf, J., McCallister, D., and Vandevelde, K. (2017). Evaluation of research careers fully acknowledging open science practices. https://ec.europa.eu/research/openscience/pdf/oas_rewards_wgreport_final.pdf. doi:10.2777/75255.

26. Scholtens, S., Jetten, M., Böhmer, J., Staiger, C., Sluwerhof, I., van der Geest, M., and van Gelder, C.W.G. (2019). Towards FAIR data steward as profession for the life sciences. Report of a ZonMw funded collaborative approach built on existing expertise. Zenodo. https://doi.org/10.5281/zenodo.3471707.

27. Barker, M., et al. (2021). Digital Skills for FAIR and Open Science: Report from the EOSC Executive Board Skills and Training Working Group (Publications Office of the European Union), https://doi.org/10.5281/zenodo.4009007.

28. DCMi: LRMI Version 1.1. https://www.dublincore.org/specifications/lrmi/1.1/.

29. Schema.Org. https://schema.org/.

30. Jones, S., Pergi, R., Houtf, R., Miksa, T., Samors, R., Ungvari, J., et al. (2019). Data management planning: how requirements and solutions are beginning to converge. Data Intell. 2. https://doi.org/10.1162/dint_a_004319.

31. Jacobson, A., Kallyperumal, R., da Bonino, L.O., Mons, B., Schultes, E., and Thompson, M. (2019). A generic workflow for the data FAIRification process. Data Intell. 2. https://doi.org/10.1162/dint_a_00028.

About the Authors

Angus Whyte (a.whyte@ed.ac.uk) is a senior research data specialist at the Digital Curation Centre, University of Edinburgh. He specializes in producing training and guidance materials on research data management, currently as task leader in EU project FAIRsFAIR. He contributes to a number of working groups, including the Research Data Alliance’s interest groups on Professionalising Data Stewardship (PDS-IG), Education and Training on Handling of Research Data Interest Group (ETHRD IG), and Active Data Management Plans. He has also contributed to the EOSC Skills & Training WG report and led work on the FAIR4s skills framework in EU project EOSCpilot.

Nancy J. Hoebelheinrich (rheobel@kmotifs.com, https://www.linkedin.com/in/nancy-hoebelheinrich-0576ba3) is a digital library, archives, and data repository consultant specializing in geospatial, cultural heritage, and educational metadata. Nancy is founder and principal of Knowledge Motifs LLC (www.kmotifs.com). Nancy has been active in a variety of information and educational technology efforts including, most recently, those of the Earth Science Information Partners, the Research Data Alliance’s Education and Training on Handling of Research Data Interest Group (ETHRD IG), and EarthCube. Nancy is Co-I, Project Coordinator and Editor of the IMLS National Leadership Grant-funded Data Management Training Clearinghouse, and a member of the core GO FAIR US team.

Hugh Shanahan has a research background in computational biology and theoretical physics. He completed his PhD in 1994 in Lattice QCD and completed postdocs in Glasgow, Cambridge, and Tsukuba before moving into bioinformatics in 1999. In 2005 he joined the department of computer science at Royal Holloway, University of London where he is now professor of Open Science. Since 2015 he has been a co-chair of the CODATA-RDA schools in research data science that has delivered training in data science methods for researchers to students from approximately 40 countries. He is a member of the FAIRsFAIR consortium.