Data Service Infrastructure for the Social Sciences and Humanities

EC FP7
Grant Agreement Number: 283646

Deliverable Report

Deliverable: D7.1
Deliverable Name: Course modules
Deadline: M36
Nature: O

Responsible: UGOE
Work Package Leader: UGOE

Contributing Partners and Editors: UiB, NSD, KCL, OEAW, Timo Gnadt, Claudia Engelhardt

SEVENTH FRAMEWORK PROGRAMME

Contents

1. Introduction

The objectives of the work package "Training and Education" (WP7) were to establish a joint domain for training and education for the SSH infrastructures; to inspire researchers and developers to come up with new research methodologies and approaches using research infrastructures and to discuss with and gather feedback from researchers in all SSH domains about the role of data infrastructures in research methodologies.

To pursue these objectives, the DASISH Description of Work (DoW) specified two tasks. Task 7.1 "Training Modules" was dedicated to developing online training modules for topics and target groups relevant to the SSH communities. The findings here also served as a basis for planning and organizing workshops for these topics in the context of Task 7.2 "Workshop Programme". This report documents the work carried out with regard to Task 7.1 and presents its results. The specific activities and results of Task 7.2 are described in a separate report (D7.2 A compendium of workshop reports). Comprehensive aspects relevant to both Task 7.1 and Task 7.2 are described in the document at hand.

In Task 7.1, several training modules were developed and made accessible to the public via the DASISH website[1]. In section 2, this document first describes the processes of

- assessing the training needs of and available material from ESFRI communities
- organizing the work for creating and publishing online training modules
- gathering feedback from the SSH communities and the other DASISH WPs
- revising the modules based on internal and external feedback
- ensuring sustainability of the material produced

After this, in section 0 the main structure of the actual training modules are described with a brief summary of their chapters' content. The actual training modules can be accessed through the training section on the DASISH website[2].

The Appendix then provides additional details and documents, and also includes a list of acronyms used in this report.

2. Methodology and work progress

a. Assessment of training needs

The target groups of the WP7 training efforts were specified in the DASISH Description of Work as "infrastructure-related initiatives within the participating SSH domains" and

[1] See http://training.dasish.eu/
[2] See http://training.dasish.eu/

"new communities in the SSH". This includes developers and managers of data archives and repositories, decision makers from research and education institutions as well as researchers.
To verify this specification, and to assess the concrete training needs of the different target groups, in its initial stage, WP7 conducted a 2-round survey among the SSH ESFRI projects. The first questionnaire included questions regarding:

- the topics related to data management for which the respondents expressed a need for training (namely the "3-4 most burning issues")
- the ideal format of training modules and workshops
- whether or not a multilingual term base for data management was needed
- data management topics which were already covered by training offers in the respective ESFRI projects
- research data that respondents could offer for use in DASISH training materials

The full first questionnaire can be found in Appendix A1.

This first round of the survey was intended to provide a first overview of topics with a need for training, as regarded by the SSH ESFRI projects. In May 2012 an invitation was sent out to representatives from different institutions belonging to one of the five SSH ESFRI projects, with a request to fill in the questionnaire from the ESFRI projects' perspective. It yielded a total of 7 responses, summarized in the following table.

ESFRI project	CESSDA	CLARIN	DARIAH	ESS	SHARE
responding institutions	GESIS, NSD	UiB, OEAW	KCL, UGOE	NSD	MPG-MEA

The answers from this first questionnaire already provided a good basis for the following, more specific survey. It became evident, however, that there was not much training material available yet among the ESFRI projects or their participating institutions, which DASISH could directly use or integrate into an existing platform such as ESS EduNet.

From the first questionnaire, 5 main topic areas were derived, each with 3-6 subtopics: "Data enrichment", "Data quality", "Data archiving", "Security & Policies", and "Management". Based on this categorization, the ESFRIs were asked to indicate the relevance for certain target groups, kinds of desired activity, available training material and give further comments. This second questionnaire was filled in by WP7 ESFRI contacts together with their respective ESFRI contact person in June/July 2012. It can be found in Appendix A2.

From the results, several highly relevant topics were identified and adopted as a first tentative list of topics to be addressed by training measures: "Access Policies", "Licensing", "Persistent Identifiers", "Data analysis/harmonization", "Workflows", "Linked Data", "Authentication and Authorization Infrastructure", "Metadata standards and usage", "Publication/Open Access" and "Deposit services and SLA negotiations". At the same time the responses again revealed a considerable lack of available training material.

To close this gap, WP7 members searched for available training modules for the identified relevant topics. As a result, several tutorial texts were collected, however none of them in the form of an online training module, which would allow for cross-referencing, context-sensitive help, target group or expert-level diversification or even interactivity.

Since the results of both the WP7 search for modules and the 2-round survey showed that there were no ready-to-publish online training modules available, WP7 decided to create the modules themselves, using the material gathered so far. The work on the first training module on "Access Policies and Licensing" was organized collaboratively, by dividing the module into 7 chapters covering different aspects of the topic, and assigning author, contributor and reviewer to each chapter.

A similar procedure was applied to involve the other DASISH work packages in determining desired topics and forms of training material. While in the first half of the project, there was not much input available from the other WPs, the cooperation intensified strongly during the second half. This was mainly triggered by two topics addressed within WP5 "Shared Data Access and Enrichment": Task T5.2 "Persistent Identifiers" and Task T5.1 "Authentication and Authorization Infrastructures". These topics also had emerged as desired topics from the ESFRIs surveys, and thus fuelled the collaborative creation of two other training modules for these topics.

b. Module compilation and publication

The creation of the training modules was a joint effort of WP7 members. After agreeing on a specific topic, e.g. "Access Policies and Licensing", the group discussed and defined subtopics and viewpoints to be used as separate chapters, in order to structure the module and enable users to directly access specific subtopics (For the detailed structure of each module, see section 3 of this report). According to this chapter structure, each chapter was assigned a responsible editor, a contributor and a dedicated reviewer. These roles were all filled by WP7 members, and enabled a more effective discussion of the module content in smaller groups.

The editing process was done via the use of MS Office documents, which were created as templates, and then filled with content for each chapter by the responsible partners. The template for training module chapters for the first module can be found in Appendix B. The exchange of the documents for the purpose of gathering and discussing feedback took place via the DASISH Portal[3]. When the WP7 partners agreed that the basic structure and scope of the chapters would not significantly change anymore, the content was transferred into the website content management system OpenCMS[4] for final editing and publication.

OpenCMS is a free content management system for creating and managing websites. It can be used in open as well as in proprietary environments. Its browser based user interface can be customized and has a WYSIWYG editor that allows easy input of content

[3] See https://theuniversityofgothenburg.basecamphq.com/
[4] See http://www.opencms.org/en/index.html

also for people who do not possess programming skills. In addition, there is also the possibility to generate customised templates.

OpenCMS was chosen for the publication of the DASISH training modules, since it is also being used for hosting ESS EduNet[5], as well as for setting up and administrating the DASISH website[6]. This made the integration of the training and education pages into the general DASISH website very easy.

NSD, one of the partners in DASISH WP7, provided expertise and support in setting up and managing the webpages and the module structure.WP7 was also able to benefit from NSD having experience with developing online training material, for example from ESS EduNet. In the process of developing the first training module, all WP7 partners were equipped by NSD with login accounts to a part of the OpenCMS system dedicated to the DASISH training modules. All editing partners were thus able to directly edit the content and structure, and notify the partner at NSD that a page or chapter had been revised, thus initiating the publication of the revised content.

The overall coherence of the module, including overlapping content, cross-references and consistent terminology and formatting, was discussed in regular virtual meetings (2-weekly via Skype and Adobe Connect) and several face-to-face meetings. For each module, an introductory page was created which briefly describes the module content and target groups, and provides hyperlinks to the chapters. Furthermore, a Glossary (see Figure 1) was set up that explains terms from all modules and is directly accessible when clicking on a marked-up term in the text. When hovering over a glossary term in the text, a box with the explanatory text from the glossary appears (see Figure 2).

[5] See http://essedunet.nsd.uib.no/
[6] See http://dasish.eu/

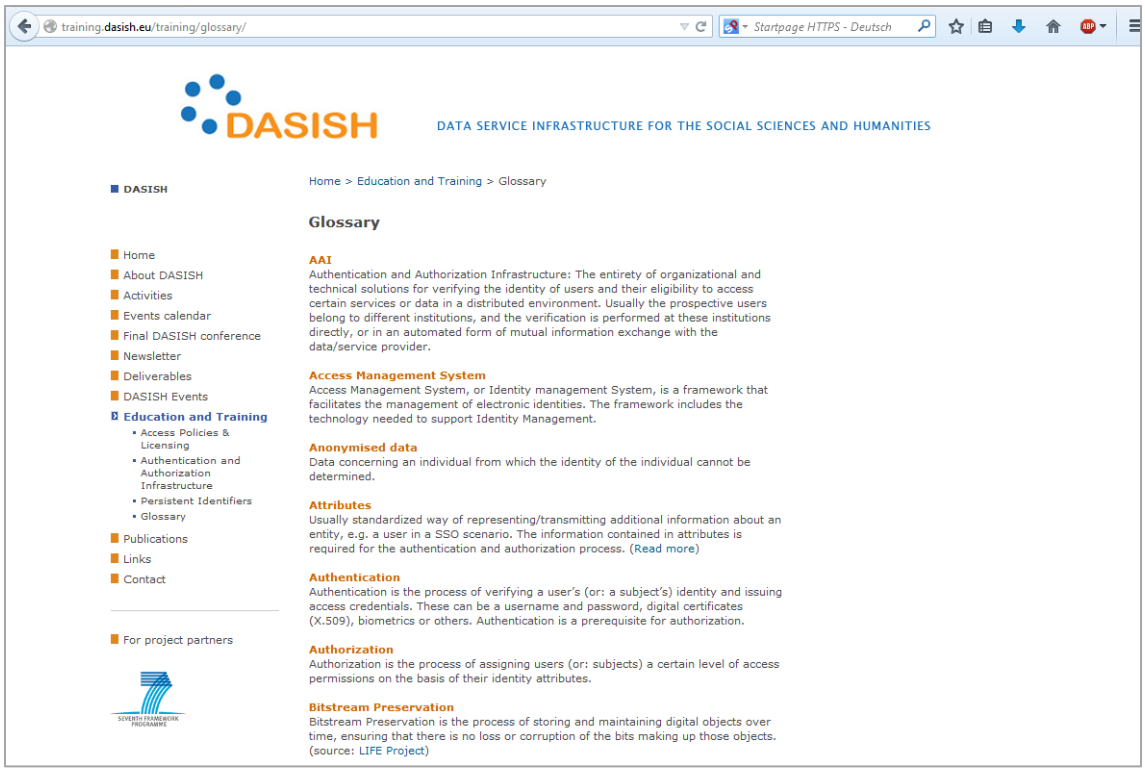


Figure 1: Screenshot of the Glossary

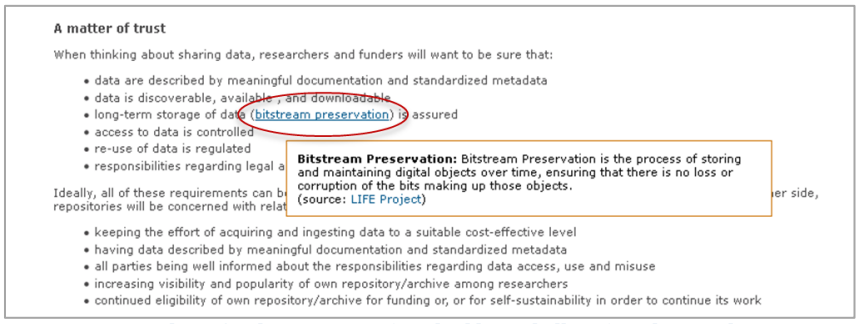


Figure 2: Screenshot of a glossary term (marked by red ellipse) and its explanation

The first module on "Access Policies and Licensing" contains a quiz (see Figure 3) and two exercises at the end, intended for self-testing of the users' knowledge after completing the training module. This concept was adopted from external training material that NSD had developed previously. However, user feedback indicated that this form of testing was not seen as appropriate for the target group (see also section 2c of this report). Accordingly, for the second training module on "Authentication and Authorization Infrastructures", an FAQ page was created instead of a quiz, addressing 12 common questions regarding the module topic in condensed form and providing hyperlinks to the specific sections of the module chapters for further information (see Figure 4).

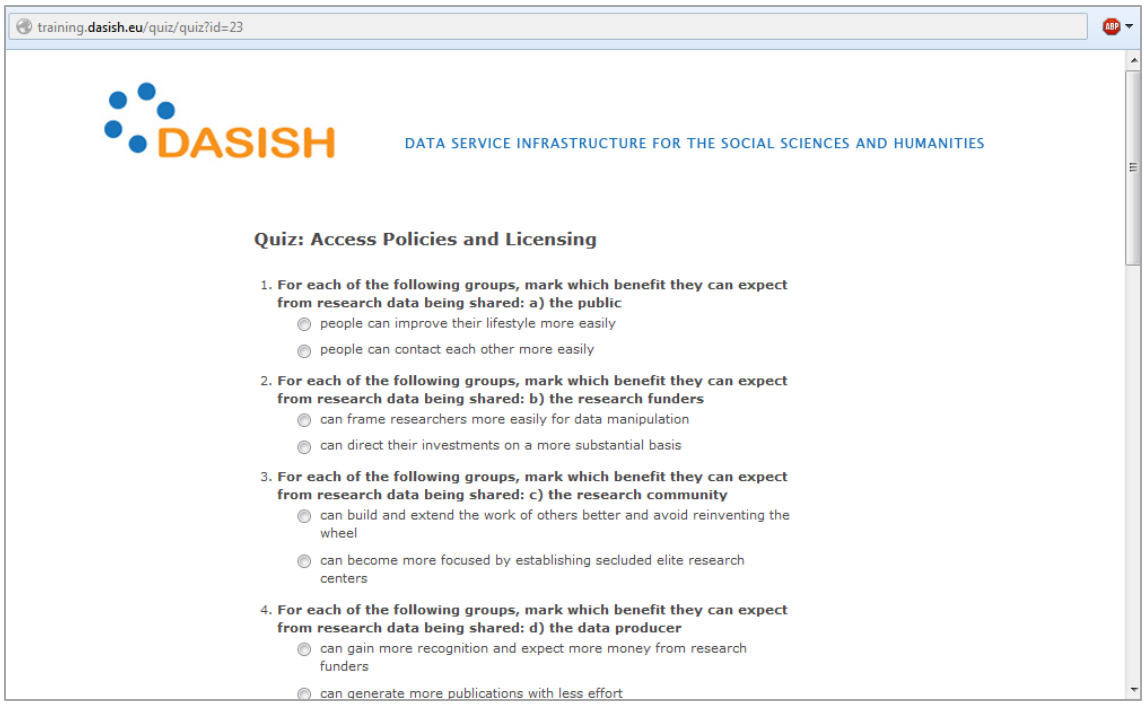


Figure 3: Screenshot of the quiz in the AAI module

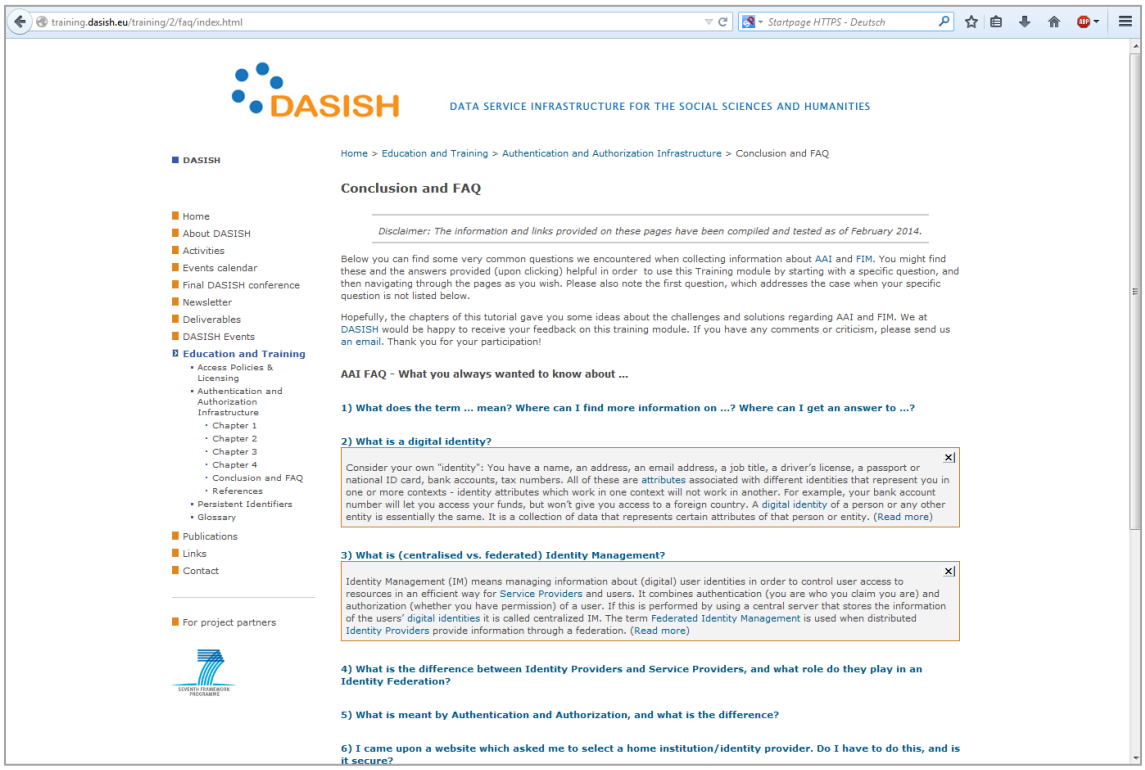


Figure 4: Screenshot of the FAQ in the AAI module

Each module can be read sequentially, by clicking on a hyperlink available on the bottom of each page (reading "Go to next page" or "Go to next chapter" respectively). Alternatively, a user can access each chapter and also sections of each chapter via using the tree menu on the left side of the training website (see Figure 5).

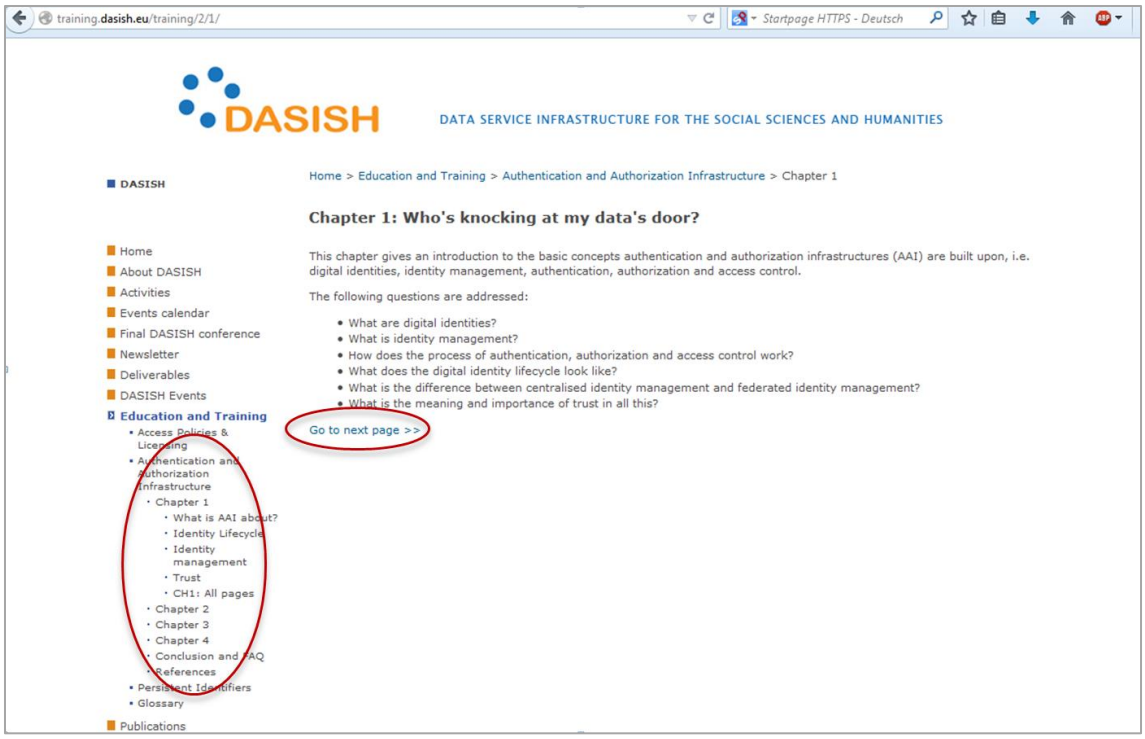


Figure 5: Means of navigating through the module (marked by red ellipses)

The modules were first presented internally to the DASISH consortium with explicit requests for feedback, comments or approval. After the internal feedback period, they were then made available publicly and announced by WP8 “Dissemination” through the DASISH website, the DASISH newsletter and also through other ESFRI channels by the WP8 ESFRI correspondents.
NSD agreed to host the modules via its OpenCMS instance for the near future, unless instructed otherwise. However, this does not include performing any content-wise maintenance or updates. These sustainability issues are addressed in section 2d of this report.

c. Feedback loops and revision processes

The first module on “Access Policies and Licensing” was created in an iterative process with WP7 internal feedback and reviews. It was presented to the DASISH consortium in January 2013 at the consortium meeting in Munich, and made available to the public in May 2013. A first revision took place in September 2013, when a few minor mistakes were corrected and ambiguities clarified. Another major revision was started in June 2014 as an effort of WP7 partner UiB, which included a significant restructuring of the entire module. Although the revision process is already in a very advanced state, it will most likely not be completed within the project runtime.

The second module on “Authentication and Authorization Infrastructures” was discussed intensively with members from Task T5.1 and revised internally several times. It was presented to the consortium in January 2014 at the consortium meeting in Barcelona, and made available to the public in March 2014. In particular, it included an FAQ section at the end instead of the Exercise section from the first module. This change

was motivated by user feedback stressing that the target audience of the DASISH training modules were not so much students and postgraduates, who would be more willing to complete a quiz, but rather researchers and scientific employees.

The third training module on "Persistent Identifiers" was also the result of an extensive revision process based on discussions with members of Task T5.2. The main part of the module was finished and presented to the consortium in November 2014 and made available online in December 2014.

A fourth training module on "Metadata Quality" was also developed on the basis of the first part of the WP5 deliverable D5.2 "Metadata Quality and Portal Progress Report". The internal revision process for this module, however, could not be completed.

d. Future plans and sustainability

The training modules created by WP7 and other contributors from DASISH will remain available to the public in their current state through the project website dasish.eu. However, in the long term, a useful provision of the information gathered and presented in the modules can only be provided through an institutional commitment which is not limited by project runtime and funding.

During the progress of DASISH, several requests for comments were sent to the ESFRIs by DASISH WP7, asking for the usefulness of the modules to the respective ESFRI's communities, desired adaptations and possible future hosting of the module content. However, no decision could be reached so far. Currently, negotiations are ongoing with partners from DARIAH-EU, who have already expressed their interest in using and maintaining the modules. Possible forms of publication would be to either integrate the modules into the DARIAH-DE Portal (similar to the already available eCodicology-example[7]), or to use the American platform for open educational resources OERCommons.org (where a DARIAH-group[8] has already been started). Because of the rather German focus of the first option and the international orientation of the second, the latter option is momentarily being preferred and evaluated.

Since DARIAH now also officially has the status of an ERIC, we hope that this discussion can be led to a successful end, and that the content and format of the DASISH training modules will form the basis of a sustainable online platform from which researchers and data managers from all SSH ESFRIs can benefit.

We hope that this will also provide a basis for finishing the work on the second revision of the first training module on "Access Policies and Licensing" as well as publishing the yet unfinished training module on "Metadata Quality".

[7] https://de.dariah.eu/ecodicology-tutorial
[8] https://www.oercommons.org/groups/dariah/229/

3. Online training modules

a. Access Policies and Licensing

When sharing research data, a number of legal and ethical considerations have to be taken into account (or at least clarified). They have a strong influence on the conditions under which data may be shared. For (potential) data re-users, it is important to have clarity in this respect; otherwise they might refrain from re-using the data. As a researcher wanting to share data, a good way to provide clarity is to attach a license to the data – so that (potential) re-users know what they are allowed to do with the data under what conditions. As a data archive or repository, rules for access to the data, i.e. data policies, have to be determined. These may differ depending on the type of data and the type of use.

The questions above are addressed in the first DASISH training module on "Access Policies and Licensing". As just explained, there are two major parties responsible for providing the research data in a form in which it can easily and safely (in a legal sense) be accessed and re-used: data creators on the one hand and infrastructure administrators and managers on the other. Therefore, both constitute the main target groups for this training module. A third target group are data re-users. The learning goals for the individual target groups are:

- Data creators:
  - to be familiar with legal and ethical considerations to be taken into account when creating data, particularly with regard to making the data available for re-use
  - to be familiar with the requirements and procedures when depositing data in a repository/data archive
  - to be familiar with licenses most commonly used in the SSH area and about the parameters to consider when choosing a license for their research data
- Infrastructure administrators/managers:
  - to be familiar with the legal and ethical considerations to be taken into account in terms of ingest and long-term preservation of research data as well as their provision for re-use
  - to be familiar with the measures that repositories/data archives should take in this respect, particularly in the context of ingest and access (e.g. deposit agreements, access policies, terms of use)
  - to be familiar with the requirements of sensitive data in terms of ingest, long-term preservation and the provision of access and measures that can be taken to make sensitive data (re-)usable
- Data re-users:
  - to be familiar with licenses most commonly used in the SSH area and the implications of license types and specifications on data re-use

The module is structured into seven chapters which are briefly summarized in the following.

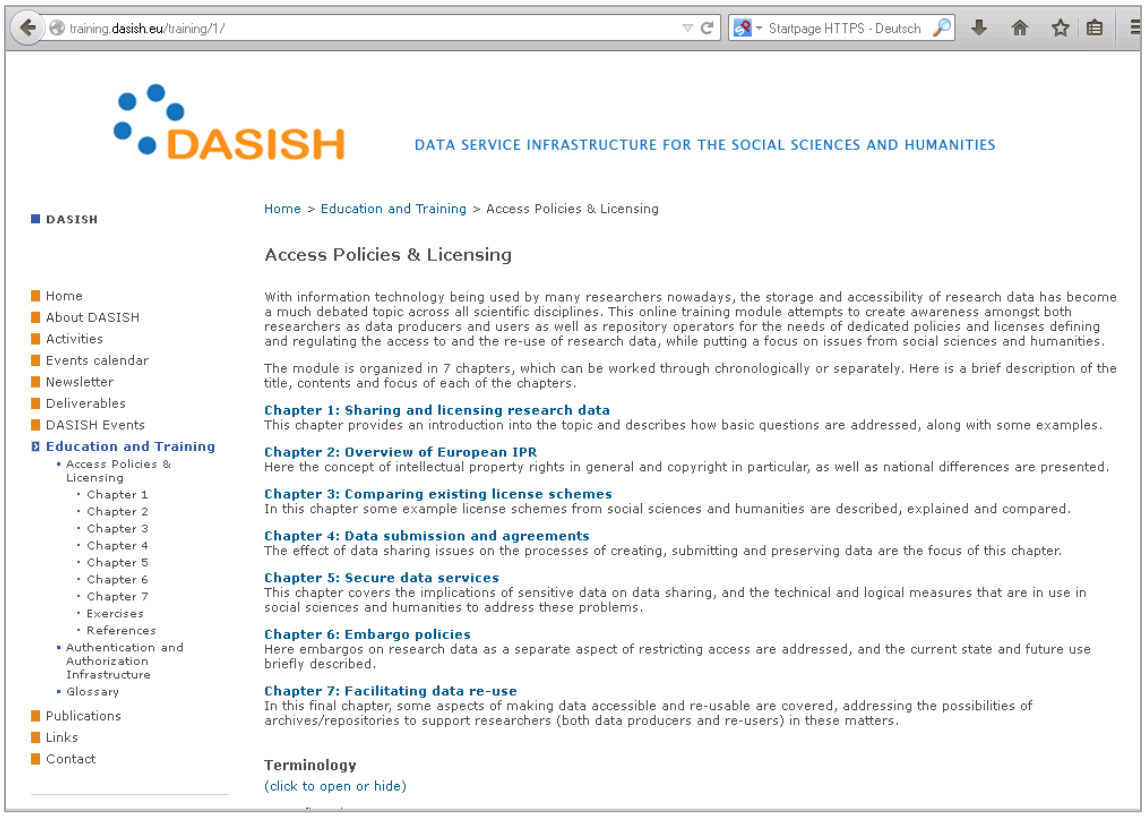


Figure 6: Screenshot of the introductory page of the module "Access Policies and Licensing"

Chapter 1: Sharing and licensing of research data

This introductory chapter to the module specifies the benefits of sharing data and explains why licensing research data is important. It introduces basic principles, actors and techniques involved in data sharing. It also addresses common concerns that many data creators prevent from sharing their data and offers advice and possible solutions. Furthermore, it looks at the relationship between repositories or data archives and users and gives some examples of repositories.

Chapter sections:

- Why share research data?
- Why license research data?
- A matter of trust
- The Data Lifecycle
- Examples for archives/repositories and their focus

Chapter 2: Overview of European IPR

Intellectual Property Rights (IPR) are a central issue in the context of data sharing and re-use. IPR legislation varies from country to country – not only worldwide, but also within Europe. This chapter gives an introduction to IPR in general and copyright in particular. It offers examples of national and European law that may apply to research. In addition, it discusses the question, if copyright does apply to data and explains under which conditions copyrighted material can be used for research.

Chapter sections:

- Intellectual property rights
- Copyright
- Databases
- Implications of data type and origin

Chapter 3: Comparing existing license schemes

Licenses are a means under which conditions and to what extent a copyrighted work may be used by others. By granting a license, the copyright holder transfers parts of the rights connected with copyright (i.e. the right of redistribution) to other people. This chapter explains in detail what licenses are and how they are used. It compares different types of licenses and gives examples of common license schemes currently in use in the European Social Sciences and Humanities domains.

Chapter sections:

- What is a license?
- Types of licenses
- Standard licenses
- Prepared licenses

Chapter 4: Data submission and agreements

In order for research data to be shareable and re-usable without problems, a thorough research data management throughout the whole research process is necessary. This chapter deals with the issues of IPR / copyright as well as the ethical questions that have to be addressed as part of this. Furthermore, it looks at the agreements that are concluded between depositor and repository before data is transferred into a repository or data archive. This document is often called deposit or license agreement and specifies the conditions of data transfer, storage, preservation and distribution for re-use, including the responsibilities of both parties involved.

Chapter sections:

- Submitting data: preliminary considerations
- On the archive's/repository's side: Things to consider in a policy
- On the researcher's side: Things to consider when creating and depositing data
- Considerations when creating data as part of research data management
- Deposit agreements
- Further reading

Chapter 5: Secure data services

In some disciplines, like the social sciences, a considerable amount of research data is of sensitive nature (for example, because it contains personal information about research objects). This puts constraints upon the possibilities to share and re-use such kind of research data. This chapter explains these constraints, and then focuses on the technical and logical measures that can be used to tackle these problems and facilitate use of confidential data. One of these is anonymization, another secure data services.

Chapter sections:

- Access to data
- Confidentiality in Quantitative Data
- Confidentiality in Qualitative Data
- Secure Data Services
- Examples

Chapter 6: Embargo policies

This chapter explains the concept of embargos and describes their current use. Embargos are a means of restricting access to research data for a certain amount of time. They are a measure to counteract a common concern that keeps many data creators from sharing their data: the fear that someone else will usurp the data that they have put so much work in.

Chapter sections:

- What is an embargo?
- Archives
- Funders
- Journals

Chapter 7: Facilitating data re-use

Chapter 7 finally puts its focus on data re-use. It describes how archives can support researchers as data producers in making their data accessible and re-usable, as well as data re-users in finding and accessing data. It also makes suggestions how re-use can be promoted. In addition, it deals with the questions, how legal requirements with regard to data protection and security can be met and what measures there are to pursue and sanction infringement of copyright, terms of use etc. or other kinds of misuse.

Chapter sections:

- Who will use the data?
- Access to research data in Social Sciences and Humanities
- Supporting data depositors in sharing their data
- Repository means for promoting re-use

Exercises

To give users the possibility to test and consolidate what they have learned, the module contains an exercise section at the end. There is a quiz with ten multiple choice questions (see Figure 7) and two more open exercises with tentative solutions.

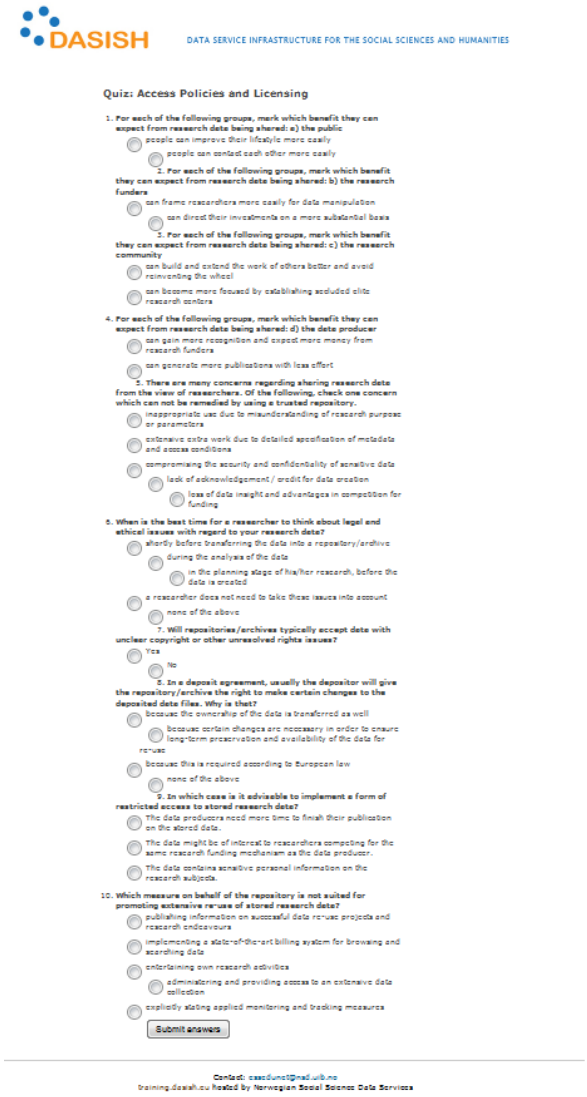


Figure 7: Quiz for training module on Access Policies and Licensing

b. Authentication and Authorization Infrastructure

In the recent years, we have seen an enormous growth of the amount of research data being produced and used. Along with that, the number of data centres and institutional repositories is increasing. Existing and emerging research infrastructures are becoming more and more interconnected. To access data and services, researchers nowadays do not only use their local infrastructures, but also repositories, data centres and other service providers. For these, the pool of potential users has increased significantly and now includes members of numerous institutions from all over the world.
For infrastructure facilities, this poses challenges in terms of user management and access control. This module deals with current approaches and solutions for user authentication and authorization, commonly referred to as Authentication and Authorization Infrastructures (AAI). It has a focus on Federated Identity Management (FIM), which is designed to cope with the developments just described. One of its most important features is to enable Single Sign-On (SSO), which allows users to log in and get access to resources and services of multiple providers with just one set of credentials

(usually provided through their home institution). At the same time, data archives, repositories etc. can make their resources available to a much larger group of users than before without having to manage additional identities.

The target groups of this module are administrators of data archives or repositories and institutional decision makers on the one hand, researchers or users of data archives or repositories on the other. The learning goals for these target groups are:

- Data archive / repository administrators:
  - to be familiar with basic concepts, terms and actors in the context of AAI
  - to be informed about the challenges and current solutions of Federated Identity Management in European SSH communities
  - to have an overview of technical challenges and solutions involved in implementing and managing FIM-based AAI
- Institutional decision makers:
  - to be familiar with basic concepts, terms and actors in the context of AAI
  - to be informed about the challenges and current solutions of Federated Identity Management in European SSH communities
- Data archive / repository users:
  - to be familiar with basic concepts, terms and actors in the context of AAI

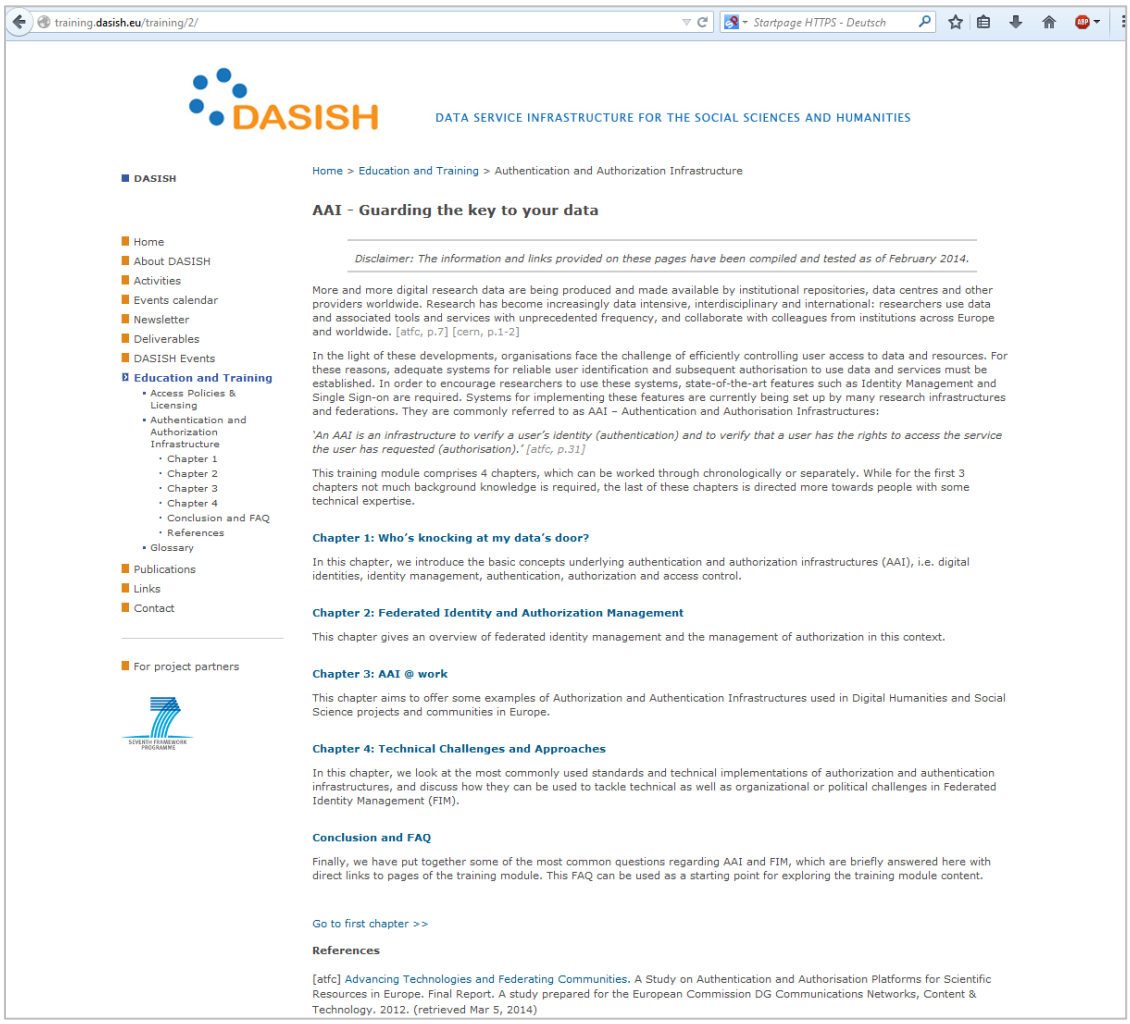


Figure 8: Screenshot of the introductory page of the module "Authentication and Authorization Infrastructure"

Chapter 1: Who's knocking at my data's door?
This chapter gives an introduction to basic concepts. Users do not use digital infrastructures in person, but with their digital identities. These consist of a number of attributes that are relevant to the specific context or system. The identity management administers the information about the digital identities of users and controls their access to the system. The three main functions of identity management are authentication, authorization and access control. Furthermore, the chapter explains the difference between centralised identity management and federated identity management and reflects on the role of trust in this context.

Chapter sections:
- What is AAI about?
- The Digital Identity Lifecycle
- Approaches to identity management
- Trust

Chapter 2: Federated Identity and Authorization Management
As the number of digital resources produced and used worldwide grows and the crosslinking within and between infrastructures increases, federated identity management (FIM) (and the function of single sign-on enabled by them) is becoming more and more important. After describing the basic mechanism behind FIM, this chapter has a look at the roles and responsibilities of and the relationship between the major parties involved in identity federations, namely Identity Providers and Service Providers. Identity management typically involves the processing of personal information. This poses challenges with respect to privacy and data protection, which are also looked upon in this chapter.

Chapter sections:
- Introduction: Why federation?
- Scenario and basic mechanism
- Trust in federations
- Privacy and Data Protection Issues

Chapter 3: AAI @ work
This chapter looks at authentication and authorization infrastructures in practice – with examples of AAI implementations from European SSH projects and communities that illustrate who is using AAI as well as why and how they are doing it. There are examples of services (OpenAthens, eduGAIN and eduroam), AAI solutions used in European infrastructure projects (CLARIN, DARIAH and EHRI) and communities active in the field of Federated Identity Management.

Chapter sections:
- Services: OpenAthens
- Services: eduGAIN
- Services: eduroam
- SSH Infrastructures: CLARIN
- SSH Infrastructures: DARIAH
- SSH Infrastructures: EHRI
- Communities and Conclusions

Chapter 4: Technical Challenges and Approaches
This chapter addresses the technological aspects of AAI solutions. There are different ways to realise authentication and authorization infrastructures technically. The most commonly used standards and technologies are described in this chapter, with a focus on the relations between them and their potential use to tackle not only technical, but also organizational and political challenges connected with Federated Identity Management. The particular aim is to provide information and guidance to technical staff, administrators or system architects who are planning to get involved (or already are involved) in the implementation of Web Service Security, and Federated Identity Management applications such as Single Sign-On.

Chapter sections:
- Web Services Security
- Single Sign-On
- SAML 2.0
- SAML Implementations
- Solutions for Authentication
- Solutions for Authorization
- Solutions for Authentication and Authorization
- Solutions for IdP Discovery

Conclusion and FAQ
For the final chapter, the key facts of the previous chapters have been extracted and condensed to be presented here in the form of Frequently Asked Questions.

c. Persistent Identifiers

As the Internet has transformed our capabilities and possibilities of finding and accessing information, it has also changed our expectations and demands regarding information findability and accessibility. If the information we want is too difficult to find, or if it is not available at all, we will look someplace else. As a consequence to information providers, who want others to use and re-use information they generated or provided, they will need to ensure that it can be easily found both by machines and by humans. Findability and accessibility of information has thus become an important aspect of information quality itself, apart from its actual content.

Apart from supplying rich metadata to enable findability, a challenge here is to ensure that references to online information remain valid. If referred-to information is being removed, modified or updated, the references are pointless or misleading.

Meeting these concerns requires providing functionality or a system that makes it possible for humans to find and cite sources found on the Internet, and it is important that this system contains information that makes it possible for machines, or technical systems, to utilize and connect the information. One strategy is to add identifiers to resources that will be preserved, often called Persistent Identifiers.

This training module gives an introduction into this topic, touches upon common problems and requirements, provides examples of systems in use and their distribution, and gives references for further reading.

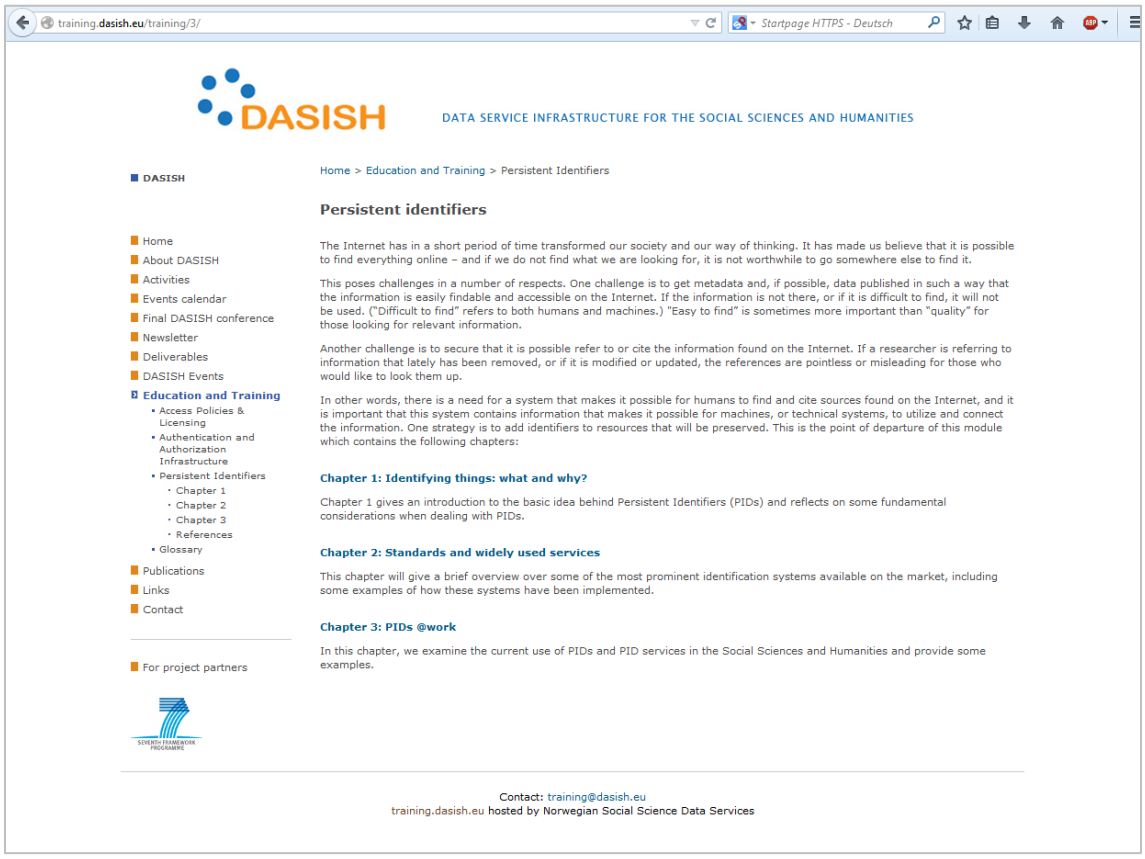


Figure 9: Screenshot of the introductory page of the module "Persistent Identifiers"

The audience addressed by this module includes SSH researchers with little knowledge of PIDs, data archivists who want to start obtaining PIDs, technical implementers or administrators of PID systems, and also policy makers in research infrastructures, research funding organizations, library and publishing sectors.

The learning goals for these target groups are:

- SSH researchers:
  - to be familiar with the general functionality of PID services
  - to be able to recognize examples from different commonly used PID systems
  - to be familiar with basic requirements of commonly used PID systems
- Data archivists:
  - to have a better overview of the different functionalities of commonly used PID systems
  - to be informed about which PID systems are used or recommended for use in which SSH community
- Technical staff
  - to be informed about which PID systems are used or recommended for use in which SSH community

    - to be aware of basic SSH specific requirements regarding the use of PID services
- Policy makers:
    - to be familiar with basic requirements of commonly used PID systems
    - to have a better overview of the different functionalities of commonly used PID systems

Chapter 1: Identifying things: what and why?

Chapter 1 gives an introduction to the basic idea behind Persistent Identifiers (PIDs). It starts out by motivating the use of PIDs through examples of attempts to access resources which are not available anymore, or cannot be located. The chapter goes on to explain the conceptual functioning of PIDs and fundamental terms such as "Reference", "Resolution" and "Intermediary". Taking the introduction one step further, Roles such as "Identifier Provider" and "Identifier Manager" are explained. Finally, the chapter reflects on some fundamental considerations and requirements when dealing with PIDs.

Chapter sections:

- Why do we need Persistent Identifiers?
- What is a PID and how does it work?
- Roles and responsibilities associated with the setup and maintenance of a PID service and the assigning of PIDs
- Considerations and Requirements when assigning identifiers
- Requirements relating to identifier providers, managers and institutional decision makers

Chapter 2: Standards and widely used services

This chapter provides a brief overview over some of the most prominent identification systems available on the market, namely ARK, Handle and Handle-based systems, PURL, URN and ISLRN. For each of these systems, a general description of format and usage is given and explained with an example PID, complemented by references and further reading. To conclude this overview, tabular comparisons of the different PID systems according to example format, Costs and Resolution are given.

Chapter sections:

- Archival Resource Key, ARK
- Handle, DOI, EPIC
- Persistent Uniform Resource Locator, PURL
- Uniform Resource Name, URN
- International Standard Language Resource Number, ISLRN
- Comparison tables

Chapter 3: PIDs @work

The last chapter takes a look at how PID systems are being used in different institutions and communities in the Social Sciences and Humanities. This partial overview is separated into looking at SSH infrastructures (in particular, the five DASISH ESFRIs), libraries, archives, museums and SSH projects. For each of these groups, the distribution of PID systems in use as well as existing PID policies are being described. References to

chapter 2 are included to directly refer back to a more detailed description of the different PID systems.

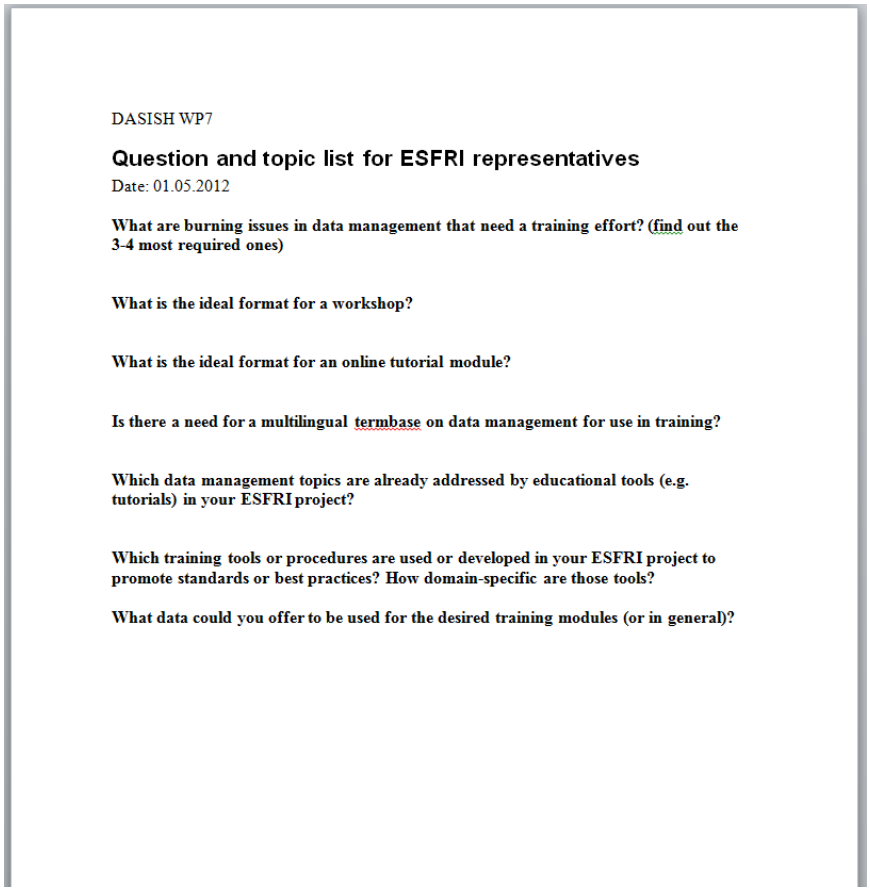
Chapter sections:

- SSH Infrastructures
- Libraries
- Data Archives
- Museums
- SSH Projects

Appendix

Appendix A

A1: First Questionnaire for SSH ESFRI members regarding training topics and measures



DASISH WP7

Question and topic list for ESFRI representatives

Date: 01.05.2012

What are burning issues in data management that need a training effort? (find out the 3-4 most required ones)

What is the ideal format for a workshop?

What is the ideal format for an online tutorial module?

Is there a need for a multilingual termbase on data management for use in training?

Which data management topics are already addressed by educational tools (e.g. tutorials) in your ESFRI project?

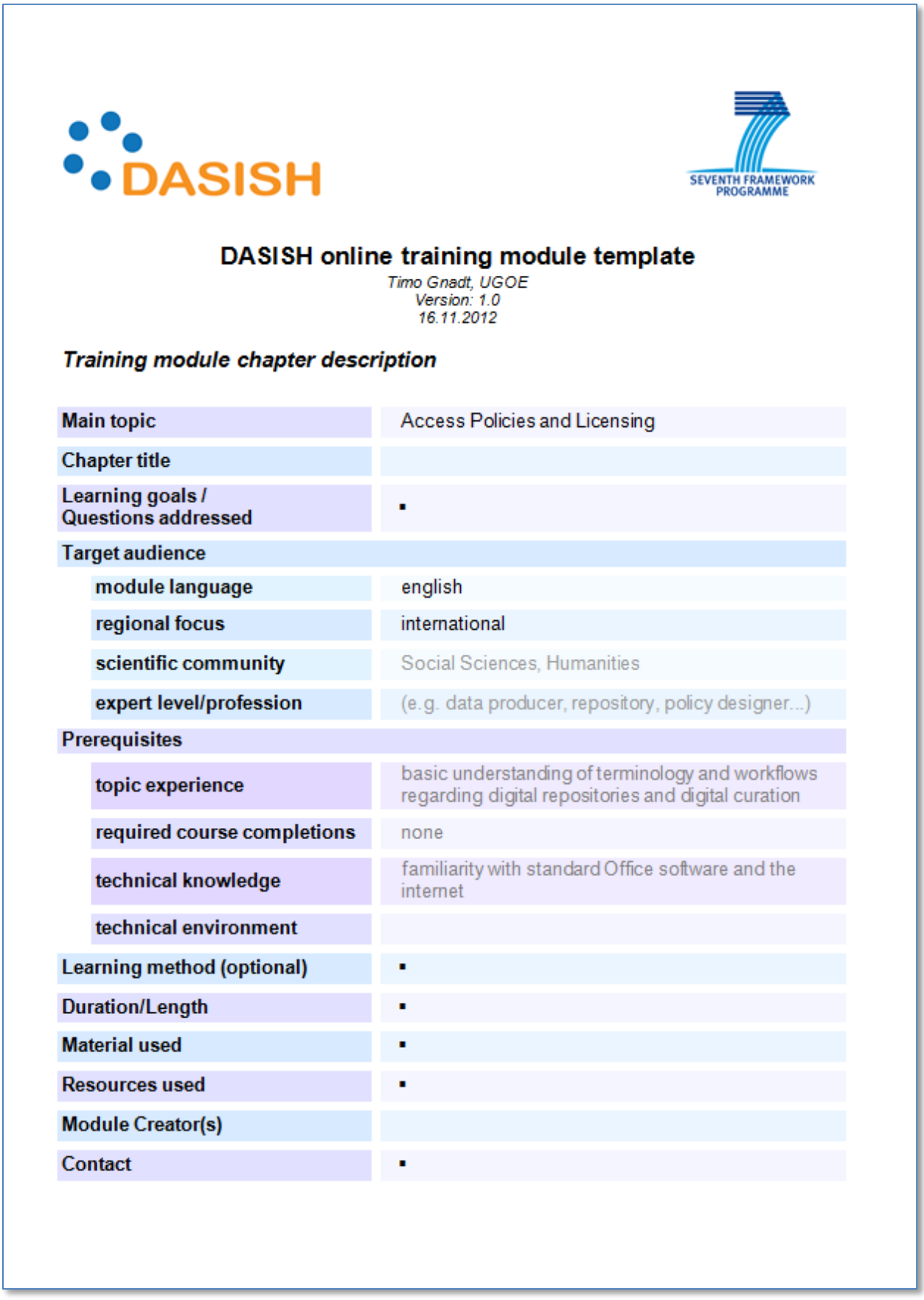
Which training tools or procedures are used or developed in your ESFRI project to promote standards or best practices? How domain-specific are those tools?

What data could you offer to be used for the desired training modules (or in general)?

A2: Second Questionnaire for SSH ESFRI members regarding training topics and measures

Date:		ESFRI project:			ESFRI representative:			DASISH representative:	
main topic	subtopics	scope / target group			desired activity/material from or through DASISH		material available from or through the ESFRI project		comments
		introduction for data users	introduction for data creators	specialization	tutorial	workshop	tutorials / exercises	available data	
		Please leave blank for no interest or enter: 1 for minor interest 2 for strong interest			Please enter yes or leave blank				e.g.: focus on national issues, specific community, links to available material, ...
Data enrichment	Data enrichment in general								
	Metadata: existing standards, usage								
	Metadata: effect of data type								
	Metadata: harvesting, storing and sharing								
	Linked data								
Data quality	Data quality in general								
	Data analysis/harmonisation/mining								
	Workflows								
	Glossary / international termbase								
Data archiving	Data archiving in general								
	Deposit services and SLA negotiations								
	Bitstream preservation								
	Certification								
	Persistent Identifiers								
Security & Policies	Security & Policies in general								
	Access policies								
	Licensing								
	Sensitive data / anonymization								
	Data security								
	Authentication & Authorization Infrastructure								
Management	Management in general								
	Funding and Project proposals								
	Publication of data (e.g. Open Access)								

Appendix B: Template for describing and creating training module chapter content



DASISH online training module template

Timo Gnadt, UGOE
Version: 1.0
16.11.2012

Training module chapter description

Main topic	Access Policies and Licensing
Chapter title	
Learning goals / Questions addressed	▪
Target audience	
module language	english
regional focus	international
scientific community	Social Sciences, Humanities
expert level/profession	(e.g. data producer, repository, policy designer...)
Prerequisites	
topic experience	basic understanding of terminology and workflows regarding digital repositories and digital curation
required course completions	none
technical knowledge	familiarity with standard Office software and the internet
technical environment	
Learning method (optional)	▪
Duration/Length	▪
Material used	▪
Resources used	▪
Module Creator(s)	
Contact	▪

[Access Policies and Licensing]

Chapter 1: [Chapter title]

[Chapter 1 content]

Exercises

[Exercises Chapter 1]

[-- New page --]

Glossary

[Glossary content]

[-- New page --]

References

[References]

Appendix C: Acronyms used in this report

AAI	Authentication and Authorization Infrastructure
CESSDA	Consortium of European Social Science Data Archives
CLARIN	Common Language Resources and Technology Infrastructure
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DASISH	Data Service Infrastructure for the Social Sciences and Humanities
EHRI	European Holocaust Research Infrastructure
ESFRI	European Strategy Forum on Research Infrastructures
ESS	European Social Survey
FIM	Federated Identity Management
GESIS	GESIS - Leibniz-Institut für Sozialwissenschaften, Germany
KCL	King's College London, UK
MPG-MEA	Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. - Munich Center for the Economics of Aging, Germany
NSD	Norsk Samfunnsvitenskapelig Datatjeneste AS, Norway
OEAW	Österreichische Akademie der Wissenschaften, Austria
PID	Persistent Identifier
SAML	Security Assertion Markup Language
SHARE	Survey of Health, Ageing and Retirement in Europe
SSH	Social Sciences and Humanities
UGOE	Georg-August-Universitaet Göttingen Stiftung Öffentlichen Rechts, Germany
UiB	Universitetet i Bergen, Norway