Project Title	**Fostering FAIR Data Practices in Europe**
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

D6.7 Report on schools run through franchising with local organisers

Work Package	WP6
Lead Author (Org)	Hugh Shanahan (RHUL)
Contributing Author(s) (Org)	Elizabeth Newbold (STFC), Joy Davidson (DCC)
Due Date	28.02.2022
Date	18.02.2022
Version	1.0
DOI	10.5281/zenodo.6043906

□ CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Authors	Notes
0.0.1	17.11.2021	Hugh Shanahan	Initial skeleton
0.5	25.11.2021	Hugh Shanahan	Draft of text without references or executive summary
0.9	30.11.2021	Hugh Shanahan	Draft for internal review
1.0	20.12.2021	Hugh Shanahan	Content ready

Disclaimer

Abbreviations and Acronyms

CODATA	The Committee on Data for Science and Technology
ECR	Early Career Researcher
EOSC	European Open Science Cloud
EOSC 5B	European Commission H2020 INFRAEOSC-05-2018-2019 projects
FAIR	Findable, Accessible, Interoperable, Reusable
HEI	Higher Education Institution
ICTP	International Centre for Theoretical Physics
RDA	Research Data Alliance
RI	Research Institution
VLE	Virtual Learning Environment

Executive Summary

The CODATA-RDA schools have since 2016 provided training to Early-Career Researchers, particularly from Low and Middle Income Countries, through a specific model of delivering training in a variety of different areas (e.g. São Paulo, Brazil; Addis Ababa, Ethiopia; Kigali, Rwanda). The number of schools run per annum has expanded slowly since then, with the primary rate-limiting steps being dependent on volunteer network and having access to mostly *ad hoc* funding that is focussed on individual schools rather than on the central organisation.

Based on the successful experience of running regional versions of schools where local[1] organisers make as much use of the local instructor capacity as possible, the concept of franchised schools, which are run by a separate project, e.g. one of the EOSC 5B projects, or a community that is based in Europe, was explored. A successful implementation of franchised schools within the FAIRsFAIR project would indicate that they could be run on a more financially sustainable footing beyond the lifetime of the project and hence in the longer term provide sustainable financial support for the running of the schools in Low and Middle Income Countries.

During the FAIRsFAIR project it was found that there was little interest in taking up this offer. Much of this is due to the pandemic as focus shifted to online delivery. On the other hand, it is apparent that offering a complete course for an organisation to take up that already has a commitment to provide its own teaching is less attractive than we had thought.

With this in mind, on top of the work of transitioning to online delivery, which has proved to be quite a positive experience, further work has been carried out on examining the financial sustainability of the schools.

There are a number of recommendations to make. In particular, teaching components (materials, curriculums etc.) should be as small as possible to encourage integration with other types of teaching. Creating new online materials or transitioning to online is not trivial and needs materials and planning. Making teaching materials FAIR will increase the visibility and use of those materials. Finally, making materials accessible (in the sense of accessibility for those with disabilities) and translating materials to widely used languages will also increase their use.

---

[1] each of these schools are hosted at one institution drawing students in from across that city, country and even neighbouring countries; the term local is reserved for the materials provided by that institution and regional is reserved for where the students are coming from.

Table of contents

1. Outline of franchising approach

The CODATA-RDA schools have run since 2016. The emphasis for these schools is to provide early-career researchers (ECRs) with a broad foundation in the principles of Data Science and to be more efficient and effective in their research. The school curriculum as a result covers a broad range of topics, from purely technical topics (the UNIX command line, git, Machine Learning, Computational Infrastructures, Author Carpentry, Information Security) to more social components (Research Data Management and Open and Responsible Research). In a physical setting this corresponds to two intensive weeks combining lectures and labs. In a virtual setting, this runs over 11 weeks with materials for students to work on with two live hours per week and questions being addressed on a forum.

The model for these schools is to run, on an annual basis, a school that is hosted by the International Centre for Theoretical Physics (ICTP) in Trieste which acts as a hub to test new developments in the curriculum and to also run an increasing number of regionally-based schools on a global basis. The regionally run schools are run in close partnership with local organisers. In particular, there is a principle of using as many locally-based instructors as possible teaching the defined curriculum of the schools. Local instructors can use as much of the materials developed for the curriculum as they see fit but can also develop or adapt the school's materials accordingly. To ensure continuity and support for local organisers all regional schools have at least one co-chair from the central organisation present. This has proved very effective in terms of minimising costs for the central organisation (namely the CODATA-RDA schools) to run these events and demonstrating the local application of these materials.

This model has enabled a certain amount of growth in the number of schools that have been run. In the absence of the pandemic there is likely to have been further linear growth in the number of schools in 2020, 2021 and beyond. The FAIRsFAIR project provided key support in running the first physical school and Data Steward instructor training event in 2019 and then supporting the virtual events in 2020 and 2021. It has also provided support to transition the materials to online delivery (to be discussed in detail in the FAIRsFAIR report D6.5: Report on three annual schools in core data skills for researchers).

Table 1. Regionally run CODATA-RDA schools

Year	Number of regionally run schools
2017	1 (São Paulo, Brazil)
2018	2 (São Paulo, Brazil; Kigali, Rwanda) + 1 shorter version (Brisbane, Australia)
2019	2 (Addis Ababa, Ethiopia; San José, Costa Rica)

2020	1 (Pretoria, South Africa)
2021	1 (Pretoria, South Africa - run virtually)

The work carried out by the central organisation, prior to FAIRsFAIR, is volunteer-based and is a key rate-limiter in the expansion of the number of schools in terms of organising the logistics. In addition funding for regional schools happens on an *ad hoc* set of grants and commitments. Further expansion requires a long-term, financially sustainable, model.

With this in mind a strategy was in place in the first year of FAIRsFAIR to run the schools and Data Steward Instructor training; Schools would continue to be run in Trieste over the duration of the project and Data Steward instructor training would be run in parallel with the schools, enabling a more efficient delivery of training and ensuring that two communities - of ECRs and Data Stewards - would have the opportunity to see how the other community works.

A variety of different EOSC-related projects, such the EOSC 5B projects and other organisations such as ELIXIR would be approached to teach the school curriculum. The additional Data Steward instructor training required from the FAIRsFAIR project could be likewise embedded into the franchised schools. The term franchising is used here as this would be a proof of concept of producing a sustainable financial model beyond the lifetime of the project, where schools run in High Income Countries could subsidise those run in Low and Middle Income Countries.

2. Implementation

2.1. Implementation outside Europe

The University of Pretoria in 2020 ran a version of the school where it took a much higher level of responsibility for organising the school teaching and made a number of small adjustments to the delivery (for example running parallel sessions for natural and social scientists in R). From April-June 2021, the same organisation ran an ambitious programme for the virtual delivery of the school. A nominal fee, used to pay incidentals, encourage attendance, and waived for many students, was charged. The model from the University of Pretoria (not including the nominal fees) was adopted by the central organisation to run the virtual version of the school that was run for FAIRsFAIR from September-November 2021. This is not matching the model of franchising described above but is an excellent example where a separate team becomes more autonomous in the delivery of the materials.

2.2. Implementation within Europe

Over the duration of the project a number of conversations were held with potential partners, in particular, EOSC Pillar, EOSC Nordic and ELIXIR. The pandemic had a serious impact on the further

development of these conversations. In particular the majority of time from the project partners was spent on transitioning the schools and Data Steward Instructor training to an online basis. The pandemic and transitioning to online delivery was not the only factor in running more of the schools (see section 2.1). The pilot school in 2019 which ran both the ECR school and Data Steward Instructor training was successful but it was clear that Data Stewards working full time could not commit the equivalent of two weeks of time on this type of training and hence it was necessary to construct a more compact form of training which prioritised specific aspects of their role as a Data Steward. Hence more time was also spent in terms of a major redesign of that curriculum in consultation with other partners, namely the institutions that hosted the Data Steward Instructor training events. Finally, prospective partners, particularly those in funded projects which already have a training budget, have an expectation to deliver materials that are appropriate to their own context and hence are unwilling to simply rerun instances of a prepared curriculum.

3. Lessons learnt

3.1. Impediments to adoption

As noted in section 2 projects that already have a requirement to develop training are not in an obvious position to simply take on a curriculum and teaching model that is already in place. Such organisations can show resistance to simply to taking up complete solutions such as this. Partnerships between projects (such as the partnership between FAIRsFAIR and EOSC Synergy on the development of the Data Steward Instructor training programme) are also likely to have more success but don't address the take up of already developed materials.

3.2. Transition to online

Despite the significant amount of work required to make full use of online delivery, the experience has been largely positive and enables a potentially much greater reach. Instructors need to spend a significant amount of time preparing the materials. Helpers are necessary to a) field questions on forums on a regular basis during the training b) act as moderators or facilitators in break out rooms and c) oversee possible questions raised in chat channels during live sessions. In particular, groups can participate in training if they have their own sets of helpers.

3.3. Sustainability

As noted in section 1, the long-term financial sustainability and scalability of the schools was the main purpose of introducing the franchising approach. It is clear that this approach is not likely to succeed as a viable option. In this light, steps have been taken to instead improve the overall governance of the schools by setting up an Advisory Board (AB) and running a workshop explicitly on sustainability models in February 2022. The prepared documents and notes from this workshop will be outputs of FAIRsFAIR. The AB will be composed of 9-15 members from a variety of different backgrounds. There will also be two ex-officio members representing CODATA and the RDA. The AB

will meet twice a year, with the first meeting soon after the workshop, and membership being a minimum of two years.

4. Recommendations

4.1. Developing materials for online teaching requires planning, skilled trainers and helpers

Although face to face training is likely to see some resurgence after the pandemic, online training will remain a key method for delivery. Some questions remain over online teaching, such as poorer engagement and concerns over connectivity. Nonetheless the advantages of online training are apparent. It can be more cost-effective than face to face teaching, though as we will note below it is certainly not cost free. It is potentially much more scalable. Materials that can be delivered completely asynchronously (e.g. a static set of materials on a Virtual Learning Environment (VLE)) potentially remove the need for any form of live interaction and represent an extension of documentation. An example of this is the MOOC, hosted at the EOSC-Synergy VLE, of the Data Steward instructor training of FAIRsFAIR.

Developing materials for online teaching is not trivial and not simply a matter of posting materials developed for face to face delivery. The simplest model of micro-lectures coupled with self-assessment exercises requires time from the instructor, familiarity with basic video-recording tools and post-editing of those videos. Hosting materials on a server is relatively inexpensive but configuring those services and maintaining them requires support time. Forums need to be regularly checked during a live training event. Team-based exercises in breakout rooms require thought and helpers (acting in a moderating role) to ensure that the attendees understand the assignment and that all members of the breakout room can feel that they can take part.

The time required for the transition of online materials from an original set of face to face materials is proportional to the time required for the development of the original materials and hence should be costed appropriately.

4.2. Small teaching components are better

As noted in section 2, the wholesale use of the schools model of teaching used during FAIRsFAIR represents a significant effort. With an enthusiastic and supportive local team, it is possible to take those materials and run them along the same lines as those done in the schools run during the project. Nonetheless, it requires a significant number of meetings and a close partnership with the original providers. The wholesale transfer of the model to another organisation or set of trainers is difficult. On the other hand, *teaching components,* (namely materials, lesson plans, assignments, etc.) are easier to take up by trainers outside of the original team. Smaller teaching components, which are closely aligned to specific learning objectives, allow trainers to easily insert them into their own materials or to adjust them to match their own context. For example, Carpentries

instructors[2] (Baker et al., 2016; Teal et al., 2015; Wilson, 2014) can and do adjust the standard materials provided by the Carpentries web site for the context that it is being taught in. Providing a more staged approach to teaching delivery represents good teaching practice.

This does not preclude longer form teaching components, but their design should aim to be modular with respect to learning outcomes.

4.3. Online teaching can scale and be run in parallel with disparate groups

The delivery of online training which has a live component depends on a hosting site for materials (e.g. a VLE or forum-based service), a forum for questions with a set of helpers who can answer the questions and conferencing software (e.g. Zoom or Microsoft Teams) for live webinars, Questions and Answer sessions or team-based assignments. *The scalability of this type of training can be achieved as the number of instructors required can remain static with the size of the cohort increasing as long as the number of helpers increases.* This was initially observed through H3Africa (Aron et al., 2021) and CodeRefinery's "Bring your own Breakout Room" model ("CodeRefinery - Lessons learned from the May 2021 online workshop," n.d.). Likewise, the last virtual CODATA-RDA school had ten MSc students from Addis Ababa University attending in parallel with local helpers present *in situ*. Very disparate groups can attend the same training as long as they have their own helpers to assist them.

4.4. Make teaching materials FAIR

Making teaching materials FAIR is not only an example of following best practice - FAIR teaching materials enable the reuse of those materials. The interpretation of how to apply the FAIR principles to teaching materials has already been discussed elsewhere (Garcia et al., 2020) but the necessary infrastructure to make teaching materials FAIR is rapidly improving. Catalogues of materials are already available but metadata standards such as the RDA Interest Group on Professionalising Data Stewardship and terms4FAIRskills ("Professionalising Data Stewardship IG," 2020; "terms4FAIRskills Initiative," n.d.) will improve their findability and hence such standards should be used in future for materials. Materials should be deposited in repositories on a regular basis even if the materials are available on a conventional web-site. This should also be carried out for objects such as materials exported from a VLE using standardised formats for an object. If VLE materials may be used by facilities with limited computing services, the possibility of also creating containers[3] that have the relevant VLE service, as well as materials, should also be considered.

---

[2] www.carpentries.org

[3] A container is a standard package of software which bundles an application's code together with the related configuration files and libraries, and with the dependencies required for the app to run. This allows developers and IT pros to deploy applications seamlessly across environments.

4.5. Enable diversity in sources

Materials are reused and adapted by research communities, Higher Education Institutions (HEIs) and Research Institutions (RIs). Such communities will find materials through their networks rather than one central location. Hence the focus should be on ensuring that materials being developed are discoverable, rather than building a single service that users must consult to find those materials. It is important to note that Low and Middle Income Countries may not have a range of language skills and hence developers should consider the possibility of spending funds to get materials translated to a set of languages (French, Spanish, Portuguese, Mandarin) to increase reach. The definition of accessibility in the sense of making the materials accessible to individuals with impaired vision, motor difficulties, cognitive impairments or learning disabilities, deafness or impaired hearing is increasingly a legal requirement across countries. Text materials developed should use accessible formats (e.g. HTML, presentation formats such as Powerpoint, flat file formats such as ASCII or Markdown but **not** PDF) and tested accordingly ("Understanding accessibility requirements for public sector bodies," n.d.). Video should have closed captioning, using automatic transcription services if necessary.

The above steps are implementable within a reasonable budget. Post-pandemic we have an increased palette of options for teaching delivery. The critical role training plays in initiatives such as EOSC is now recognised and the community of instructors should work with the full range of this palette.

2. Bibliography

Aron, S., Chauke, P.A., Ras, V., Panji, S., Johnston, K., Mulder, N., 2021. The Development of a Sustainable Bioinformatics Training Environment Within the H3Africa Bioinformatics Network (H3ABioNet). Front. Educ. 6, 356. https://doi.org/10.3389/feduc.2021.725702

Baker, J., Moore, C., Priego, E., Alegre, R., Cope, J., Price, L., Stephens, O., Strien, D. van, Wilson, G., 2016. Library Carpentry: software skills training for library professionals. Liber Q. 26, 141–162. https://doi.org/10.18352/lq.10176

CodeRefinery - Lessons learned from the May 2021 online workshop [WWW Document], n.d. URL https://coderefinery.org/blog/2021/11/25/lessons-learned-may-2021/ (accessed 11.28.21).

Garcia, L., Batut, B., Burke, M.L., Kuzak, M., Psomopoulos, F., Arcila, R., Attwood, T.K., Beard, N., Carvalho-Silva, D., Dimopoulos, A.C., Angel, V.D. del, Dumontier, M., Gurwitz, K.T., Krause, R., McQuilton, P., Pera, L.L., Morgan, S.L., Rauste, P., Via, A., Kahlem, P., Rustici, G., Gelder, C.W.G. van, Palagi, P.M., 2020. Ten simple rules for making training materials FAIR. PLOS Comput. Biol. 16, e1007854. https://doi.org/10.1371/journal.pcbi.1007854

Professionalising Data Stewardship IG [WWW Document], 2020. . RDA. URL https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig (accessed 11.28.21).

Teal, T.K., Cranston, K.A., Lapp, H., White, E., Wilson, G., Ram, K., Pawlik, A., 2015. Data Carpentry: Workshops to Increase Data Literacy for Researchers. Int. J. Digit. Curation 10, 135–143. https://doi.org/10.2218/ijdc.v10i1.351

terms4FAIRskills Initiative [WWW Document], n.d. . GitHub. URL https://github.com/terms4fairskills (accessed 9.12.21).

Understanding accessibility requirements for public sector bodies [WWW Document], n.d. . GOV.UK. URL https://www.gov.uk/guidance/accessibility-requirements-for-public-sector-websites-and-apps (accessed 11.17.21).

Wilson, G., 2014. Software Carpentry: lessons learned. F1000Research 3, 62. https://doi.org/10.12688/f1000research.3-62.v1