The Objectives, Scope and Activities of a Possible GO TRAIN Implementation Network

This short document **for discussion**, is the outcome of a workshop hosted by CODATA at the International Council for Science in Paris on Friday 3 February 2017.

Introduction

The objective of the workshop was 'to inform the activities and focus of a possible GO TRAIN Implementation Network (IN) as part of the wider GO FAIR initiative.'

GO FAIR is an early-mover-driven 'bottom up' initiative to start working on a trusted environment where public and private sector partners can deposit, find, access, exchange and re-use each other's data, workflows and other research objects. The core of GO FAIR is a federation of existing topical networks of excellence that collectively commit to the FAIR approach and capitalise on their critical mass to make choices in the implementation of the FAIR principles in terms of standards, protocol and best practices.

The workshop was structured around group discussion sessions addressing the following issues: 'the scope of data stewardship training; the modes of delivery, progression and assessment; and how the logistics (namely instructors, materials, etc.) would be implemented, coordinated and sustained. These topics provide the structure of what follows.

An effort was made to include in the workshop representatives of many of the major European (and global) initiatives relating to training in data skills and data stewardship. Our objective was to make a useful first step towards defining a GO TRAIN IN. Some important training initiatives were not represented: practical issues of numbers constrain all meetings. We do not pretend, therefore, that this paper presents a final, definitive view, or that it represents all important potential contributors to a GO TRAIN IN. Rather, it is a small practical step towards defining what such an IN would do. Efforts will be made to reach out and to discuss this paper and the further development of a vision for a GO TRAIN IN with a wider group of stakeholders and potential participants.

The preparation documents for the workshop used the term 'a GO TRAIN Node'. Since then, the GO FAIR initiative has adopted the term 'Implementation Network' and consequently this is used here. Further definition is still required (see Sections 4 and 5) but on the face of it, the notion of an Implementation Network fits better with the discussions at the workshop.

The agenda is provided in Appendix A and the list of participants in Appendix B.

Summary and Preliminary Conclusions

A GO TRAIN IN is conceived as the means to train those individuals who will make the best use of FAIR data services and the data stewards capable of providing those services.

1) **Gaps to be addressed:** A GO TRAIN IN will be most effective if it focuses on the following five roles or 'skills areas':
   - Skills needed by the research data specialist function based in institutions;
   - Skills needed by the institutional research data advisor and contact point;
   - Skills needed by the long term stewardship function;
   - Skills needed by all researchers.
   - Capacity building and skills management best practice for organisation management roles

2) **Approach to be taken:** A GO TRAIN IN should:
   - address the needs of researchers and of professionals in an increasingly non-traditional academic sector;
   - perform an advocacy function as well as training;
   - offer a coordination function;
   - provide a form of certification or endorsement of materials and activities;
   - promote and encourage the following training approaches: blended learning, train-the-trainers, supporting peripatetic career pathways and be adaptive to new and innovative mechanisms for skills acquisition.

3) **Organisations to be involved and their current activities:** The organisations that we hope will participate in a GO TRAIN IN are listed below with their current 'offers' in data training. Organisations that were not represented at the workshop but have expertise or an offer of significance for GO TRAIN are also listed. A matrix should be developed against an improved version of the typology of activities suggested.

4) **Functions and structure of a GO TRAIN IN:** The primary objective of the GO TRAIN IN is to provide a horizontal coordinating function in order to provide direction and to accelerate the widespread implementation of data training in the five areas identified. The optimal institutional arrangement needed to provide such a function requires further discussion. The five priority activities identified were:
   - Refine and focus requirements.
   - Signpost and endorse to assist discovery and adoption of materials, create or use existing training materials directories.
   - Training advice, career and capacity building.
   - Coordinated and scaling delivery of training, create and maintain trainers network.
   - Encourage and perform train-the-trainer activities.

5) **Funding and sustaining the IN:** Further work is required on the characteristics of the GO TRAIN IN as a real, virtual or consortium/federated organisation. The criteria for these characteristics relate both to the delivery of the IN's coordinating functions and to the possible business model to sustain those functions. A number of possible business models are listed.

It is clear that there is significant interest from the attendees in moving this forward; next steps involve reaching out to other interested parties and putting more detail into the discussion outlined below.

Detailed Workshop Recommendations

1) What are the roles or 'skills areas' that need most attention from a GO TRAIN IN?

The discussion group identified five roles or 'skills areas' in relation to which capacity needs to be increased. These roles or 'skills areas' were determined by considering the research process and asking: 'where can something effective be done to bridge the gap between where we are now and where we want to be in terms of effective data stewardship'. This is deliberately not the same as asking 'what are the skills required of data stewards?' In other words, our contention is that capacity for data stewardship is best improved by addressing the gaps in roles, as described below, rather than just by training 'data stewards' with a specific list of skills.

By the same token, the five 'skills areas' *could* largely correspond to particular roles in the research process and provided by institutions. However, we have retained the terms 'roles' and 'skills areas', because there are potentially a number of ways of ensuring that the functions required are performed.

From the discussions, the roles and skills areas were characterised as follows:

1. **Specialist embedded research data skills:** these are the specialist data skills required in most research groups and span the skills relating to the 'science of data' (data analysis visualisation, statistics etc) and those relating to data management (covering planning for curation, annotation and metadata etc). The important points here are that: 1) the skills are embedded in the research group, 2) the skills are specialised and not necessarily expected of all subject experts. The role must have sufficient subject knowledge and will contribute to research design specifically from the perspective of statistical analysis, database design, data management, data analysis and integration, visualisation, etc. The role or roles works at the interface between the subject objectives and the challenges of data science and data management. They could be performed by a 'data scientist' and a 'data manager' or by a combination of both.

2. **Institutional research data advisor and contact point:** this cluster of skills forms a role that may be filled by data librarians in a research institutions/universities, or by the 'front office' function in the Research Data Netherlands model[1]. The role provides advice to researchers on a range of data related topics (ethos of Open Science, funder and journal requirements, data management plans, data curation, discovery of data resources for access to data and for deposit, etc.).

3. **Long-term stewardship:** the skills that are required to see to the long term preservations, archiving, accessibility and availability of the data. These skills relate to the functions provided by trusted digital repositories; they correspond to the 'back office' function in the Research Data Netherlands model. There is a recognised need to develop these skills (and the associated functions) to address the long tail of digital research data (c.f. P.B. Heidorn, 2008, doi:10.1353/lib.0.0036) in universities and other research institutions.

4. **Data skills of benefit to all researchers:** all researchers now need a range of skills relating to digital data. In some cases these might be termed 'awareness-level' competencies:

---

[1] http://datasupport.researchdata.nl/en/start-de-cursus/vi-data-support/front-office-back-office/

knowing when and how to call on additional expertise. In other cases the skills required may go further than this. The demands of specialisation are such that it is not proposed that all researchers should also be expert data managers, data stewards and data scientists. But what some have called '**research data literacy**' is now a fundamental requirement and part of the core set of research skills that is needed by all researchers. Accordingly, curricula and training efforts have emerged which offer foundational data management and data science skills for *all* researchers.

5. **Skills management for organisations:** at an organisational level there is a need for managers to be aware of the above requirements and to play a role in advocating their importance within their organisations. Without support at this level, researchers will be stymied in their progress.

Each of these skills areas can be fleshed out with a series of topics. Those identified in the workshop are listed in Appendix C.

Conclusion: A GO TRAIN IN will be most effective if it focuses on the five areas described.

2) What are your desiderata for a GO TRAIN IN?

The discussion group considered what a GO TRAIN IN would do (i.e., what mechanisms (certification or endorsement, examination), what initiatives, what sort of coordination activity etc is needed) and came up with a list of desiderata covering the content, approach and mechanism of training. These should be considered in relation to the foci described above.

1) **Address Needs of Researchers within and outside the academic sector:** Need to think beyond researchers (in the university/grant-funded research performing organisation sector) and address other communities. This might suggest two strands of activities, one directed primarily at academic researchers, the other at research professionals in the government, private and SME sectors that would benefit from these skills. *More needs to be done on defining the priorities in the latter case.*

2) **Perform Advocacy:** Both within and outwith the core research sector, advocacy is of great importance and must not be neglected. Without sufficient advocacy there will not be uptake of the skills and training and the necessary change will not be achieved. In some instances, there is a 'fuzzy boundary' between advocacy and the 'need-to-know' style briefing or training designed for middle and upper management. GO TRAIN needs to coordinate a spectrum of activities, not just core provision of skills training. The relationship with GO-CHANGE will be important in this respect.

3) **Coordination Function:** A lot of initiatives are underway and a lot of materials already exist. GO TRAIN must resist the temptation to start things afresh. GO TRAIN *could* play a valuable role in providing mechanisms for coordination. It should do so by encouraging collaboration, alignment with current standards and best practices in training, reuse and adaptation of (properly meta-tagged) materials, rather than offering top-down direction.

4) **Certification or Endorsement Function:** One valuable way of assisting coordination, could be through an certification or endorsement function. GO TRAIN could provide various courses and activities and offer a lightweight endorsement as 'FAIR-ready' to those courses and also to existing courses. *The mechanisms for this need to be considered and feasibility-tested, but in principle such a function could assist coordination and the more widespread adoption of more effective training and skills development.*

5) **Training Approaches:** The GO TRAIN IN should address the five roles/skills areas described above. Three training approaches were recommended.

a) Blended learning: provision of online materials and courses is a necessary but not sufficient activity to achieve training at the scale required. Although online courses are very useful they are most effective when combined with face-to-face learning in a [blended learning](#) approach.

b) Train-the-trainers: because hands on and face-to-face activities remain important and cannot be entirely replaced by online courses there is an important challenge of increasing the number of instructors. A considerable train-the-trainer function will be essential. There are established Train the trainer programmes e.g in Software and Data Carpentry and in ELIXIR. These programmes use practitioner-based approaches such as Bring Your Own Data (BYOD) and paired learning.

c) Supported career pathways: as the challenge does not just consist of creating a discrete new profession of 'data stewards' or 'data scientists', but in increasing the skills of people performing new/additional functions within existing roles, then the career pathways for those who invest time in the acquisition of such skills needs to be considered and supported. A GO TRAIN IN can play an important advocacy function in this regard. The proposal here is not to create a new profession but to add skills which are best described as a combination of 'data stewardship' and 'data science' to existing roles and in particular to the 'specialist embedded research data skills' role.

Conclusion: GO TRAIN should:

1) address the needs of researchers and of professionals outside the traditional academic sector;

2) perform an advocacy function as well as training;

3) offer a coordination function;

4) provide a form of certification or endorsement of materials and activities;

5) promote and encourage the following training approaches: blended learning, train-the-trainers, supporting career pathways.

3) Which organisations do we hope will contribute towards GO TRAIN?

As noted above the meeting sought to bring together a number of organisations, initiatives and projects engaged with training activities and which could form an effective GO TRAIN IN. We recognise that not *all* the organisations involved in data skills training in Europe were represented.

Organisations Represented
The organisations or projects involved in training represented (though not simply including institutional affiliations) were:

CODATA-RDA Schools of Research Data Science, an initiative to provide a reusable course and a scalable network of schools providing foundational training in core data skills
http://www.codata.org/working-groups/research-data-science-summer-schools
CSC
CSC is a Finnish center of expertise in ICT that provides ICT expert services at an internationally high level of quality for research, education, culture, public administration and enterprises, to help them thrive and benefit society at large.
DANS
DANS (Data Archiving and Networked Services) is the Netherlands Institute for Permanent Access to Digital Research Resources. DANS encourages researchers to make their digital research data and related outputs Findable, Accessible, Interoperable and Reusable. They provide expert advice and certified services. Their core services are: DataverseNL for short-term data management, EASY for long-term archiving, and NARCIS, the national portal for research information. By participating in (inter)national projects, networks and research, DANS contributes to continued innovation of the global scientific data infrastructure. Open if possible, protected where necessary. DANS is an institute of the Dutch Academy KNAW and funding organisation NWO.
Digital Curation Centre
The DCC is a consortium run from the University of Edinburgh whose mission is to help others develop the skills and capacities to increase the value of data for reuse. It conducts research, consultancy, advocacy and training in support of these aims http://www.dcc.ac.uk/
DTL
The Dutch Techcentre for Life Sciences (DTL) is a public-private partnership of more than 50 life science organisations in the Netherlands. The majority of Dutch universities and university medical centres are DTL partners and a growing number of companies are joining the organisation.
DTL is organised as a network of experts and policymakers affiliated with the DTL partners. This DTL network is supported by a small facilitating team located in Utrecht. (Read more about the organisation.) DTL connects scientists, data experts, technical experts, and trainers that are specialised in a variety of high-end wet lab and data technologies, and working in life science domains ranging from health to nutrition, agro, biotech, and biodiversity. Together, these professionals interconnect and improve their research infrastructure to enable cost-effective cross-technology life science research in national and international collaboration.
EDISON - http://edison-project.eu
The EDISON project has been funded to address the significant challenge of ensuring that there is a sufficient number of appropriately trained experts in the field of data science to meet the needs of research infrastructures and other employers in Europe and beyond, now and in the future. The two-year project has been funded under the European Commission's Horizon 2020 research and

innovation programme for the European Union. Sustainability of the EDISON project outcomes will be ensured by the EDISON consortium members in cooperation with other interested parties.

EIFL - http://www.eifl.net/

EIFL (Electronic Information for Libraries) is a not-for-profit organization that works with libraries to enable access to knowledge in developing and transition economy countries in Africa, Asia Pacific, Europe and Latin America.

ELIXIR

ELIXIR is a European research infrastructure with a mission to manage and safeguard the increasing volume of data generated by life science research. It coordinates and sustains bioinformatics resources across its member states and help researchers to more easily find, analyse, share data and exchange biological data.The mission of the ELIXIR Training Platform is to establish an interactive, ELIXIR-wide training community to deliver ELIXIR-related training across Europe. ELIXIR Training targets developers, researchers and trainers within the ELIXIR Nodes, providing them with skills to effectively exploit the data, tools, standards and compute infrastructure offered by ELIXIR.

EUDAT - https://eudat.eu/

EUDAT's vision is *Data is shared and preserved across borders and disciplines.* Achieving this vision means enabling data stewardship within and between European research communities through a Collaborative Data Infrastructure (CDI), a common model and service infrastructure for managing data spanning all European research data centres and community data repositories.

GOBLET, a foundation established to cultivate the global bioinformatics trainer community, set standards and provide high-quality resources to support learning, education and training: www.mygoblet.org. In its mission to professionalise training, GOBLET articulated a Joint Training Strategy with ELIXIR, to align training portals, collaborate on train-the- trainer/-researcher initiatives, explore certification or endorsement mechanisms and share best practices: https://www.elixir-europe.org/news/elixir-and-goblet-publish-joint-training-strategy

LEARN

The EU H2020 LEARN project is building on the LERU Roadmap for Research Data to deliver:
1. A model Research Data Management (RDM) policy;
2. A Toolkit to support implementation of the policy, and;
3. an Executive Briefing in five core languages so as to ensure wide outreach.

LIBER - http://libereurope.eu/

LIBER (Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries) is the main network for research libraries in Europe.

They work to represent the interests of European research libraries, their universities and their researchers in several key areas. They lobby policymakers on issues such as Copyright and Open Access. They collaborate with their member libraries on European-funded projects and through events such as our Annual Conference we create opportunities for library professionals to meet and learn from each other.

FOSTER - https://www.fosteropenscience.eu/

FOSTER Plus (Fostering the practical implementation of Open Science in Horizon 2020 and beyond) is a 2-year, EU-funded project, carried out by 11 partners across 6 countries. The primary aim is to contribute to a real and lasting shift in the behaviour of European researchers to ensure that Open Science (OS) becomes the norm.

Research communities, research performing institutions, and research funders have each recognised that OS skills are increasingly essential for researchers to undertake responsible research and innovation. While there is increasing agreement around the need to improve OS skills amongst all stakeholders, the adoption of OS approaches has been quite limited to date. Indeed

general awareness of OS approaches has improved among EU researchers. However, there is still a lack of practical guidance and training to help researchers learn how to open up their research within a particular domain or research environment. For this reason, FOSTER Plus places specific emphasis on creating discipline-specific guidance and is partnering with expert organisations representing the scientific areas of life science, social science and humanities.

Open-AIRE - https://www.openaire.eu/
OpenAIRE is an EC-funded initiative which aims to support the Open Access policy of the European Commission via a technical infrastructure.

50 partners make up this collaborative effort working towards a common goal to bring a change in realising open science for the benefit of society, innovation and industry.

OpenAIRE currently operates an interoperable and validated network of more than 520 repositories and OA journals, integrating more than 9 million OA publications and 1,000 datasets, with 50,000 organizations and 30,000 projects from two funders. It has identified over 100,000 FP7 publications from about half the 26,000 FP7 projects, and offers literature-data integration services.

RDA IG on Data Education -
https://www.rd-alliance.org/groups/education-and-training-handling-research-data.html
The objective of this IG is the exchange of information about existing developments and initiatives and promotion of training/education to manage research data throughout the data lifecycle. Concretely, it will make the case for creating taxonomies of the skills required by different group of data management specialists/professionals and elaborating reference models as a basis to:

1. enable the setting of quality standards for appropriate education and training programmes aimed at researchers and the professionals that support them, at all career stages;
2. encourage the recognition of data skills amongst employees, employers, and professional bodies.
3. prepare the ground for practical applications applying these standards in educational environments

Existing Offers
The existing 'offers' of these organisations in respect to data training are given in Appendix D. This should be refined as a matrix in relation to the activities or functions listed below.

In discussions it was suggested that we could advance coordination and the construction of a GO TRAIN IN by preparing a matrix of the things that various organisations and project could bring to a coordinating activity. The preliminary typology of activities was listed as follows:

1. Advocacy
2. Materials
3. Training capacity
4. Train the trainer capacity
5. Access to disciplines (advocacy)
6. Access to research institutions (advocacy)
7. Review / Evaluate / Accredit
8. International networks

Organisations not represented that should be involved:

● European and national data infrastructures and initiatives: e.g., CESSDA, GESIS, UKDA and etc. for other disciplines;

- Domain-specific networks of excellence: e.g., ESFRIs (i.e., BBMRI, EPOS), GODAN, CERN, etc;
- e-infra organizations: EGI (they have another role, supporting infrastructure);
- University organisations: e.g., LERU, EUA, etc. (also clearly related to GO-CHANGE);
- Software Carpentry and Data Carpentry foundations
- Providers of (platforms for) large online courses: e.g., Coursera, Udacity;
- Commercial training companies;
- Statistical organisations that providing training.

Conclusion: Organisations that were not represented at the workshop but have expertise or an offer of significance for GO TRAIN should be contacted. A matrix should be developed against an improved version of the suggested typology of activities.

4) What would the GO TRAIN IN do?

The groups discussed 1) what the functions of a GO TRAIN IN optimally would be; and, 2) what sort of organisation would be required.

What would the optimal activities of a GO TRAIN IN be?
The task of determining the optimal activities of a GO TRAIN IN should take into account the focus on the five roles/skills areas described and the desiderata in terms of approach. It should also build on and refine the typology of activities developed above.

The workshop discussions identified the following five priorities for a GO TRAIN IN:

1. **Help refine and focus requirements:** Taking the typology of five gaps, the IN should *identify in more detail the skills, training framework and the existing materials*, and thereby provide a structure for a series of training modules to address the gaps.
2. **Signpost and endorse:** The IN should provide a *framework for discovery and certification or endorsement*. To do so it should build on or partner existing work (e.g. FOSTER and ELIXIR-TeSS). The GO TRAIN IN could play an important role to provide certification or endorsement of training materials (e.g., a training seal of approval to indicate quality in training) if it is judged that this will help coordination and avoid duplication of effort and materials.
3. **Training advice, career and capacity building.**
4. **Coordinated and scaling delivery of training.**
5. **Encourage and perform train-the-trainer activities.**

What organisational form would the GO TRAIN IN need to take?
There were varied discussions on this question. One the one hand, it was urged that the functions of the GO TRAIN IN should not duplicate or reproduce the work of existing organisations and initiatives. On the contrary, the benefit and justification of such a IN was precisely to provide a coordinating function. By the same token, the value of an organisation/central office dedicated to providing a horizontal coordinating function to help provide direction and to accelerate the widespread implementation of data training in the five areas identified was recognised. This points to the need for dedicated and funded effort.

Whether the GO TRAIN objectives and functions could best be achieved through a real or 'virtual' organisation, through seconded or dedicated staff requires further discussion. Where possible the IN should build on, use expertise from existing organisations/partners. The IN should provide coordinating functions and should also have a mechanism for distributing resource to partners delivering training or developing materials.

Conclusion: The primary objective of the GO TRAIN IN is to provide a horizontal coordinating function in order to provide direction and to accelerate the widespread implementation of data training in the five areas identified. The optimal institutional arrangement needed to provide such a function requires further discussion. The GO TRAIN IN should focus on the five priority activities identified above.

5) What needs to be funded and sustained?

Based on the discussion above - which needs to be further refined both in terms of the precise functions and the structure of the organisations - the following funding requirements were identified:

1. Initial investment to establish the coordinating function and its core mechanisms;
2. Initial capacity building investment to assist partners delivering training or developing materials, including review of existing materials and creating community maintainable directory of training materials;
3. Create and maintain trainers network;
4. Ongoing funding to sustain the horizontal coordinating function for an appropriate period.

Options for Sustainability

The case is very strong for establishing a coordinating function as described, and for sustaining it while there remains a need to support the wider and accelerated development of data skills and roles. This would require initial investment, as described above, but models for sustaining the IN should be considered: the GO TRAIN IN should not be sustained for its own sake, but only for so long as it provides an important function.

The discussion of sustainability and funding models was limited by time and attenuated by some uncertainty as to the precise characteristics of the GO TRAIN IN, the optimal organisation structure and the opportunities for support. The following models for sustained funding were identified and should inform further discussion.

- Contributions/membership from national partners.
- Contributions/membership from research institutions.
- Secondments and/or contributions from existing institutions.
- Create a commercial market for delivery - free materials, charge for delivery of training events. Delivery could be by the 'IN' organisation or by the consortium of partners. Charging needs to sustain the IN.
- Software and Data Carpentry model: materials are free, delivery is branded, members make contributions to the process which accredits trainers.
- Train-the-trainer consortium model: funding drawn from membership fees, which provide members with discount on train-the-trainer activities.
- Service provided for and recognised by National and European Commission programmes: funded by a top-slice of project funding, perhaps using the mechanism of 'cloud coins' (voucher mechanisms).
- Brokerage for a pool of experts for emergency data skills: one of the benefits of membership, provided only to paying members?

Conclusion: further work is required on the characteristics of the GO TRAIN IN as a real, virtual or consortium/federated organisation. The criteria for these characteristics relate both to the delivery of the IN's coordinating functions and to the possible business model to sustain those functions.

Appendix A: Workshop Agenda

Workshop to Discuss the Formation of GO TRAIN Node, 3 February 2017, Paris

The meeting is to discuss the focus and logistics of a GO FAIR Training Node. The intended output is a short document, for discussion, which will help inform the activities and focus of GO TRAIN.

The discussion session will identify the scope of the training; the modes of delivery; the progression and assessment; and how the logistics (namely instructors, materials, etc.) will be implemented. The workshop is an opportunity for brainstorming and sharing knowledge on these issues and will produce a concise and high-level summary.

Time	Content
9:30-10:00	Talk from Barend summarising GO FAIR, goals, likely budget, management structure (i.e., GO FAIR is more like a loose collection of nodes with some coordinating at the centre as opposed to top down management like ELIXIR). The need for data stewards.

10:00-12:00	**What do we need?**
	Two break out groups (need a chair and rapporteur):
	    1. What are the core skill sets for Data Stewards? The group will produce a structured list of skills. Iryna Kurchma Rapporteur
	    2. What logistics are required to enable the training of Data Stewards? The group will describe what a GO TRAIN node would do: i.e. what mechanisms (certification or endorsement, examination), what initiatives, what sort of coordination activity etc is needed? What are your desiderata for a GO TRAIN node? Laura Molloy Rapporteur
	
	*(allow 15 minute coffee break at **11**)*

12:00-12:30	Report back from break out groups. (10 minutes per group + 10 minutes discussion)
12:30-13:30	*Lunch*
13:30-15:30	**How is it delivered? Who does what?** Another two break out groups (chair and rapporteur required). Each group will consider the same set of issues. What is the progression in delivering the skills (from foundational understanding for researchers to expert level)? **What organisations can contribute towards GO TRAIN? What are the roles that they would play within GO TRAIN? What type of organisations will have to be set up (i.e., physical or virtual institutes, working within third level institutes or as a separate professional body, etc.)? Consider sustainability.** *Rapporteurs Steve Brewer and Celia Van Gelder* *(15 minute coffee break at 14:30)*

15:30-16:00	Report back (10 minutes per group + 10 minutes discussion)
16:00-16:30	Develop structure and notes for a draft report (1-2 page) summarising conclusions from the meeting. This should lay out recommendations, possible structure and actions to achieve a GO TRAIN node.

Appendix B: Participants

1. Kevin Ashley <kevin.ashley@ed.ac.uk>, Digital Curation Centre
2. Teresa Attwood <teresa.k.attwood@manchester.ac.uk>, GOBLET
3. Steve Brewer <S.Brewer@soton.ac.uk>, EDISON
4. Yuri Demchenko <y.demchenko@uva.nl>, EDISON
5. Ingrid Dillo <ingrid.dillo@dans.knaw.nl>, DANS
6. Celia van Gelder <celia.van.gelder@dtls.nl>, ELIXIR, DTL
7. Clare Gryce <c.gryce@ucl.ac.uk>, LEARN, UCL
8. Simon Hodson <simon@codata.org>, CODATA
9. Iryna Kuchma <iryna.kuchma@eifl.net>, EIFL, FOSTER, Open-AIRE
10. Laura Molloy <laura.molloy@rsa.ox.ac.uk>, RDA IG on Data Education; Oxford Internet Institute.
11. Barend Mons <barendmons@gmail.com>, Leiden University Medical Centre; GO FAIR
12. Per Oster <per.oster@csc.fi>, EUDAT; CSC.
13. Susan Reilly <susan.reilly@kb.nl>, LIBER
14. Eloy Rodriguez <eloy@sdum.uminho.pt>, FOSTER
15. Jonathan Rans <J.Rans@ed.ac.uk>, DCC
16. Hugh Shanahan <Hugh.Shanahan@cs.rhul.ac.uk>, CODATA-RDA Schools of Research Data Science; RHUL.

Appendix C: List of Skills for the Five Skills Areas identified as needing most attention from a GO TRAIN Implementation Network

Further work should refine these lists building on EDISON (http://edison-project.eu/edison/edison-data-science-framework-edsf); FOSTER (https://www.fosteropenscience.eu/images/documents/D2.3_ContentandLearningObjectives.pdf); COAR (https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf http://datasupport.researchdata.nl/en/about-the-course/competenties/); ISCB (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003496) and other resources.

A lot of work has already been done in the area of identifying data skills from a variety of perspectives. A useful and tractable task in the development of the GO TRAIN IN will be to build the skill sets for five 'skills areas / roles' on the basis of existing work as outlined by the preliminary lists provided below.

1. Specialist embedded research data skills: There is a significant challenge in how to develop the combination of subject knowledge and data skills.
- Open science skills
- Data modelling/analytics
- Selection of data / choosing what to keep
- Long-term archiving - reg. Knowledge
- Communication skills
- Experiment design / statistics
- Data curation

2. Institutional research data advisor: This cluster of skills is increasingly being embodied in the data librarian role in the libraries of research institutions; it also covers the skills of the Front Office function in the Research Data Netherlands model.
- How to promote data
- ICT skills
- Consulting
- Discovery
- Metadata
- Provenance assessment
- Legal
- Open science skills
- Research method

3. Long-term stewardship: The skills that are required to see to the long term preservations, archiving, accessibility and availability of the data. These skills relate to the functions provided by trusted digital repositories; they correspond to the 'back office' function in the Research Data Netherlands model.
- Understand high-use material

- Superficial domain knowledge
- Assessment of "value of data"
- Knowledge of data infrastructure and services (global and local)
- Selection / appraisal
- Data publishing and data curation
- Data presentation and data visualisation
- Active data reuse
- Open Science and Open Data skills
- Research methods

4. Data skills of benefits to all researchers: all researchers now need a range of skills relating to digital data. In some cases these might be termed 'awareness-level' competencies: knowing when and how to call on additional expertise. The demands of specialisation are such that it is not proposed that all researchers should also be expert data managers, data stewards and data scientists. But what some have called 'research data literacy' is now a requirement. Similarly, the curricula and training efforts have emerged which offer foundational data management and data science skills for all researchers.
- principles and practice of Open Science and Open Data
- Basic research data management and curation (knowledge of curation lifecycle, DMPs)
- use of a range of data platforms and infrastructures
- large scale data analysis tools and technologies
- Statistics and statistical methods
- Data visualisation and modelling techniques
- basic computational skills, software development: familiarity with one or several data analysis languages and platforms (e.g. R, python, Julia, RapidMiner, SPSS, Tableau, etc.),
- data annotation and metadata
- Involvement into related professional network

5. Skills management for organisations: managers within organisations must understand the importance of all the above roles and be willing to advocate their importance internally. In this respect, such managers need to have an outline understanding of all of the above (i.e. what is involved in the specialist roles and the need for data skills for all researchers). These managers needs to have an understanding of the cost implications and benefits of these roles and best financial practices in this area.

Appendix D:

Basic listing of the 'offer' provided by organisations and projects that participated in the workshop. This should be extended to a matrix of organisations and the typology of activities developed above

CODATA-RDA Schools of Research Data Science	● Generic, foundational, training; ● Short format (2 week school), but can be split into sub-units; ● Primary targets are ECRs, but also train-the-trainer component; ● Target group: mainly those who will be embedded in research groups; ● Aims to scale up, particularly building on model of SC/DC.

Digital Curation Centre (DCC)	- RDM, data curation, open research, policy; - Generalist, but also addresses domain areas; - Over 10 years of delivery as an organisation; - Aim is to upskill researchers and institutions as well as performing advocacy function; - Transitioning; - Training is provided as a self-sustaining activity, both on-site at locations around the world in response to customer requests and as open events and in shorter forms such as webinars; - Training provision in institution has included train-the-trainer elements.
DTL	- Provide practical support; - Seek to support new profession of data steward: learning by doing; Have run c. 10 BYOD (Bring Your Own Data) workshops, which 1) bring data specialist together with content people; 2) participants are together for few days, make their sample data sets FAIR, run analytics on day three to show results; - Sustainability for the workshops addressed along from the beginning: paying model for the BYOD, 25 keuro; - Teachers and linked data experts are being paid; - Further developing business model; - Can help build the sustainable, scalable business model for GO TRAIN, e.g. build a franchisable package, put certification in place (e.g. the materials should be Open and free, but the IN can charge for the stamp, certification process).
EDISON	- Project finishes in August 2017; - Amsterdam School of Data Science will guarantee EDISON maintenance for at least 3 years; - EDISON portal will be available for 5 years (no further development but will be maintained); - Offers framework: defines curriculum, competences, Body of Knowledge; - Curriculum profiling; - Capacity building; - Started to work with other domains related to competences and to implement framework.
ELIXIR Training	- Domain-specific, life sciences; - Aim to define that 'border' between generic and domain-specific requirements and build on the generic materials of CODATA-RDA and the framework provided by EDISON; - Working together with EDISON and CODATA-RDA related to data science competences; - Target audience includes end-users=researchers AND bioinformaticians/developers/admin; - Currently defining skills requirements for the different rolesand will then develop training; - Training Portal TeSS: importance of findability, descriptions of materials should also be FAIR, described with metatags, http://bioschemas.org/index.html ; 21 countries in Europe involved (ELIXIR nodes), each with national training coordinator, pool of combined expertise, delivering trainings, best practices, mechanism to disseminate in place;

	- Train the trainer programme and framework with materials; - Putting in place ELIXIR Industry Link for training. - Metrics and KPIs have been defined and are being monitored - Collaboration agreement with GOBLET and with BD2K TCC
FOSTER	- First stage is ending; in May the 2nd stage starts and will run for 2 years; - Approach: collecting material, categorizing, according to subject, target audience, level, language; > 2000 entries; mostly introductory level; 2nd stage will focus on intermediate and advanced materials; - Train the trainer approach; - Network of FOSTER trainers; - Badging mechanisms; - MOOC in preparation.
GOBLET	- Established global training network, members are organisations (US, Australia, Africa, Asia, South America); - See http://www.mygoblet.org ; - Portal in place, repository, not specific for data science but extends to training in bioinformatics; - Collaboration agreement with ELIXIR in place; - Setting up minimal descriptors for training materials, in liaison with http://bioschemas.org/index.html ; - Setting up guidelines for teachers in preparing courses and materials.
LEARN	The primary deliverable of the LEARN project has been the publication of a Toolkit of Best Practice in Research Data Management, which includes: 23 Best-Practice Case Studies from institutions around the world, drawn from issues in the original LERU Roadmap; 8 Main Sections, on topics such as Policy and Leadership, Open Data, Advocacy and Costs; - One Model RDM Policy, accompanied by guidance and an overview of 20 RDM policies across Europe; - An Executive Briefing in six languages, aimed at senior institutional decision makers. Issues of roles, responsibilities and skills (including training) cross cut these outputs.