D2.1–Technical protocol for rich metadata categorization and content classification delivered to WP3

WP2 - Content

University of Glasgow

18/06/2014

DOCUMENT DESCRIPTION

Deliverable number:	D2.1
Deliverable title:	Technical protocol for rich meta-data categorization and content classification delivered to WP3
Actual Date of Delivery:	10/07/2014
Author(s):	Joy Davidson, Iryna Kuchma, Vanessa Proudman, Astrid Orth, Sarah Jones
Workpackage:	WP2 -Content
Distribution:	Public
Version/Revision:	1.0

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612425

TABLE OF CONTENTS

Summary 3

Participants and timeframe 4

Background for metadata scheme 4

Sub task 1: Identifying training resources 5

Sub task 2: Describing training resources 6

Sub task 3: analysis of training resources 6

Links with WP3 10

Links with WP4 10

Next steps 10

Appendix 1: FOSTER classification scheme 11

Appendix 2: FOSTER content identifcation and mapping form 16

Appendix 3: FOSTER criteria for describing, reviewing and evaluating courses and resources – preliminary checklist 20

Appendix 4: DaMSSI-ABC Classficaition Scheme 23

Appendix 5: FOSTER Content Map 24

GRAPHICS

Graphic 1 - Resources by Discipline 6

Graphic 2 - Resources by Audience 7

Graphic 3 - Resources by Topic 7

Graphic 4 - Resources by Level of Knowledge 8

Graphic 5 - Resources by Language 9

Graphic 6 - Resources by Format 9

FIGURES

Figure 1 - FOSTER Content Identification and Mapping Form 5

SUMMARY

FOSTER aims to set in place sustainable mechanisms for EU researchers to FOSTER OPEN SCIENCE in their daily workflow, supporting researchers optimizing their research visibility and impact and the adoption of EU open access policies in line with the EU objectives on Responsible Research & Innovation.

More specifically, the FOSTER objectives are to:

- Support different stakeholders, especially young researchers, in adopting open access in the context of the European Research Area (ERA) and in complying with the open access policies and rules of participation set out for Horizon 2020;
- Integrate open access principles and practice in the current research workflow by targeting the young researcher training environment;
- Strengthen the institutional training capacity to foster compliance with the open access policies of the ERA and Horizon 2020 (beyond the FOSTER project);
- Facilitate the adoption, reinforcement and implementation of open access policies from other European funders, in line with the EC's recommendation, in partnership with PASTEUR4OA project.

As stated in the project Description of Work (DoW) these objectives will be pursued and achieved through the combination of 3 main activities: **content** identification, repacking and creation; creation of the **FOSTER Portal**; delivery of **training**.

The core activity of the Task T2.1 will be to define a basic quality control protocol for content, and map available content by target group, and content type in parallel with WP3 Task 3.1.

Training materials include the full range of classical (structured presentation slides) and multi-media content (short videos, interactive e-books,) that clearly and succinctly frames a problem and offers a working solution, in support of the learning objectives of each target group, and the range of learning options to be used in WP4 (elearning, blended learning, self-learning).

The map of existing content metadata will be delivered to WP3 for best choice of system requirements for continuous and sustainable content aggregation, enhancement and content delivery via "Tasks 3.2 e-Learning Portal" and "Task 3.4 Content Upload". The resulting content compilation will be tailored to each Target Group and delivered to WP4.

PARTICIPANTS AND TIMEFRAME

Participants in the task included:

Task leader - DCC-HATII
Task Contributors – University of Minho, UGOE, DTU, SPARC-E, TUDelft, OU, CSIC, DCC-UEDIN

Work began in M2 and finished in M5. Two PM were allocated for this task across the nine partners. At the end of M5, we have some unused effort remaining. We propose to allocate this towards updating the content map over the life of the project as new resources are released and to attempt to fill any gaps that were identified in topics covered and content for specific groups.

BACKGROUND FOR METADATA SCHEME

DCC-HATII led this task and, as described in the DoW, suggested that we build upon our previous efforts to classify open science and in particular research data management (RDM) training outputs in the Jisc-funded Data Management Skills Support Initiative – Assessment, Benchmarking, Classification (DaMSSI-ABC) project. DaMSSI-ABC aimed to classify course offerings produced by the Jisc 01/12 RDMTrain projects[1] to ensure that the anticipated outcomes of these and other training interventions could be clearly set out to allow participants to select the training that best meets their particular learning objectives. The DaMSSI-ABC team developed the classification scheme based on reviews of the Research Information and Data Literacies Coalition (RIDLs) criteria for describing and evaluating which had been considered alongside Vitae's Researcher Developer Framework (RDF) and the European Commission-funded DigCurV curricula lenses. The following is the list of the metadata fields developed for use within the project and subsequently added to the controlled vocabularies chosen for the Jorum RDM Portal. A full description of the fields is included in Appendix 4.

- Intended audience
- Ability level
- Delivery
- Range
- Skills

WP2 participants reviewed the DaMSSI-ABC classification scheme in March 2014 and revised it to reflect Open science more generally and to highlight our four target groups. The full classification scheme is included in Appendix 1.

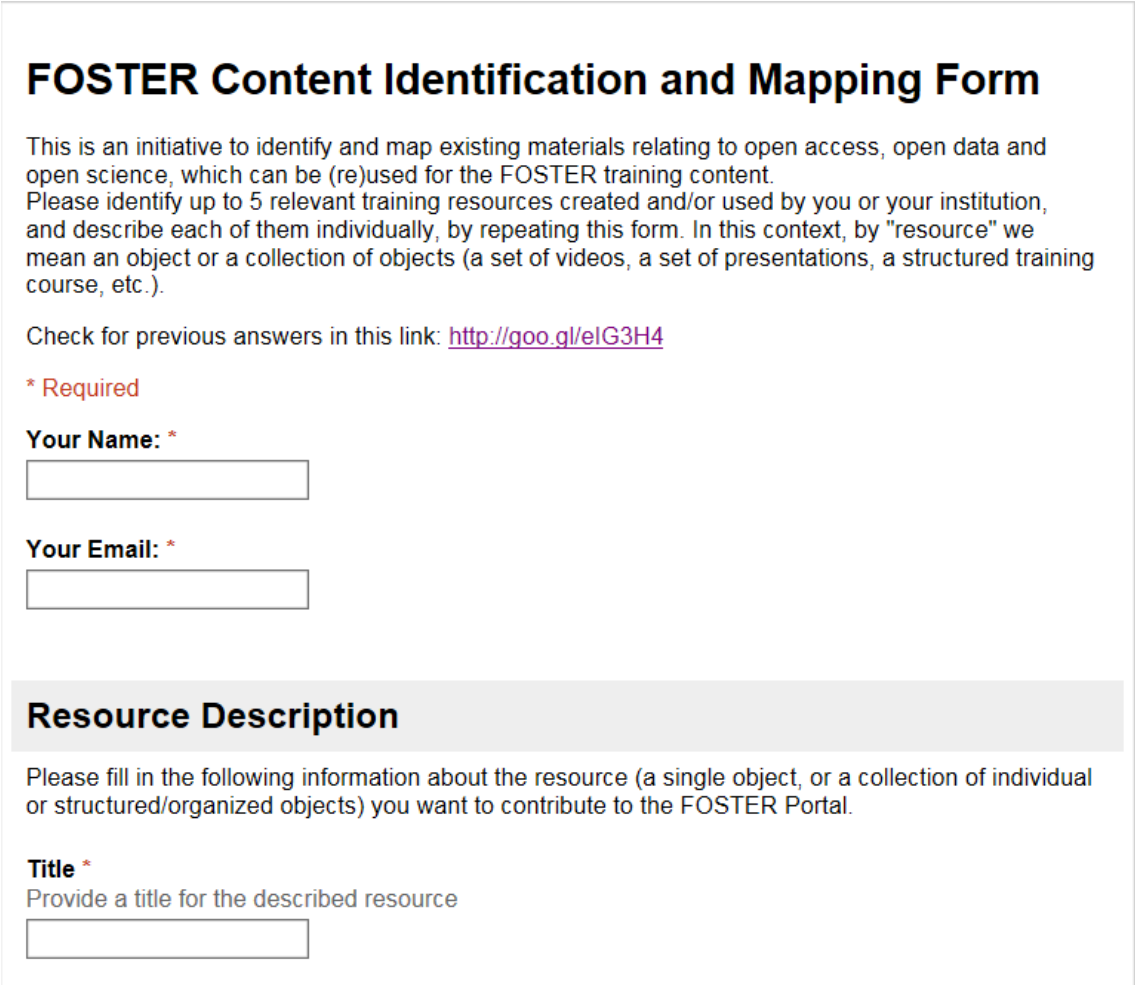
- Title
- General description of the resource
- Author and creator
- Date
- URL of the resource
- Language
- Target audience
- Scientific discipline
- Level of Knowledge
- Main topic
- Secondary topic
- Resource licence
- If there is any specific license associated, please specify

1

- Media type
- File type
- Size

In early April, UMINHO developed a draft content collection form based on the agreed classification scheme – which was also presented by Sarah Jones at the kick-off WP2 workshop in Minho in February.

Figure 1 - FOSTER Content Identification and Mapping Form



FOSTER Content Identification and Mapping Form

This is an initiative to identify and map existing materials relating to open access, open data and open science, which can be (re)used for the FOSTER training content.
Please identify up to 5 relevant training resources created and/or used by you or your institution, and describe each of them individually, by repeating this form. In this context, by "resource" we mean an object or a collection of objects (a set of videos, a set of presentations, a structured training course, etc.).

Check for previous answers in this link: http://goo.gl/elG3H4

* Required

Your Name: *

Your Email: *

Resource Description

Please fill in the following information about the resource (a single object, or a collection of individual or structured/organized objects) you want to contribute to the FOSTER Portal.

Title *
Provide a title for the described resource

The form is available to add additional content at https://docs.google.com/forms/d/18SqFx57oitBOcFn7u7qcbR56rb7j0Qfg5aUMFocoaMo/viewform

SUB TASK 1: IDENTIFYING TRAINING RESOURCES

Each partner involved in Task 2.1 was asked to identify at least **20 training resources** as part of our initial sweep of the training landscape. Choices were to be made based on the expertise of the partners involved and their particular areas of interest (e.g., Open Sciences, Research Data Management, Policy). The quality controls developed for the identification of training content were:

- Content is available online
- Content is openly available (e.g. CC licence)
- Content related to an aspect of open science

SUB TASK 2: DESCRIBING TRAINING RESOURCES

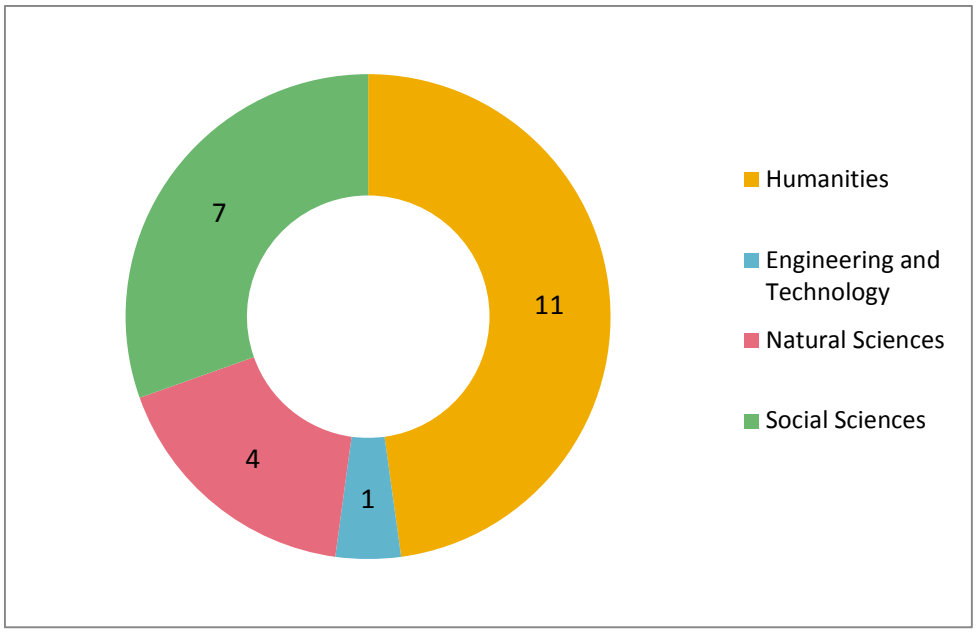
Once the initial mass of resources were entered into the content form at a high level, participants began the second phase of the work which was adding value by improved description of the resources using the agreed metadata scheme. A final check of the content was carried out by DCC-HATII to refine links to multiple resources and to ensure that each resource identified met the basic quality control set in place.

SUB TASK 3: ANALYSIS OF TRAINING RESOURCES

DISCIPLINE

A total of 146 training resources have been collected during task 2.1. The majority of the resources (119) are multidisciplinary in nature. This means that we have a broad base of resources that can be adapted for use within a number of disciplines. However, as we progress work in Tasks 2.2, 2.3, 2.4 and 2.5 we may seek to identify – or source from the second call for FOSTER sponsored training in WP4 - more discipline-specific resources.

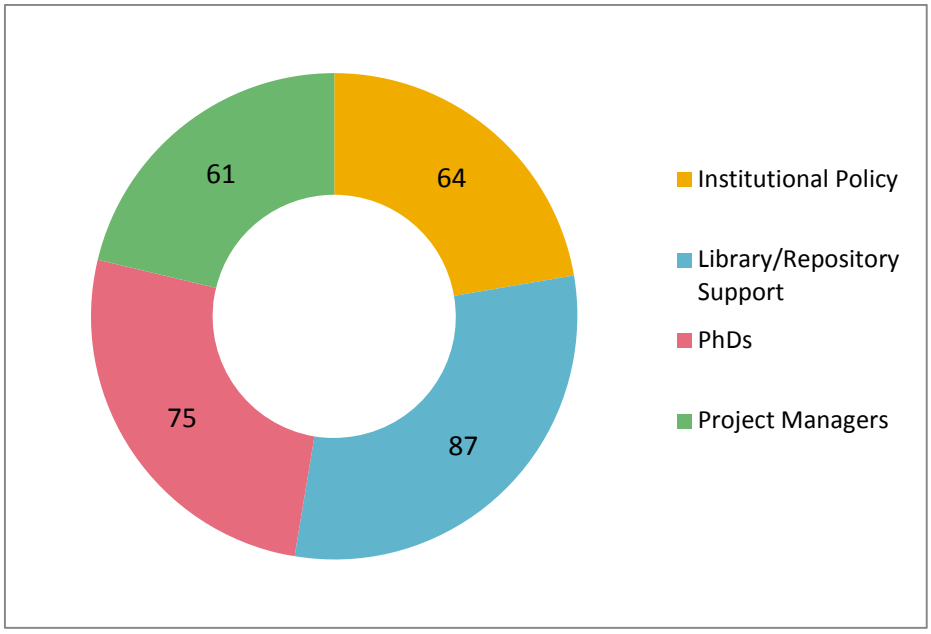
Graphic 1 - Resources by Discipline



TARGET AUDIENCE

There was a very good split of resources identified across our four key target groups. We identified a slightly higher number of resources targeted to Libraries/Repository Support staff which is beneficial as in most cases it is this group that will be responsible for providing training to staff and students within European HEIs. Interestingly, most of the resources identified were applicable to more than one target audience which means that with slight revisions, the materials should be widely applicable for reuse.

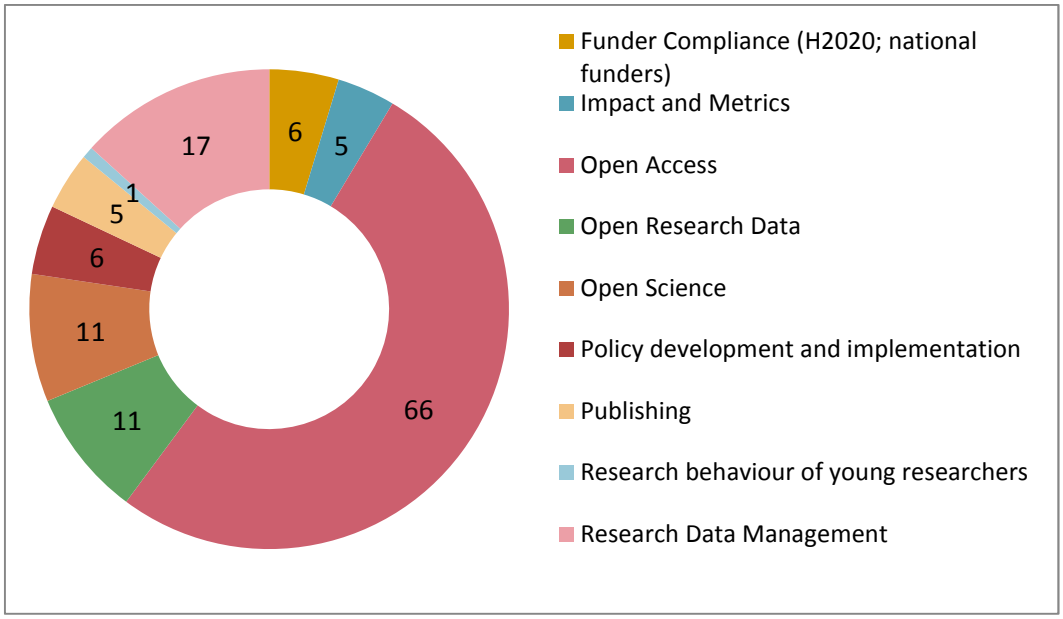
Graphic 2 - Resources by Audience



TOPICS

The range of topics covered by the resources we identified was quite broad. However there was a clear majority of resources relating to Open Access in general. Again, this is probably quite good as we will be spending quite a bit of time increasing advocacy and general awareness of Open Access with our target groups in WP2. There were quite a few resources identified that relate to research data management, open data and open science. These will be beneficial as we try to encourage researchers to consider sharing their research data as well as their publications.

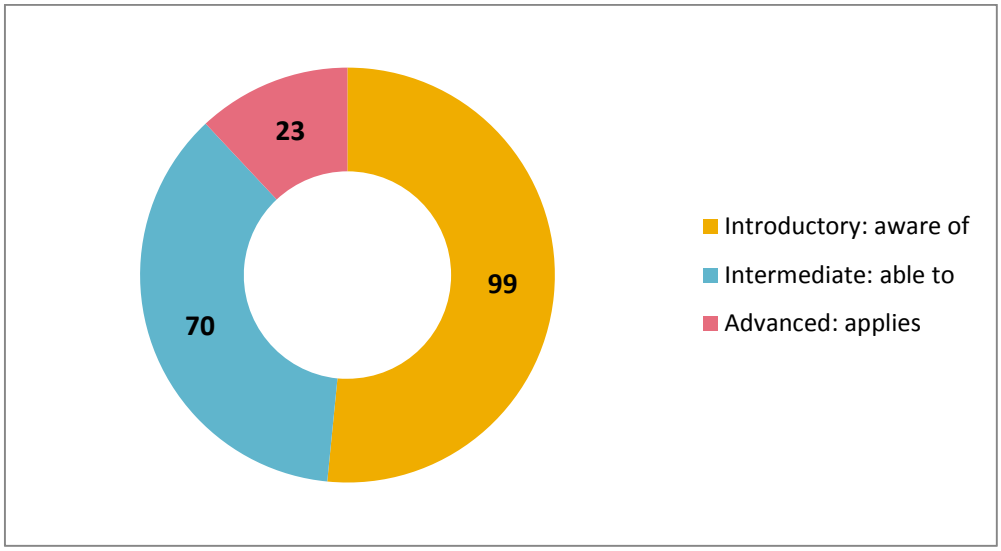
Graphic 3 - Resources by Topic



LEVEL OF KNOWLEDGE

The metadata scheme attempts to benchmark the training materials to ensure that they are targeted to specific levels of knowledge. Introductory level materials provide an awareness of key issues. Intermediate level materials ensure that participants are able to do something particular after completing the course. Advanced level materials ensure that those completing the course can apply their knowledge in their own particular setting. The majority of the training resources identified in task 2.1 fell with within the introductory or intermediate categories. This is about right as much of the work we need to do is awareness raising at this stage. We may seek to increase the level of advanced materials within the FOSTER portal in years to come as the cohort of EU researchers, project managers, and library/repository support staff become more skilled in dealing with Open Science issues.

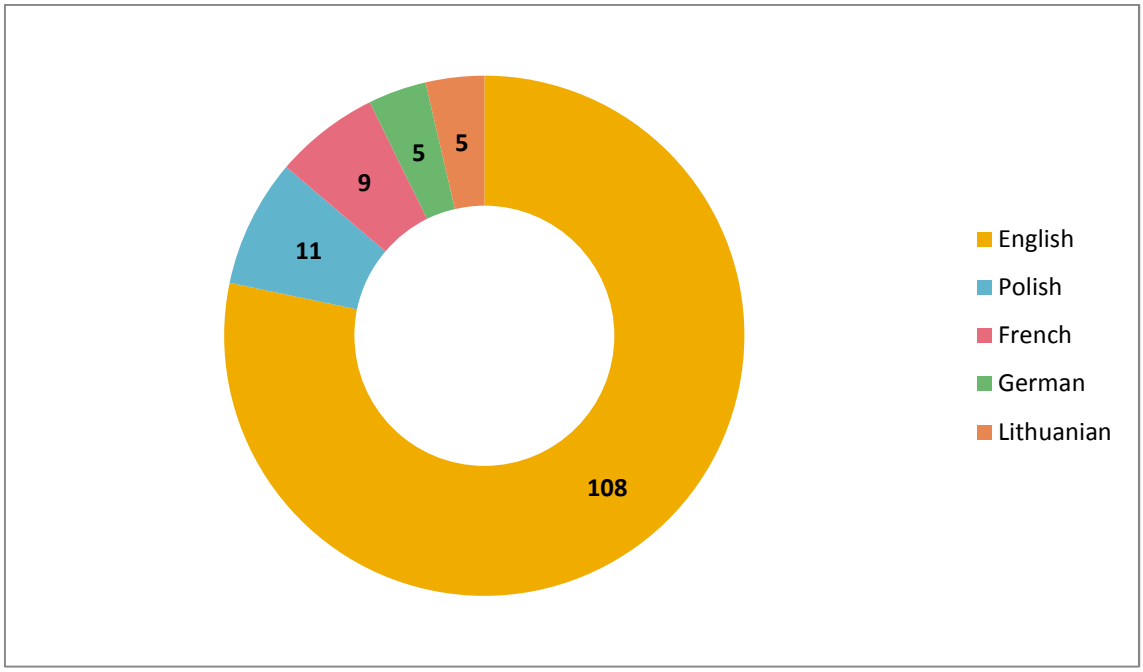
Graphic 4 - Resources by Level of Knowledge



LANGUAGE

The vast majority of the resources are in English which is what we anticipated. However, we do have quite a bit of content in Polish, Lithuanian and French. We may seek to identify additional content for other language groups as we proceed with FOSTER.

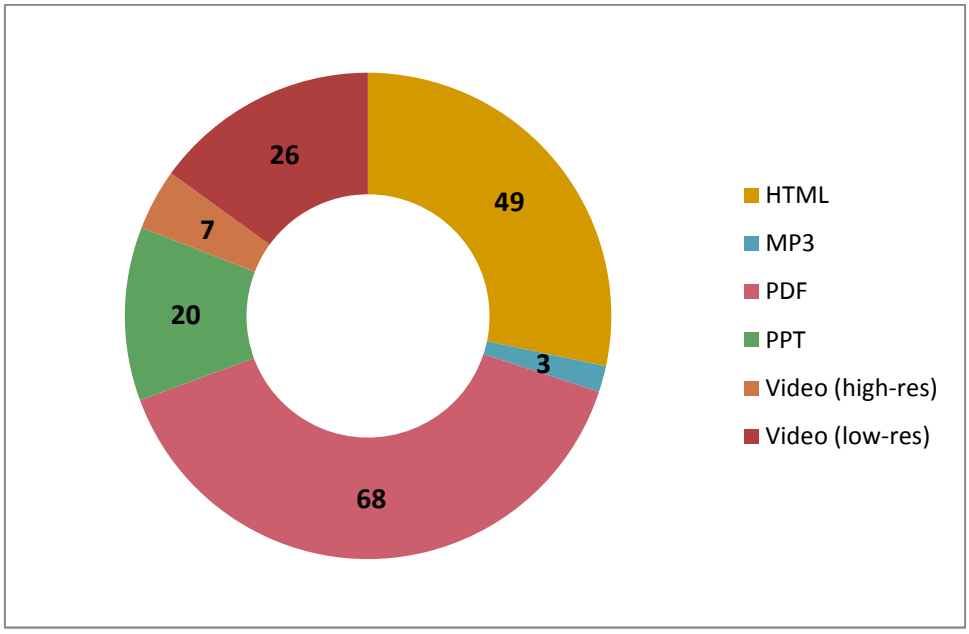
Graphic 5 - Resources by Language



FORMAT

Most of the resources identified in task 2.1 were PDF files, HTML pages and PowerPoint slides. However, in many cases the identified resources were comprised of more than one format (e.g., HTML page with a link to a PPT file and a low-resolution video). As we proceed to populate the FOSTER portal over the life of the project, we should seek to source additional video content.

Graphic 6 - Resources by Format



LINKS WITH WP3

From the start of the project participants in WP2 and WP3 have been keen to share progress to ensure that the FOSTER portal development can progress smoothly. Participants from both work packages joined in early telecons and virtual meetings to ensure that these two tasks progressed in parallel and informed each other. It has been agreed that the online FOSTER portal will not include a mandatory option for discipline for depositors at this stage due to the overwhelming majority of resources that are multidisciplinary in nature. However, we recommend that this be maintained as an optional field within the system. During early discussions about the content form and metadata fields in Task 2.1 we decided to allow content depositors to select more than one target audience for each resource identified for the content map rather than assigning both primary and secondary target audiences for each resource. The vast majority of the resources identified were subsequently allocated to between 4-5 target groups. This made the analysis rather difficult and means that FOSTER portal users searching for targeted materials may not get back anything particularly useful to their needs but more generic in nature. Based on this experience, we recommend including a mandatory primary target audience field allowing only one choice and then allowing depositors to complete an optional secondary audience field where they ,may choose more than one option.

LINKS WITH WP4

The first tranche of FOSTER sponsored training sessions have been agreed. The outputs will be uploaded into the FOSTER portal using the same metadata scheme used in task 2.1. The review of the content map will help to inform priority areas for the second call for FOSTER sponsored training session that will come out in October 2014. We have a fairly good split of training materials across each stakeholder group. We have a large number of resources aimed at the introductory level so, as we proceed with the second call for FOSTER sponsored training, we may wish to emphasise materials targeted towards intermediate and advanced levels of study. All FOSTER-sponsored trainers in the first round have been asked to complete a recording agreement so that we have an audio resource to share via the FOSTER portal. However, we may wish to investigate the potential for acquiring more video content in the second round – perhaps through animations or short lectures rather than recorded workshops.

NEXT STEPS

The initial content map was handed over to WP3 in mid May along with the metadata schema used to describe training content. From month 6-9, participants in task 2.1 with unused effort will carry out a secondary sweep of the landscape to consider how the second call for FOSTER sponsored training might complement the gaps that have emerged from the review carried out in task 2.1. Tasks 2.1 participants recommended that WP4 might consider applying similar basic quality control checks as were carried out in WP2. In addition, task 2.1 recommended the use of the RIDLs criteria for describing and evaluation good practice which has been amended for use in WP4.

APPENDIX 1: FOSTER CLASSIFICATION SCHEME

This classification scheme will enable users of the FOSTER portal to better identify relevant training for specific topics, stakeholders and skills levels.

FOSTER classification scheme					
Field	**Field description**	**Mandatory field?**	**Format**	**Drop-down / tick-box list options**	**Specification**
Title	Provide a title for the described resource	Yes	Free text	N/A	
General description of the resource	Provide a brief description of the resource	Yes	Free text	N/A	2000 characters
Author and creator		Yes	Free text	N/A	
Date	Provide the date of creation or last update of the resource	Yes	Drop-down list	2014	Select only one
				2013	
				2012	
				2011	
				2010	
				2009	
				before 2009	
URL of the resource	If the resource is available online, please provide a URL	No	Free text	N/A	
Language	Select the main language for the resource	Yes	Drop-down list	Bulgarian	Select only one
				Czech	
				Danish	
				Dutch	
				English	
				Estonian	
				Finnish	
				French	
				German	

				Greek	
				Hungarian	
				Irish	
				Italian	
				Latvian	
				Lithuanian	
				Maltese	
				Polish	
				Portugese	
				Romanian	
				Slovak	
				Slovene	
				Spanish	
				Swedish	
				Other	
Target audience	For which stakeholders is your resource intended? You can select more than one item.	Yes	Tick box option	Researcher/Students	Can select more than one option
				Library/Repository support	
				Project Manager	
				Institutional Policy Maker	
				Funding body	
				Publisher	
				PhD student	
Scientific discipline	Select a specific scientific domain or select multidisciplinary.	Yes	Drop-down list	Natural sciences	Select only one
				Engineering and technology	
				Medical and health sciences	

				Agricultural sciences	
				Social sciences	
				Humanities	
				Multidisciplinary	
Level of Knowledge	Please select the scope level of your resource	Yes	Tick box option	Introductory: aware of	Select only one
				Intermediate: able to	
				Advanced: apply	
Main topic	Select the main coverage of your resource based on the following topics.	Yes	Tick box option	Open access	Select only one
				Open research data	
				Copyright & IPR	
				Funder compliance (H2020, national funders)	
				Impact and metrics	
				Open science	
				Policy development and implementation	
				Research data management	
			Free text	Other	
Secondary topic	Select the coverage of your resource based on the following topics.	No	Drop-down list	Open access	Select only one
				Open research data	
				Copyright & IPR	
				Funder compliance (H2020, national funders)	
				Impact and metrics	
				Open science	
				Policy development and implementation	

				Research data management	
Resource licence		Yes	Tick box option	Yes	Select only one
				No	
				Don't know	
If there is any specific license associated, please specify		No	Drop-down list	None (all rights reserved)	
				Attribution -Non Commercial -Share Alike - CC	Select only one
				Attribution -Non Commercial - CC	
				Attribution - Non Commercial -CC	
				Attribution - Non Commercial - No Derivs - CC	
				Attribution - Creative Commons	
				Attribution - Share Alike - CC	
				Attribution - No Derivs - CC	
				CC0	
			Free text	Other	
	If there isn't any specific license associated would it be possible to associate an open license (like CC-By or similar)?	No	Tick box option	Yes	Select only one
				No	
				Don't know	
Media type	In what types of media was your resource created?	Yes	Tick box option	Text	Can select more than one option
				Videos	

				e-Book	
				Multimedia Objects	
				Interactive Objects (Quiz)	
				Audio	
				Image	
			Free text	Other	
File type	Select the physical type of the resource.	Yes	Tick box option	PDF	Can select more than one option
				PPT	
				DOC	
				MP3	
				HTML	
				Video (low resolution - less than 720 pixels height)	
				Video (high resolution - more than 720 pixels height)	
			Free text	Other	
Size	Can you estimate the volume of your resource?	Yes	Drop-down list	Up to 10 megabytes (MB)	
				Between 10 and 100 MB	Select only one
				Between 10 and 100 MB	
				Between 100 and 500 MB	
				Between 500 and 1 Gigabyte GB	
				Between 1 GB and 5 GB	
				More than 5 GB	
				Don't know	

APPENDIX 2: FOSTER CONTENT IDENTIFCATION AND MAPPING FORM

This is an initiative to identify and map existing materials relating to open access, open data and open science, which can be (re)used for the FOSTER training content. Please identify up to 5 relevant training resources created and/or used by you or your institution, and describe each of them individually, by repeating this form. In this context, by "resource" we mean an object or a collection of objects (a set of videos, a set of presentations, a structured training course, etc.).

Check for previous answers in this link: http://goo.gl/eIG3H4

---

* Required

Your Name: *

Your Email: *

RESOURCE DESCRIPTION

Please fill in the following information about the resource (a single object, or a collection of individual or structured/organized objects) you want to contribute to the FOSTER Portal.

Title *

Provide a title for the described resource

General description of the resource *

Provide a brief description of the resource (max. 2000 characters)

Author & Creator *

Date *

Provide the date of creation or last update of the resource.

URL of the resource

If the resource is available online, please provide a URL.

Language *

Select the main language of your resource.

Target Audience *

For which stakeholders is your resource intended? You can select more than one item.

- o ☐ Researcher / Students
- o ☐ Library / Repository Support
- o ☐ Project Manager
- o ☐ Institutional Policy
- o ☐ Funding
- o ☐ Publishing
- o ☐ PHD Student

Scientific Discipline *

Select a specific scientific domain or select multidisciplinary.

Level of Knowledge *

Please select the scope level of your resource

- o ☐ Introductory: aware of
- o ☐ Intermediate: able to
- o ☐ Advanced: apply

Main Topic *

Select the main coverage of your resource based on the following topics.

- o ☐ Open Access
- o ☐ Open Research Data
- o ☐ Copyright & IPR
- o ☐ Funder Compliance (H2020; national funders)
- o ☐ Impact and Metrics
- o ☐ Open Science
- o ☐ Policy development and implementation
- o ☐ Research Data Management
- o ☐ Other: [ ]

Secondary Topic

Select the coverage of your resource based on the following topics

[ ]

Resource license? *

- o ☐ Yes
- o ☐ No
- o ☐ Dont' know

If there is any specific license associated, please specify:

[ ]

If there isn't any specific license associated would it be possible to associate an open license (like CC-By or similar)?

- o ☐ Yes
- o ☐ No
- o ☐ Don't know

Media Type *

In what types of media was your resource created?

- o ☐ Text
- o ☐ Videos
- o ☐ e-books
- o ☐ Multimedia Objects

- ☐ Interactive Objects (Quizz)
- ☐ Audio
- ☐ Image
- ☐ Other: _____

File Type *

Select the physical type of the resource

- ☐ PDF
- ☐ PPT
- ☐ DOC
- ☐ MP3
- ☐ HTML
- ☐ Video (Low Resolution - less than 720 pixels height)
- ☐ Video (High Resolution - 720 pixels height or more)
- ☐ Other: _____

Size *

Can you estimate the volume of your resource?

[,,"-77233915830	0	false	-7723391583071

Submit

APPENDIX 3: FOSTER CRITERIA FOR DESCRIBING, REVIEWING AND EVALUATING COURSES AND RESOURCES – PRELIMINARY CHECKLIST

The FOSTER criteria enable practitioners to design, review and assess courses and resources.

Part A: criteria for describing and reviewing your course or resource

1.1 Learners: who is it aimed at? Name of course/resource:		
• Researchers • Postgraduate students • Institutional decision makers and staff • Librarians and repository managers • Funders (decision makers and staff) • Project managers	☐ ☐ ☐ ☐ ☐ ☐	
1.1.1. Discipline or professional body (if applicable)		

1.2 Need	
1.2.1. Have you properly assessed learners' needs?	Y/N
1.2.2 Does the course fit the broader professional needs of the learners?	Y/N/NA
1.2.3. Does the course fit with your institutional and/or departmental policy and practice on researcher development?	Y/N
1.2.4. Does it relate also to expectations or requirements from the ERA and Horizon 2020?	Y/N
1.2.5. Can the course be transferred or adapted to suit other needs or contexts from the one for which it was designed?	Y/N
1.2.6. Is the course or resource accessible, particularly for learners with special needs or disabilities such as visual impairment?	Y/N

1.3 Demand	
1.3.1. Is the course a response to demand from learners?	Y/N
1.3.2. Is the course a response to an institutional request/ policy?	Y/N
1.3.3. Is the course a result of your own perception of a need?	Y/N

1.4 Meeting requirements	
1.4.1. Does the course or resource take into account	
• Current level of skill	Y/N
• Years of experience	Y/N
• Disciplinary areas	Y/N
1.4.2. Are there any prerequisites?	Y/N

2.1 Areas of open science, open access and open data covered		
• Open access • Open research data • Open sciences • Policy development and implementation • Research data management • Copyright and IPR	☐ ☐ ☐ ☐	

• Impact and metrics • Funder compliance	☐ ☐ ☐ ☐	
2.2 Learning outcomes		
2.2.1. Do you have explicit learning outcomes? If yes, please list them here:		Y/N
2.3 Use of FOSTER portal/training materials		
2.3.1. In the course have you used FOSTER portal and training materials? 2.3.3. If yes, have you adapted FOSTER training materials to suit your training needs?		Y/N Y/N
3.1 Approach		
3.1.1. Format of the course: • Classroom-based courses (lecture or workshop) • Individual tuition • Online courses • Training material (printed or digital) • Other	☐ ☐ ☐ ☐ ☐	
3.2 Training providers		
3.2.1. Who designs and delivers the course? .3Library .3Graduate school .3IS department .3Other (please specify)	☐ ☐ ☐ ☐	
3.2.2. Do the trainers all have the relevant level of skills and qualification?		Y/N
3.3 Support, dissemination and promotion		
3.3.1. Is there sufficient support required to run the course in terms of:		
• Personnel		Y/N
• Facilities		Y/N
• Finance		Y/N
3.3.2. Is the course well-advertised?		Y/N
3.3.3. Is the course/resource available outside the institution?		Y/N
3.3.4. Has it been deposited in online facilities such as the FOSTER portal?		Y/N

Part B: criteria for evaluating your course or resource

4.1 Numbers	
How many learners have taken part in course or used resource?	
4.2 Feedback	
4.2.1. Did >75% of participants give positive feedback on	
Overall training course satisfaction	Y/N
Relevance of the training materials	Y/N
4.2.2. Has feedback been used to redesign elements of the course?	Y/N
4.2.3. Have you received higher level feedback from departments/ institution? (if applicable)	Y/N
4.2.4. Has this feedback been used to redesign elements of the course? (if applicable)	Y/N
5.3 Knowledge and skills	
5.3.1. Have you been able to assess any changes in learners' knowledge, skills and competencies resulting from course or resource	Y/N

APPENDIX 4: DAMSSI-ABC CLASSFICAITION SCHEME

Intended audience

- Doctoral researcher
- Librarian/information professional
- Postgraduate researcher
- Professional/senior level
- Research staff
- Support staff
- Technical staff
- Tenured researcher (lecturer, reader)
- Trainers/facilitators
- Undergraduate
- Other

Ability level

- Introductory (aware of)
- Intermediary (understands)
- Advanced (able to)

Delivery

- Classroom-based courses
- Individual tuition
- Online courses
- Training materials

Range

- Department
- Discipline
- Faculty
- Institution
- Inter-institutional
- International (open to non UK HEIs)
- National/regional
- N/A

Skills

- Project planning
- Information searching, review and discovery
- Assessment and analysis of information sources
- Creation and collection
- Description
- Attribution and citation
- Ethics & Data Protection
- Open data and Freedom of Information
- Legal issues (licensing, copyright, fraud)
- Selection and retention
- Storage and preservation
- Publishing and dissemination of research results
- Other

APPENDIX 5: FOSTER CONTENT MAP

The content map is available on the FOSTER dropbox deliverable folder at https://www.dropbox.com/s/z7qshhrg6ud81yl/FOSTER%20Content%20Map_For%20Report%20Appendix_30062014.xlsx?n=71153400.