

netCommons
Network Infrastructure as Commons

Deliverable 7.3 **Data Management Plan (v2)**

Deliverable Number D7.3
Version 1.0
June 24, 2018



Co-Funded by the Horizon 2020 programme of the European Union.
Grant Number 688768



Project Acronym: netCommons
Project Full Title: Network Infrastructure as Commons.
Call: H2020-ICT-2015
Topic: ICT-10-2015
Type of Action: RIA
Grant Number: 688768
Project URL: <http://netcommons.eu>

Editor: Renato Lo Cigno, UniTN

Deliverable nature: Report (R)

Dissemination level: Public (PU)

Contractual Delivery Date: December 31, 2017

Actual Delivery Date: June 24, 2018

Number of pages: 34

Keywords:

Authors: Roberto Caso, UniTN
Christian Fuchs, UoW
Federica Giovanella, UniTN
Roberta Guidolin, UniTN
Renato Lo Cigno, UniTN
Leonardo Maccari, UniTN

Peer review: Melanie Dulong de Rosnay, CNRS
Francesca Valentini, UniTN
Virginie Aubree, UniTN

History of Revisions

Rev.	Date	Author	Description
v0.1	20/12/2017	Renato Lo Cigno	First draft derived from D7.1
v0.4	20/04/2018	Leonardo Maccari	Zenodo usage added
v0.5	15/05/2018	Renato Lo Cigno	Discussion on Data Copyright started
v0.9	15/06/2018	Roberto Caso	Complete Discussion on Data Copyright
v1.0	22/06/2018	Renato Lo Cigno	Final organization of the document and proofreading

Executive summary

This document presents the policies adopted by the project consortium to ensure that rules on Open Access to Publications and Open Research Data are met. It extends D7.1[1] and complete both the discussion on Open Science and the details of the copyright contracts by major publishers, as some of these have changed since the publication of D7.1 A technical (in legal terms) discussion on the process of publishing Open Data to achieve Open Science is also included in the document to highlight the conceptual problems that may be encountered, by netCommons, but also by any other research project, in managing data repositories as well as using existing ones due to the fuzziness of the legal framework in the sheer definition of “data”. This part is entirely novel compared to D7.1

This deliverable sets the final principles used by the consortium in managing the data produced during the project and the policies and technical measure taken to protect people privacy and data security in case personal data is collected during the research. In particular this Deliverable details the procedures that the consortium adopts to make all scientific publications, both peer-reviewed and non peer-reviewed, fully available according to Open Access best practices, guaranteeing that they are correctly archived in multiple repositories, available through the project web site and properly indexed through metadata.

Regarding Data Management, the document details the types of data that netCommons is handling and will handle till the end of the project, and the procedures undertaken to guarantee that no personal or sensitive data are discovered beyond the boundary clearly defined during the data collection procedure, and that this boundary, as well as the person responsible for the data collection and protection, are clearly understood by the people involved in the data collection.

Contents

1. Introduction	7
2. netCommons Open Access Policy	8
2.1. Data, Property & Copyright	8
3. Open Access to scientific publications	11
3.1. Documents Subject to Open Access	11
3.1.1. Green and Gold Open Access	11
3.1.2. Open Access Repositories	12
3.1.3. Accepted version and published version	12
3.2. Implementation of the Open Access Policy to Publications	12
3.2.1. Procedures for PR publications	12
3.2.2. Procedures for NPR Publications	13
3.3. Current Policies by some of the Major Scientific Publishers	14
3.3.1. Elsevier	14
3.3.2. ACM	15
3.3.3. IEEE	16
3.3.4. Springer	16
3.3.5. IFIP	17
3.3.6. SAGE	18
4. netCommons Open Data Policy	19
4.1. Licenses Analysis and Platform Selection	19
4.2. Using the Zenodo Interface	21
5. Data Security and Privacy Provisions in netCommons	27
5.1. Scope of the privacy/security model	27
5.2. General Principles	27
5.3. Security Framework	28
5.3.1. Authentication	28
5.3.2. Authorization	28
5.3.3. Accounting and Auditing	28
5.4. Summary of Technological solutions	29
6. Conclusions	30
Bibliography	32
A. Appendix: Template of the Informed Consent Form	33

List of Figures

4.1. Zenodo log-in interface.	21
4.2. Zenodo interface for uploading new material.	22
4.3. List of uploads the user started in Zenodo.	23
4.4. Zenodo interface for files upload.	24
4.5. Zenodo form for data fields.	25
4.6. Zenodo interface for optional fields.	26

List of Acronyms

ACM	Association for Computing Machinery
CA	Consortium Agreement
CAPS	Cooperative Awareness Platforms for Sustainability
CERN	Conseil Européen pour la Recherche Nucléaire
CN	Community Network
DL	Digital Library
DOI	Digital Object Identifier
EC	European Commission
FAQ	Frequently Asked Question
GA	Grant Agreement
IEEE	Institute of Electrical and Electronics Engineers
IFIP	International Federation for Information Processing
NIH	National Institutes of Health
NPR	non peer-reviewed
OA	Open Access
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OKI	Open Knowledge International
ODbLv1.0	Open Database License version 1.0
ORD	Open Research Data
ORDP	Open Research Data Pilot
PDF	Portable Document Format
PR	peer-reviewed

1. Introduction

netCommons operates only marginally with personal data, while in most cases the data produced by the project refer to technical measures or experiments not involving human beings. Nonetheless netCommons seeks the adoption of proper security and confidentiality standards for the data collected as well as proper Open Access (OA) policies to maximize the impact of the research carried out, as we are well aware that at the heart of modern research is an extensive scientific dialogue, with a timely sharing of data and experiences.

Proper data sharing accelerates innovation, allows researchers to build on previous work improving the quality of the results, fosters collaboration and avoids duplication of work. The necessity of Open Access and Open Research Data (ORD) adoption has gained momentum and it is influencing the political choices of all the main public agencies funding and sponsoring research. The European Commission (EC) is no exception to this general international trend, which has been first spawned in the U.S. by the National Institutes of Health (NIH). The commitment of the EC toward Open Access of the research results is reflected in official guidelines [2] and in the wording of Grant Agreements (GAs) (e.g., Art. 29.2). In addition, from the specific nature of netCommons and from its being part of the “societal challenges” programme, we derive a particular emphasis on the involvement of citizens, economic stakeholders, governmental agencies and charities. All these considerations require the adoption of liberal standards for the scientific dissemination of information, in accordance with the mandate in Art. 29.2 of the netCommons grant agreement.

In order to avoid problems and misunderstandings and to streamline the whole process of data collection and of dissemination of results, this document seek to define clear guidelines on how to treat data and on how to disseminate the results. This document is extended following the EC Guidelines [3], but it also includes a discussion of the current regulatory framework and the ambiguities it still contain, as well as an analysis of the copyright contracts available from the major scientific editors and how they may affect Open Access policies.

2. netCommons Open Access Policy

netCommons is part of the H2020 Open Data Pilot [4], thus the access policy to the project result must deal both with the publications produced by the project and with the data upon which these publications are based. Moreover, given the interdisciplinary approach of the project and its societal importance, we foresee additional data to support general findings and to build a base for dissemination of the project outcomes, as well as setting the ground to build the advocacy capabilities and support the impact-oriented actions of netCommons.

One of the key challenges for a Cooperative Awareness Platforms for Sustainability (CAPS) research project like netCommons is to produce scientific knowledge that is persistent, that goes beyond the restricted scientific communities and that fosters the benefit of the individuals, of the communities and of the European society at large. Furthermore, having its roots in Internet Science [5, 6] netCommons findings are conceived to foster and benefit the development of Community Networks (CNs) also beyond the European Union.

These ambitious goals require a thorough dissemination activity of the research results, and a careful management of general data, including the information collected, to maximize the impact of the project efforts. For this reason, netCommons has opted for, and included in the Consortium Agreement, a fully open model of results and documents dissemination, including deliverables that are all public.

We treat in two separate Chapters, namely Chapter 3 and Chapter 4 the two topics of:

1. Open Access to scientific publications,
2. Open Access to research data,

as they definitely have different goals, problems, and facets, but before delving with the details of our management policy we discuss in Sec. 2.1, the implications of data management policies on open science and their relationship at a theoretical level.

2.1. Data, Property & Copyright

The topic of Open Research Data publication is much less debated, understood, and agreed upon compared to scientific publication Open Access, where a long tradition of copyright law and regulation helps framing the discussion and clearly defines moral ownership, property exploitation and the like. In particular, the license of Data (open or not) is far more difficult, as Data are not subject to standard Intellectual Property rules. Moreover, the sheer definition of data property rises issues that are closer to problems of industrial property (the difference between an invention and a discovery), rather than to works of the intellect and genius. One may simply recall the problem whether the decoded data describing a person genome belongs the person himself or to the person/team/industry decoding it.

From a regulation point of view, the concept of “data” is very complex. Currently we do not have a legal and uniform notion of “data”, although the word “data” is mentioned in a high number of laws. In fact, data subject to many different legislations, such as intellectual property law (copyright and database *sui generis* right) and trade secret; privacy and personal data protection; competition law; consumers’ protection laws; Public Open Data (access to re-usable public sector information), and more.

Such a complex setting is worsened by the inconsistent approach of European Union policies on data. On the one hand, the EU encourages and fosters Open Access and Open Sciences; on the other hand, it enacts or plants to enact laws on intellectual property and on trade secrets that actually favor private control over data. In this framework, public entities that grant funding as well as scientific and academic institutions make use of policies (such as H2020 projects’ regulation), grant agreements and open licenses—such as Creative Commons

Licenses—with the goal of giving way to data-sharing, applying to data the same logic that lies behind the idea of Open Access to scientific publications.

This is however a “second best” solution: given the lack of a consistent and systematic legal regime on data openness, the use of contract remains the best possible solution whose final goal is to trigger virtuous processes of data sharing both within the scientific community and outside it so to make it available for the society at large, for citizens and for firms.

As a second best solution, it solves some problems, but at the same time it creates new ones. For instance, there is the problem of coordinating different open licenses related to different dataset that belong to different academic, scientific or entrepreneurial institutions. It may happen that each institution for its own dataset a given license that differs from those used by other institutions working in the same field. By the way, the biggest problem is probably represented by the role of non-profit organizations. When open access laws and technologies started to be developed—around twenty years ago—non-profit academic and scientific institutions were thought as main characters in the process of building open science. Today, we must ask ourselves whether these institutions actually have the force and the will to represent an instance independent from the market. To say it better, we should ask ourselves whether Open Science can be a tool to defend the autonomy of science and academic freedom or it is instead only going to be another gearwheel of market mechanism—more precisely of the huge commercial platforms that govern the Internet through a de-facto oligopoly. We should consider four main aspects if we want non-profit science to play an important role to defend its own autonomy and freedom through open access and open data management, keeping in mind that “open” is often a tricky word as the flourishing of “pay-to-publish” journals selling as Open Access, but actually hampering the peer review process show. These aspects are:

- a. The governance of the Internet and the management of data shared through on-line platforms. The Internet seems to increasingly betray the spirit that existed at the dawn of the Net, as it is more and more in the hands of a few incumbent operators and information intermediaries. In this sense it is very important to properly understand and govern the data sharing and distribution models (through dedicated and appropriate licenses) rather than simply mandate or suggest the “open” publication of data that can actually be monopolized (de-facto) by the incumbents managing connectivity and information flows.
- b. The reform of the systems for the evaluation of science. The current system centers on two main points: First, the application of business logic (“quality assurance”) to scientific works in order to ensure their quality; second, the use of metrics in order to substitute—de facto—peer evaluation with rankings based on quantitative indexes often derived through secondary, non-controllable metrics. Both methods lead to the poisoning of the evaluation process and yields to a science that falls into line with the trends of the moment, with the needs of the market and with the mainstream thought. It is essential to give transparency to the entire process of evaluation, put human judgment at the center and delete the incentives that provoke the logic of quantitative “publish or perish,” and the marketing of papers rather than scientific results.
- c. The teaching of critical thinking. The emphasis on science as a product (scientific publications, patents, datasets, software and so on) moves the value of oral transmission of knowledge by teaching and research communities to the background. There is the need to put the teaching of critical thinking back at the center of science.
- d. The reform of intellectual property law. The proposal for a reform of copyright in the digital single market (Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market – COM(2016)593) currently pending before the EU parliament is a paradigmatic example of the umpteenth legislative tool that goes against scientific progress. The proposal contains both norms that heavily threaten the neutrality of the Internet and freedom of speech and norms that do not seriously help the development of the openness of science. An organic reform of intellectual property to help the development of science it is highly desirable.

While these four points seems to attain more to the sphere of tradition intellectual property, rather than data, today they influence more the publication of data, where there is far less tradition and background, rather than

the publication of articles and books, where the long legislation history makes at least quite clear who owns what and what sentences like “share alike” or “no derivatives” means. When data is concerned, as we discuss in some detail in Chapter 4, the concepts of derivation and manipulation are still not completely well defined, as they may even imply the possibility of falsification: Indeed, if a scientist is accused of data manipulation it normally means he has falsified experimental data to twist it to support one conclusion rather than another one, so in general one would go for “no derivatives” licenses. At the same time, a different analysis, e.g., different clustering or different correlation parameters, of data can also fall into the category of manipulation and derivative, so there is the risk that some legal wording of licenses can be interpreted in such a way as to restrict the freedom of research and science.

3. Open Access to scientific publications

One of the cornerstones of our dissemination strategy is to secure a timely and regular publication of the scientific findings in peer-reviewed, high impact journal and conferences. This will ensure a proper consideration of netCommons results in the scientific communities of interest. All scientific publications will be available in Open Access, providing archival Portable Document Format (PDF) versions of the published document. As specified in the H2020 Guidelines on Open Access publishing [2], by this term we mean the practice of providing free and unrestricted access to scientific publications to read and download. Sec. 3.3 provides a detailed analysis of all the major scientific publishers copyright policies concerning both Gold and Green Open Access (see Sec. 3.1.1 for proper definitions).

According to the contractual obligations specified in the GA Art. 29.2, “*Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.*” We will obviously comply with this obligation and we are implementing a specific policy and best practice based on OpenAIRE Sec. 3.2 to ensure and almost automatic propagation of the Open Access version of the publication to repositories compliant with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) standard [7] and to the web site of the project.

3.1. Documents Subject to Open Access

The Open Access policy for scientific publications *applies whenever a partner of the project or a group of partners decides to produce a scientific publication* containing the results of a research activity. This decision is taken on the following grounds:

- The publication is scientifically relevant and brings forth significant advances in the state of the art of the interested discipline;
- The data contained in the publications fulfill the requirements specified by the Ethical Committees of the partner/partners that collected the data (if applicable).

It must be noted that, due to the societal and open nature of netCommons research, as provided by the Consortium Agreement (CA) in Sect. 8 and in particular in Sect. 8.3.1, netCommons Parties are not subject to prior notice to the Consortium or any other legal body.

The Open Access policy does not apply to partial results which are produced at intermediate steps of the project and are not deemed scientifically relevant.

3.1.1. Green and Gold Open Access

The H2020 guidelines [2] refer to the two main procedures to enforce Open Access to scientific literature.

The Green Open Access: this procedure is based on re-publishing (often indicated as self-archiving) of the published article or the final peer-reviewed manuscript without the graphical imprints of the commercial publisher. Some journals also allow the deposit of the published version with the publisher imprinting. The manuscript is archived into an OAI-PMH-compliant repository (see Sec. 3.1.2) by the authors; some publishers could require an *embargo* period of time before the paper is made concretely available to the public: netCommons will try to minimize both the use of publishers that require an embargo and the duration of the embargo, that will in any case abide to the requirements of the Commission [2] as stated in Art. 29.2 of the GA.

The Gold Open Access: the article is provided in Open Access directly by the publisher, which normally (but not always) enables also re-publishing with the same means of the 'green' method. We note that while some publishers (most notably the International Federation for Information Processing (IFIP)) maintain a fully open Digital Library (DL) without any fees, many others require a fairly expensive fee to publish in Open Access. Many scientific communities regret and discourage 'pay-to-publish' procedures, specially in mixed publication venues (i.e., journals that allow both traditional and OA publications) where authors must declare their desire to publish in Open Access before the peer-review.

3.1.2. Open Access Repositories

A repository for scientific publications is generally defined as an online archive, but this condition is not enough to make a repository Open Access. The most known Open Access repository is probably arXiv (<http://arxiv.org>), maintained by the Cornell University. The H2020 guidelines give full freedom on the choice of the repository: it can be an Institutional Repository or a subject-based centralized repository. If the Institution the authors belong to does not have a specific infrastructure of this kind, the EU is funding the OpenAIRE effort (<http://www.openaire.eu>), which provides APIs to a comprehensive list of public repositories and in general means to foster Open Access policies. OpenAIRE plays a central role in netCommons best practices for Open Access, since it provides means to automatically link the repositories of most institutions, and it can thus be used to provide suitable visibility and linking to all the published material. In particular the Zenodo (<http://www.zenodo.org/>) repository is strictly related to OpenAIRE and is maintained by CERN, thus providing a suitable means for archival for all European institutions that cannot (or have not yet) set up an institutional repository. Other lists of repositories and further information on Open Access are available at <http://roar.eprints.org> and <http://www.openoar.org/>.

3.1.3. Accepted version and published version

An *accepted paper* is a version which has been revised by the author to incorporate review suggestions, and which has been accepted by the publisher for publication.

The final, *published version* is the reviewed and accepted article, with copy-editing, proofreading and formatting added by the publisher.

3.2. Implementation of the Open Access Policy to Publications

The Open Access policy will be applied both to peer-reviewed publications (i.e., publications that are evaluated by "peers") and to other types of publications such as books, white papers, and all other documents that the consortium deems valuable of dissemination. In the following we refer to the first type of publications as "peer-reviewed (PR)" and to the second as "non peer-reviewed (NPR)". Deliverables will be initially available through the project web site in the specific area collecting all of them¹. After review it will be decided if they deserve dissemination through OAI-PMH compliant repositories (Sec. 3.1.2).

3.2.1. Procedures for PR publications

The authors of netCommons publication have the freedom to opt for either a Green or for a Gold policy. In case of a *Green Open Access policy* the procedure is as follows:

¹netCommons has designed and implemented a specific page (<https://netcommons.eu/?q=content/deliverables-page>) where all deliverables often also in preliminary form are made available to the research community as working documents, and where it is possible to follow clearly the development of the project.

1. As soon as the paper is accepted, the draft of the accepted paper is stored in one or more repositories of the authors' choice among those supported by OpenAIRE along with bibliographic metadata;
2. The paper publication is notified to the project coordinator and to the exploitation and dissemination list (netcommons.wp6@list.disi.unitn.it);
3. Within a few days the manuscript becomes visible automatically through OpenAIRE reporting the proper reference to netCommons;
4. A script parses OpenAIRE daily (or weekly) to retrieve novel manuscripts and upload them automatically on the netCommons web site in the proper section;
5. If requested by the publisher, the paper is left unpublished for the duration of the embargo period; such period cannot exceed 6 months or 1 year in exceptional cases;
6. After the embargo period expires, the Open Access is granted to every one via the repository;

This procedure guarantees the highest visibility and dissemination as well as consistent and coordinated referencing, linking and availability.

In case of a *Gold Open Access policy* the procedure is:

1. As soon as the paper is accepted, and according to the publisher's Open Access policy, the draft of the accepted paper is stored in a repository of the authors' choice among those supported by OpenAIRE along with bibliographic metadata;
2. The paper publication is notified to the project coordinator and to the exploitation and dissemination list (netcommons.wp6@list.disi.unitn.it);
3. Within a few days the manuscript becomes visible automatically through OpenAIRE reporting the proper reference to netCommons;
4. A script parses OpenAIRE daily (or weekly) to retrieve novel manuscripts and upload them automatically on the netCommons web site in the proper section;
5. After the final publication the authors also add the publisher digital library information to ensure that the gold access policy is correctly advertised and accomplished, the publisher may request a different version to be uploaded.

The costs incurred for publication are eligible for reimbursement as long they are incurred before the end of the project; however netCommons will try to avoid all venues that apply publication fees that can rise suspicions that the publication does not follow an ethically consistent peer-review process. If the publication of a work supported by netCommons with a publisher that does not comply with EU rules is deemed by the Management Board of the utmost importance for its dissemination, the netCommons Coordinator will write a formal request to the publisher to comply with EU regulations.

3.2.2. Procedures for NPR Publications

The researchers in netCommons will publish all NPR under one of the Creative Commons licenses and they will adopt an Open Access policy also for NPR publications such as technical reports and white papers.

The procedure in this case is simple and similar to the Gold Open Access case:

1. When a technical report is published (e.g., on an institutional website), the authors store a version of the paper, along with the available metadata, in one or more repositories of her/his choice among those supported by OpenAIRE;
2. The paper publication is notified to the project coordinator and to the exploitation and dissemination list (netcommons.wp6@list.disi.unitn.it);
3. Within a few days the manuscript becomes visible automatically through OpenAIRE reporting the proper reference to netCommons;

4. A script embedded in the netcommons web-site and compliant with OpenAire APIs, parses OpenAIRE daily (or weekly) to retrieve novel manuscripts and upload them automatically on the netCommons web site in the proper section.

Exception may apply to these rules and procedure for contributions to newspapers and dissemination magazines.

3.3. Current Policies by some of the Major Scientific Publishers

Clearly, the choice of whether to take a Green or a Gold Open Access policy is also determined by the specific publisher and by the scientific field. Self archiving is today compatible with the most important publishers, as far as it is limited to the *accepted version* of the paper, but publishers as IFIP and Association for Computing Machinery (ACM) go definitely beyond, as described below. With other publishers, the evaluation should be made on a case by case basis. Details on most publishers and journal policies can be found on the Sherpa Romeo portal (<http://www.sherpa.ac.uk/romeo/index.php>). In the extreme case in which self archiving is prohibited and commercial open access options are not available, the authors should avoid the journal.

For the authors' convenience and for general reference, we report here the current policy contained in the copyright agreement or on web-pages of some of the most relevant publishers at the moment of writing, though it is strongly recommended to check the single journal OA policy on the Sherpa Romeo database and/or on the journal website. The information in the following sub-sections is mostly taken verbatim from publishers web pages, thus may contain advertisement-like information and in general the publisher visions, which are not necessarily reflected or agreed-upon by netCommons consortium.

3.3.1. Elsevier

The Elsevier policy on authors right can be found in the website <http://www.elsevier.com/about/company-information/policies/sharing>. Elsevier supports Green Open Access, but maintains a number of journals (<http://www.elsevier.com/embargoperiodlist>) with an embargo policy. Though these journals can be used for netCommons publications, we suggest to avoid those that have an embargo period longer than 12 months. In any case also journals subject to embargo allows pre-prints to be shared in private repositories. Citing from Elsevier's Frequently Asked Questions (FAQs) page:

Q. Have you removed an author's right to self-archive in their institutional repository?

A. No. We have removed the need for an institution to have an agreement with us before any systematic posting can take place in its institutional repository. Authors may share accepted manuscripts immediately on their personal websites and blogs, and they can all immediately self-archive in their institutional repository too. We have added a new permission for repositories to use these accepted manuscripts immediately for internal use and to support private sharing, and after an embargo period passes then manuscripts can be shared publicly as well.

Regarding the author rights on the *accepted versions* of the manuscripts of journals not subject to embargo, we find the following wording:

Authors can share their accepted manuscript:

Immediately

- via their non-commercial personal homepage or blog by updating a preprint in arXiv or RePEc with the accepted manuscript
- via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research

collaboration work-group

[omissis]

After the embargo period

- via non-commercial hosting platforms such as their institutional repository
- via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license

[omissis]

The CC-BY-NC-ND license can easily be obtained through the website <http://creativecommons.org/licenses/> and is explicitly recommended by the EC to *enable open access in its broadest sense*.

3.3.2. ACM

The ACM policy can be found in the website https://www.acm.org/publications/policies/copyright_policy. ACM today adopts a very flexible scheme that ACM itself summarizes as follows:

“Authors have the option to choose the level of rights management they prefer. ACM offers three different options for authors to manage the publication rights to their work.

- *Authors who want ACM to manage the rights and permissions associated with their work, which includes defending against improper use by third parties, can use ACM’s traditional copyright transfer agreement.*
- *Authors who prefer to retain copyright of their work can sign an exclusive licensing agreement, which gives ACM the right but not the obligation to defend the work against improper use by third parties.*
- *Authors who wish to retain all rights to their work can choose ACM’s author-pays option, which allows for perpetual Open Access through the ACM Digital Library. Authors choosing the author-pays option can give ACM non-exclusive permission to publish, sign ACM’s exclusive licensing agreement or sign ACM’s traditional copyright transfer agreement. Those choosing to grant ACM a non-exclusive permission to publish may also choose to display a Creative Commons License on their works.”*

We notice that also in case of the traditional copyright transfer all ACM publications allow Green Open Access without any embargo. Generally, the publisher’s version/PDF cannot be used, but the author’s refereed post-print can be uploaded for non commercial use on author’s personal website, institutional repository, open access repository, the employer’s website or the funder’s mandated repository. Publisher copyright and source must always be acknowledged, and there must be a link to the publisher version with a statement that this is the definitive version and Digital Object Identifier (DOI). A set statement must be added on the website/in the repository:

“©ACM, YYYY. This is the author’s version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in PUBLICATION, {VOL#, ISS#, (DATE)} <http://doi.acm.org/10.1145/nnnnnnn.nnnnnn>”

Statement reported on the Sherpa Romeo web site (<http://www.sherpa.ac.uk/romeo/pub/21/> as of May 30th 2018).

3.3.3. IEEE

The Institute of Electrical and Electronics Engineers (IEEE) specifies its policy in two companion documents that can be found in the association website [8, 9]. In summary:

Generally, authors have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice (see below) and, when published, a full citation to the original IEEE publication, including a Digital Object Identifier (DOI) and a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their articles.

The following copyright notice must be displayed on the initial screen displaying IEEE-copyrighted material:

“©20xx IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the IEEE work with a Digital Object Identifier (DOI), or (2) the accepted version only with the DOI (not the IEEE-published version).

IEEE also has an open access program for *Gold Access Policy*, which at the moment is limited to the societies journals. In any case IEEE always allows its authors to follow mandates of funding agencies and post the accepted version into publicly available repositories limiting the embargo to what admitted by the funding agency².

3.3.4. Springer

Generally, authors can archive post-print (i.e., final draft post-refereeing) on author’s personal website immediately and on any open access repository after 12 months after publication. Publisher’s version/PDF cannot be used; published source must be acknowledged and there must be a link to the publisher version, with a set phrase to accompany link to published version. Articles in some journals can be made Open Access on payment of additional charge. (see: <http://www.sherpa.ac.uk/romeo/pub/74/> as seen on May 30th 2018).

As far as Springer LNCS is concerned (see <http://www.sherpa.ac.uk/romeo/pub/2765/> as of May 30th 2018), authors can archive post-print (i.e., final draft post-refereeing) on author’s personal website, institutional repository or funder’s designated repository. Publisher’s version/PDF cannot be used; published source must be acknowledged and there must be a link to the publisher version with DOI and a set phrase to accompany link to published version.

If Springer Open is chosen (see <http://www.sherpa.ac.uk/romeo/pub/948/> as of May 30th 2018), authors can archive post-print (i.e., final draft post-refereeing) and publisher’s version/PDF. The published source must be acknowledged; authors retain copyright and a Creative Commons Attribution License must be attributed.

²Albeit less permissive than ACM, IEEE seeks large diffusion of its own publications and is open to re-publish on Xplore also papers that are published elsewhere (e.g., IFIP Digital Library) in Gold Open Access. IEEE rights management procedures can be found at http://www.ieee.org/publications_standards/publications/rights/rights_policies.html and in the PDF document <https://pspb.ieee.org/images/files/files/opsmanual.pdf>. In general IEEE collects copyrights through the on-line form at <https://www.ieee.org/publications/rights/copyright-main.html>. All links have been visited on June 20 2018.

3.3.5. IFIP

IFIP has in general a very liberal publication policy and is the only major publisher we considered that has an effective Gold Open Access policy without a pay-to-publish model, for some of its publications. In general, publications (mainly proceedings of conferences and workshops) that are available through the IFIP Digital Library (<http://dl.ifip.org>) is available in Gold Open Access, free-to-read basis. For other publications different policies apply, depending on the publisher collaborating with IFIP and also the policy of other scientific societies co-sponsoring the publication together with IFIP. The full description of these policies is available at http://www.ifip.org/index.php?option=com_content&task=view&id=219&Itemid=564 (visited June 20th 2018). Interestingly, as can be found on the web site, IFIP is also making its old publications available in Gold Open Access, in particular:

“Open Access to IFIP Publications is currently realized via three channels:

1. *The upcoming IFIP Digital Library (<https://hal.inria.fr/IFIP/>), which currently has the author’s files of IFIP Springer publications (AICT, LNCS, LNBIP) from 2014, 2013, 2012 and more and more earlier proceedings. It also has the references to such publications up to and including 2016.*
2. *Springer Link (<https://link.springer.com>) for IFIP Springer publications (AICT, LNCS, LNBIP) with free access for all articles 4 years old and older, almost complete back till 2005 and on top many publications back till 1995, when Kluwer became the official IFIP publisher. NB: Not the “books” are free, but the individual articles. So don’t get confused by the fact that there is a lock on the book download. Just move on to the table of contents and download the articles individually.*
3. *The current IFIP Digital Library (dl.ifip.org/) for some non-Springer publications and also Springer publications between 2005 and 2012.*

Combined, (almost) all IFIP works published with Springer from ca. 2005 to 3 years old are available for open and free access (one way or the other). In addition many proceedings back till 1995 when Kluwer became the official IFIP publisher are available (NB.: Kluwer, the then official IFIP publisher, was merged into Springer back in 2004).”

Some IFIP journals published by Springer and Elsevier have a paid Open Access option , such as:

Journal: Computers and Security (ISSN: 0167-4048)

Journal: International Journal of Critical Infrastructure Protection (ISSN: 1874-5482)

Journal: Entertainment Computing (ISSN: 1875-9521, ESSN: 1875-953X)

- Authors can archive post-print (i.e., final draft post-refereeing) on author’s personal website immediately and on open access repository after an embargo period of between 12 months and 48 months; it must link to publisher version with DOI and must be released with a Creative Commons Attribution Non-Commercial No Derivatives License
- Authors cannot archive publisher’s version/PDF;
- Permitted deposit due to Funding Body, Institutional and Governmental policy or mandate, may be required to comply with embargo periods of 12 months to 48 months.

Journal: Education and Information Technologies (ISSN: 1360-2357, ESSN: 1573-7608)

- Authors can archive post-print (i.e., final draft post-refereeing) on author’s personal website immediately and on any open access repository after 12 months after publication. It must link to publisher version; the published source must be acknowledged with a set phrase to accompany link to published version;
- Authors cannot archive publisher’s version/PDF.

3.3.6. SAGE

Journals published by “SAGE-Hindawi Access to Research” have a paid Open Access option. Authors retain the copyright of their article, which is freely distributed under the Creative Commons Attribution License, permitting the unrestricted use, distribution, and reproduction of the article in any medium, provided the original work is properly cited. In order to cover the costs of publication, Article Processing Charges are required for accepted manuscripts. (<http://www.hindawi.com/memberships/> as of June 10th 2018).

In subscription journals published by “SAGE Publications (UK and US)”, authors can deposit the version of the article accepted for publication (version 2) in their own institution’s repository. Authors may not post the accepted version (version 2) of the article in any repository other than those listed above (i.e., you may not deposit in the repository of another institution or a subject repository) until 12 months after first publication of the article in the journal. Authors may not post the published article (version 3) on any website or in any repository without permission from SAGE. When posting or re-using the article authors must provide a link to the appropriate DOI for the published version of the article on SAGE Journals (<http://online.sagepub.com>). (see <https://uk.sagepub.com/en-gb/eur/the-green-route-%E2%80%93-open-access-archiving-policy> as of June 10th 2018).

In Sage Pure Gold Open Access Journals, all articles provide worldwide, barrier-free access to the full-text of articles online, immediately on publication under a creative commons licence. All articles are rigorously peer-reviewed retaining the quality hallmarks of the academic publishing process that authors would experience in publishing in any traditional SAGE journal. Most SAGE pure Gold Open Access journals are supported by the payment of an article processing charge (APC) by the author, institution or research funder of the accepted manuscript.

Some journals (8 titles: <http://www.sherpa.ac.uk/romeo/journals.php?id=1581&fIDnum=|&mode=simple&letter=ALL&la=en>) published by SAGE Publications (UK and US) with the 12 month Embargo option let authors post on any non-commercial repository or website the version of their article that was accepted for publication – ‘version 2’. The article may not be made available earlier than 12 months after publication in the Journal issue and may not incorporate the changes made by SAGE after acceptance. When posting or re-using the article, authors should provide a link/URL from the article posted to the SAGE Journals Online site where the article is published: <http://online.sagepub.com>, and make the following acknowledgment:

“The final, definitive version of this paper has been published in ;journal;, Vol/Issue, Month/Year by SAGE Publications Ltd, All rights reserved. ©[The Author(s)]. Authors may not post the final version of the article as published by SAGE or the SAGE-created PDF – ‘version 3’.”

See https://mc.manuscriptcentral.com/societyimages/wes/WES_ExclusiveLicense.pdf as of June 10th 2018.

4. netCommons Open Data Policy

4.1. Licenses Analysis and Platform Selection

An interesting novelty of H2020 is the platform known as Open Research Data Pilot (ORDP) [4] for the dissemination of the data that could be used by different researchers to replicate the experiments or the analysis presented in the scientific publications. Given its scope netCommons obviously participates in this pilot, and committed to publish all the data-sets produced in the project with some open data license. Most of the data is being published in the third year of the project, in this deliverable we document the process we started with a data-set used in one of our technical publications. This Chapter serves as a reference to the project partners, in which we document the process, outline the key elements and some problems we encountered. We discussed the concept of data intellectual property and license of use in Sec. 2.1, and we detail here some of those general principles and considerations in light of the research process and the possible licenses publicly available, assuming that in general researchers are not in the position to produce their own license, even if this is technically and legally possible. This part can probably be useful to other projects too as well as to the EC to refine the policies on Open Data, and it is definitely within the scope of research of netCommons as it touches on the concepts of privacy, sharing, as well as freedom and rights of communication.

Among the various platforms available for publishing open data, we selected Zenodo¹, for simple reasons. First, it is used and supported by the EU, and it has a direct connection to the project database of H2020. This makes it easier to connect the published data-set to the project and ensures the availability of data for a long time after the project ends. Second, Zenodo produces DOIs that can be used to reference to the data-set and to connect already published papers to the data-set itself. Last, its interface is easy to use, and supports a variety of licenses for the data-set. Yet the procedure is not bug-free as we document in Sec. 4.2, and the licenses available are not entirely satisfactory.

Industrial and intellectual (intended as a creative process of the human mind) property are fairly well understood and regulated; data production, property and protection is instead far less debated and understood. Based in the UK, the Open Knowledge International (OKI)² is a global no-profit that offers some interesting information and some resources from the management of open data. They claim to be the first organization that started the analysis and definition of open data, issuing “The Open Definition”³ in 2005; however, it is clear even from a simple reading of the “Open Definition 2.1”⁴, currently the last definition available, that all their work and concepts are heavily derived from the free software movement, which clearly falls into the intellectual property as defined above, and treat data as a derivative of a creative process, which may not be necessarily the case.

This said, their “Open Data Handbook”⁵, is a valuable tool to start understanding the concepts around open data and the legal ambiguities existing. At the same time the licensing philosophy they propose is questionable from the very beginning. For instance the Universal Participation principle⁶, as an obligation, is highly restrictive and also poses ethical issues depending on the origin and semantics of the data itself. Tapping again at the genome example introduced in Sec. 2.1: If I openly and freely publish my personal genome data, is it ethical to

¹<https://zenodo.org>

²<https://okfn.org/>

³<https://opendefinition.org/>

⁴<https://opendefinition.org/od/2.1/en/>

⁵<http://opendatahandbook.org/>

⁶Quoting from the Open Data Handbook identified above “*Universal Participation: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.*”

build a commercial use of it? It is debatable, but clearly in face of such an obligation I may simply decide not to publish free and open, obtaining exactly the opposite result as the one intended.

Another traditional source of best practices and policies for open data and data sharing is Conseil Européen pour la Recherche Nucléaire (CERN) that maintains a portal to share under specific conditions that data obtained in experiments⁷. Clearly they are focused on data coming from physics experiments, and their vision on Open Data management is definitely different from the OKI one, more concentrated on science and education. Not all data is open and use is often restricted, with a clear hype on data citation⁸ to guarantee that CERN gets the deserved visibility.

We think that this discussion is sufficient to understand the issues and problems that still surround Open Data publication. A complete and detailed discussion, both in legal and technical terms is beyond the scope of this document; however, before proceeding with the details on how netCommons handle data publication, a short discussion on specific licenses (or the lack of them) is useful.

For instance, most of Creative Commons licenses⁹ may not apply to data as “derivative work” on Data is not clearly defined and manipulating a data set with purposes different from rendering may be inappropriate; sometimes even rendering and statistical analysis may change the actual meaning of the Data published. Remaining in the Creative Commons licensing framework, the ShareAlike¹⁰ seems acceptable for sharing data as it limits the possibility of manipulating it beyond the intention of the data collector. However, it is clear that this license focus is on “work” and not “data”, thus its interpretation remains ambiguous. From the perspective of reducing the data manipulation possibility, meaning that data semantics should not be changed to bend it to a different goal or result, the NoDerivative¹¹ option may be more appropriate. Once more the focus is on “work” not “data”, thus a restrictive interpretation may mean that the data cannot be used at all, because simply plotting it, or plotting with a different graphical layout rather than suggested by the author/owner, is a modification of the work.

Similarly, licenses like Open Database License version 1.0 (ODbLv1.0)¹² may not apply in many cases for both technical inconsistency (e.g., the wording “intermixing with other datasets” is a technically inconsistent definition) and it contains also semantic ambiguities. Furthermore, the researchers of netCommons agree that it may not be acceptable that all produced Data should be released also for commercial purposes. We finally notice that ODbLv1.0 separates the Format of the Data from the Content (which is an ambiguous definition once more), and covers only the Format, while the Content must be covered by a separate license, introducing further ambiguities, because the same “object” (as defined by a DOI) is covered by more than one license. This license seems to be intended to cover two specific aspects of data collection, but not the data itself. In particular the aspects that seems of most concern for this license are:

1. Structure of a database, meaning its technical implementation;
2. The work of collecting items of the database, i.e., the individual (or team) effort of gathering, massing up, and organizing the content of the database, as evident from Sections 4.3 and 4.4 of the license.

Once more a complete analysis of the open data licensing issues goes beyond the scope of this document, we simply signal that there are other licenses and many additional resources to be considered when publishing Open Data. In particular, the “Public Open Data: access to re-usable public sector information”¹³ refers to

⁷<http://opendata.cern.ch>

⁸<https://www.force11.org/datacitationprinciples>

⁹<https://creativecommons.org/share-your-work/>

¹⁰Quoting Creative Commons web site ShareAlike means “*You let others copy, distribute, display, perform, and modify your work, as long as they distribute any modified work on the same terms. If they want to distribute modified works under other terms, they must get your permission first.*”

¹¹Quoting again Creative Commons web site NoDerivatives means “*You let others copy, distribute, display and perform only original copies of your work. If they want to modify your work, they must get your permission first.*”

¹²<http://opendatacommons.org/licenses/odbl/>

¹³<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2017/03/13/Public+Open+Data%3A+access+to+re-usable+public+sector+information>

the process and obligations of Public Administrations in the EU in managing and publishing data pertaining to endeavors achieved with public funding, whatever it is. Indeed, the regulations at the root of the ORDP are part of this process.

Some national licensing policy for public administrations, like the Italian one¹⁴, derive from the obligations of the EU directives, and refer specifically to data produced by public administrations. Interestingly, this specific case (it is in Italian and binds Italian administrations only), is an example of license that correctly comprises the structure of the data, the data itself and the information (semantics) contained in the data, but it seems that there are no such examples of general purpose licenses with international scoping.

Finally, the EU provides a portal¹⁵ where public data can be collected, making in practice a collector hub where many (some?) Open Data can be found; however once again this is focused on data produced by public administrations and does not shed additional light in the license selection problem.

Given this situation, netCommons decided that it is not possible to identify a license that apply to every case, so each dataset will be published selecting carefully, on a case-by-case basis, the most appropriate license and the most appropriate level of aggregation and detail, as well as the most appropriate set of repositories where the Data produced during the research can be archived and made public, provided that the dataset is published at least on Zenodo.

4.2. Using the Zenodo Interface

Zenodo allows uploading data only to registered and identified users. Log-in can be done using two external sources (Github and ORCID) as Fig. 4.1 shows. Anyone in possession of a valid email can create an internal Zenodo account, if unwilling to use the other two platforms as single-sign-on.

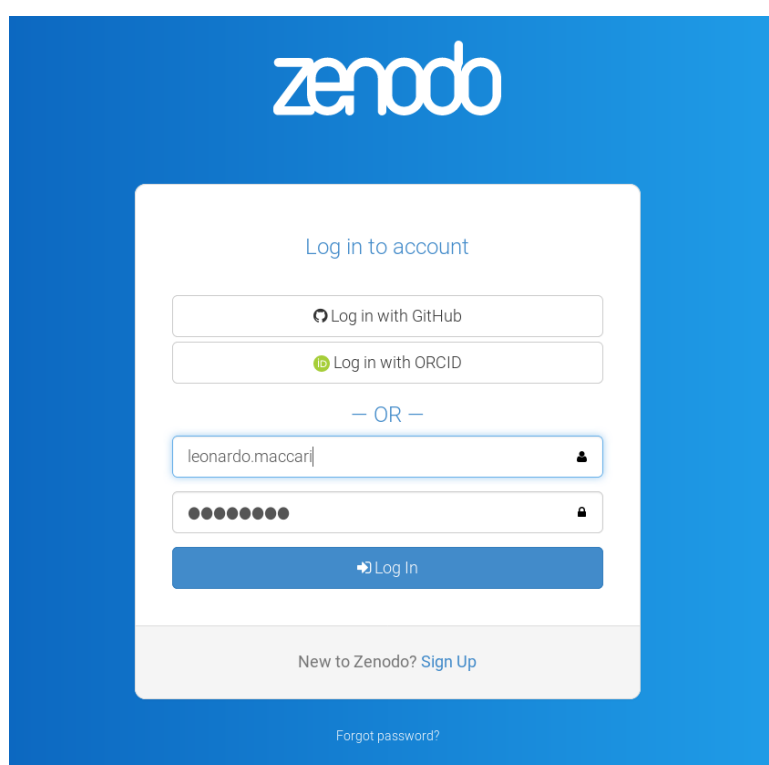


Figure 4.1: Zenodo log-in interface.

¹⁴<https://www.dati.gov.it/content/italian-open-data-license-v20>

¹⁵<https://www.europeandataportal.eu/>

Once logged-in, as depicted in Fig. 4.2 one can view recent information on the Zenodo platform and decide to upload new material with the top “upload” button.

The screenshot shows the Zenodo website interface. At the top, there is a navigation bar with the Zenodo logo, a search bar, an 'Upload' button, and a 'Communities' dropdown menu. The user's profile 'leonardo.maccari@unitn.it' is visible in the top right corner. Below the navigation bar, the 'Recent uploads' section displays three items:

- Reproducibility Package for "Reproducible research and GIScience: an evaluation using AGILE conference papers"** (April 23, 2018 v2). This item is marked as 'Open Access' and has a green checkmark. A 'View' button is present.
- Airgun-ocean bottom seismometer survey dataset** (April 24, 2018 v1). This item is marked as 'Dataset' and 'Open Access'. A 'View' button is present.
- microPIECE (microRNA pipeline enhanced by CLIP experiments)** (April 16, 2018 (v1.5.2)). This item is marked as 'Software' and 'Open Access'. A 'View' button is present.

Each item includes the author's name, a brief description, and the upload date. On the right side of the interface, there are three informational panels:

- Zenodo now supports DOI versioning!** with a link to read more.
- Using GitHub?** with a link to check out GitHub integration.
- Zenodo in a nutshell** with a list of features:
 - Research. Shared.** – all research outputs from across all fields of research are welcome!
 - Citeable. Discoverable.** – uploads get a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
 - Communities** – create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads.
 - Funding** – identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
 - Flexible licensing** – because not everything is under Creative Commons.
 - Safe** – your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

Figure 4.2: Zenodo interface for uploading new material.

Clicking on the upload button, the interface reported in Fig. 4.3 shows up, which displays the list of the uploads initiated so far. In Fig. 4.2 four initiated uploads are shown; only the first one (topmost) is completed as indicated by the small green tick on the left-hand side. The others are not completed, and can not be removed due to a bug in the Zenodo interface documented later on.

Clicking on “new upload” (top right) an upload form is presented as in Fig. 4.4, we describe briefly the main features to be considered for the upload. As a first point, one has to choose the “dataset” option among the possible ones in the “upload type” menu, which will change the upload form below it. In the “Files” tab one can upload the dataset, and in the “Communities” one has to choose the correct community the project falls in, in our case the “European Commission Funded Research” community.

Below these tabs, a comment tab is included, reported in Fig. 4.5. Most of the fields are self-explanatory, but two of them are worth a comment. The DOI field can be used to specify a DOI if already existent. The dataset in fact can be already published in some other platform, and may be re-used here. Furthermore, if one wants to reserve a DOI in the moment of starting the process, one can do it without it being actually published, until the process is completed. This is also useful to include the DOI in a paper that is under review, so that when the paper is published also the data-set can be published and the paper includes the correct reference. The description of the upload is a sensitive field, in which one should document the data published. In the current version of Zenodo (accessed April 2018) this field has a bug, as the text that one inserts in the field is wrongly interpreted by the backend. In our experience adding a BibTex citation (some text included in nested brackets) in this form produces an error in the parser, so that when the page is refreshed it can not be displayed. This

Figure 4.3: List of uploads the user started in Zenodo.

prevented us from editing an upload we already drafted, and in general prevented any interaction with it. This is the reason why in Fig. 4.3 there is a list of unfinished uploads, simply because if we click on them a blank page is shown, and we can not modify or delete them. We notified the administrators of this problem, but in the meantime we strongly suggest to use only the most simple notation for the content of this form.

Note also that removal or modification are not allowed after an upload has been published. If one wants to update some files, he/she needs to create a new version which corresponds to a new DOI.

The form in Fig. 4.6 reports the rest of the mandatory fields, among which the license with which the data will be released, and most important, the association with the netCommons project.

The following discussion on the license refers to the process of depositing the dataset “Topologies collected from 3 Community Networks”¹⁶, and is intended as a guideline and not as mandatory for other datasets. We choose a CC-BY license, which is the most permissive CC available. This choice was led by the fact that the published data-set is interesting for scientific use only (as such, it requires re-elaboration and does not need protection for commercial re-use), and it comes from the elaboration of network topologies extracted from the networks managed by various communities, as such, we did not consider correct to restrict the use of data that communities generally consider as public domain.

In the “Grant” field one has to choose the EU funded projects, and then type “netCommons” in the project name. The form supports auto-completion, so the correct grant number will appear.

In the rest of the form there are optional fields that one can add. Among them, the relevant ones are:

- A list of alternative or related identifiers for the data-set, which may already be public. These identifiers are DOIs of alternative version of the data-set, and/or DOIs of scientific publications that use the data. If these publications have a DOI already available this is the place where to specify the identifiers.
- The rest of the fields (but the last) are alternative version of the identifiers, in case the DOIs are not already available, or one wants to explicitly specify the references of the related works.
- The last field is a set of keywords one can add to characterize the data-set.

¹⁶<https://doi.org/10.5281/zenodo.1218746>

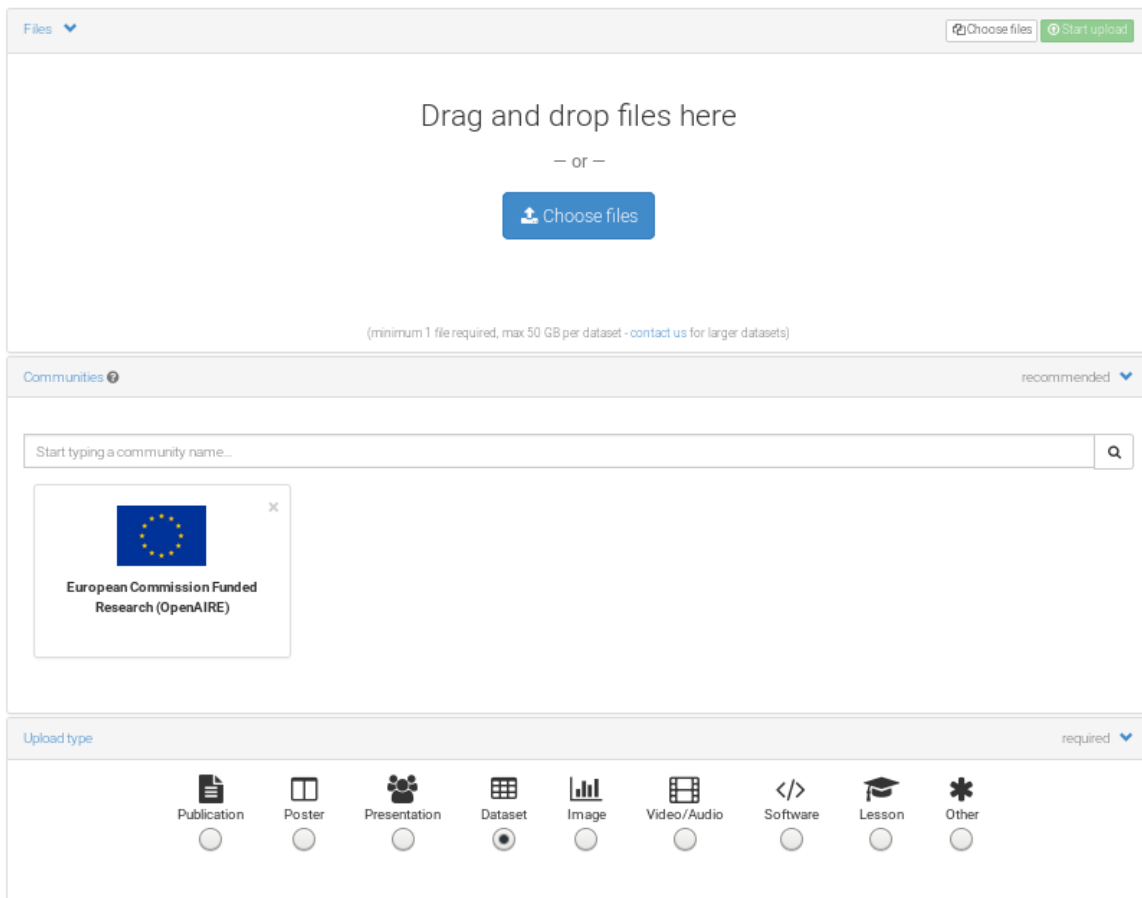


Figure 4.4: Zenodo interface for files upload.

Basic information
required ▼

Digital Object Identifier

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.

Publication date *

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

Title *

Required.

Authors *

⇅ ×

Optional.

+ Add another author

Description *

🔗 📄
B I S x_c x^c
☰ ☲
⋮ ⋮ ⋮ ⋮ ⋮
⌕ 🔍
↩ ↪
Σ Ω
📄 Source
🔗

Required.

Version

Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended. See [semver.org](#) for more information on semantic versioning.

Language

Optional. Primary language of the record. Start by typing the language's common name in English, or its ISO 639 code (two or three-letter code). See [ISO 639 language codes list](#) for more information.

Keywords ⇅ ×

+ Add another keyword

Additional notes

Optional.

Figure 4.5: Zenodo form for data fields.

Licenserequired ▾

Access right *

Open Access
 Embargoed Access
 Restricted Access
 Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

License *

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the 'Other' licenses available ('Other (Open)', 'Other (Attribution)', etc.). The supported open licenses in the list are harvested from opendefinition.org. If you think that an open license is missing from the list, please [contact us](#).

Fundingrecommended ▾

Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](#). Specify grants which have funded your research, and we will let your funding agency know!

Grants

Optional. OpenAIRE-supported projects only. For other funding agencies, please use the 'Other' option. Note: a human Zenodo curator will need to validate your uploads.

[+ Add another grant](#)

- netCommons 088798
- network infrastructure as commons

Related/alternate identifiersrecommended ▶

Contributorsoptional ▶

Referencesoptional ▶

Journaloptional ▶

Conferenceoptional ▶

Book/Report/Chapteroptional ▶

Thesisoptional ▶

Subjectsoptional ▶

Figure 4.6: Zenodo interface for optional fields.

5. Data Security and Privacy Provisions in netCommons

5.1. Scope of the privacy/security model

The activities of netCommons involve only marginally the direct interaction with people and do not require to collect directly any personal or sensitive information. People involved in Community Networks follow the usual policy of these associations, and netCommons in general does not require CNs to transfer any personal or sensitive data. Nonetheless, UoW will run some surveys and some interviews or video recording may be useful during the project.

In such cases the netCommons practice will always abide to the principle of informed consent and to the ethical annex reported as Sect. 5 of the project Document of Work (Annex 1 to the Grant Agreement), furthermore the actions and operations of the researchers will always comply with the national legislation and with the internal regulations of the partners involved in the project. In particular, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) will be strictly followed by the project security model. In the Spanish case, the Data Protection Act number 15/1999 (December 13th) adapted to the European directives by the Royal Decree 1720/2007 promulgated on 21st December 2007 will be applicable to the project. The specific goal of this document is to present and discuss the issues related to the treatment of the collected data in electronic form, their storage on different media (Hard Disks, Storage Units, CD/DVD/SD/USB peripherals), and their distribution using network connections.

5.2. General Principles

The privacy protection and operational model of netCommons rests on three pillars:

- Data anonymity,
- Informed consent;
- Circulation of the information limited to the minimum required for processing and preparing the anonymous open data sets.

Data anonymity will be guaranteed whenever possible. The only exemption to anonymity can be in some cases for the researcher directly interacting with the participants in surveys. When data must be presented in non-aggregate ways for research purposes, the data will be anonymized following the best practices of non-invertible hashing functions applied to all personal information. Furthermore, provisions will be taken to avoid the possibility of information linkage.

The *informed consent* policy requires that each participant will provide his/her informed consent prior to the start of any activity involving him/her. Appendix A reports a template of the informed consent form that will be completed by participants in surveys and interviews. Public distribution of elements of information that can reveal the identity of the users (e.g., videos or pictures) for scientific dissemination purposes will be explicitly authorized by the participant as part of this process.

To achieve a *limited circulation* of the information, the database containing in anonymous form the data collected from the users (e.g., the results of questionnaires and of laboratory experiments) will be distributed to

the partners, if needed at all, through protected and encrypted Internet connections; the raw data will only be shared if it is required for the development. The researchers will never pass on or publish the data without first protecting participants' identities. No irrelevant information will be collected; at all times, the gathering of private information will follow the principle of proportionality by which only the information strictly required to achieve the project objectives will be collected. In all cases, the right of data cancellation will allow all users to request the removal of their data from the project repository at any time.

The final datasets fully anonymized will be published as Open Data as described in Chapter 4.

5.3. Security Framework

In order to accomplish the creation of a security framework it is essential to focus on the issues of access and identity authentication, authorization and auditing (AAA). Therefore, our main objective is to develop a base security system that standardizes the processes of Authentication, Authorisation and Auditing of the various information sources involved.

5.3.1. Authentication

The Data Protection Act requires that any operator who is granted access to sensitive data must be authenticated. Authentication technology should be strict when dealing with sensitive and confidential data available to the users of the platform. To do this, a username and a password will be used so that the person who wants to access the raw data of surveys and interviews confirms that he has authorized access to the system. If deemed necessary by sensitive collected data, which is not foresees now in netCommons, we will use an RSA encryption mechanism, with each operator receiving a personal private key.

5.3.2. Authorization

The objective of the authorization is to determine the rights of a user of an information system. For each researcher, we will specify which content can be accessed based on functionality, security and confidentiality criteria.

5.3.3. Accounting and Auditing

netCommons should not deal with sensitive data, in any case logging of the personal data will be enforced to prevent abuses, and in case of necessity proper auditing measures as provided by the Data Protection Act shall be put in action.

5.4. Summary of Technological solutions

We report below a table of the main technological solutions used for the different security issues mentioned in Sec. 5.2.

GOAL	Technological Solution
Guaranteeing complete anonymity where required	The collected data will be labeled with participant codes. Participant consent forms will be held separately and will not reference the participant code. These will be paper based and held in a locked filing cabinet on the researchers site
Safe keeping of the documentation on informed consent	The informed consent will be provided by the interested subject by filling an appropriate form as reported in Appendix A. The authorized personnel must keep this physical document under lock and key. Information on the interested person can also be stored in electronic form in an database or in a spreadsheet. The spreadsheet or the database will be encrypted and its access will be password-protected and granted only to authorized operators
Remote access	In the general case, the “raw” data related to the participant to the project, will be handled only by the researchers interacting with the participant and made available to the rest of the consortium only in anonymous form. In particular any personal data contained in the collected data will be handled only by the researchers interacting with the participant. If, for special cases, some other researchers should need to access to the “raw” data, the interested participants will be informed. Only after their consent is extended to the requiring researcher, can he/she have access to the data. In this case, if the access is remote, the system has to have the following Researchers in the consortium can have access through an Internet SSL connection

6. Conclusions

The topics of Open Access and Open Research Data is one of the key debates open in the scientific world, specially in case of research project that are funded with public money. netCommons is not only a research project funded by the EC, but it is also a project that deals with societal challenges, socio-economic sustainability, the construction of a commons, and techno-legal provisions. As such its effort to disseminate and propagate results and findings must be, and it is indeed maximal.

This deliverable described the policy that netCommons has finally identified and approved for its own best practices in scientific Open Access dissemination and in data collection and management to achieve Open Research Data.

Regarding Open Access to publications, on the one hand, we have clearly identified that most leading scientific publishers provide appropriate licenses and means to achieve Open Access, either through Green Open Access (i.e., re-publication on OAI-PMH compliant repositories) or through Golden Open Access. On the other hand Golden Open Access is still very often ambiguous on the peer-review process, and the publication fees required to authors are difficult to justify by the cost of electronic publishing, specially in light of modern editing technologies that hardly require any intervention of the publisher on the material provided by the authors. In conclusion, netCommons does not see any obstacle to achieve a complete Open Access dissemination for all its scientific publications, and has actually so far complied to this policy with all the publications and deliverables too.

Regarding Open Research Data accessibility and licensing, we have found that the situation is far less clear, and that most Institutions are still most unaware of the problem and they do not provide appropriate repositories, but also that the repositories available, like Zenodo and others, are sometimes at odds with the complexity of data, as contrasted to the simple publishing of a PDF file that is normally done for Open Access to literature. At the same time, also the concept of license and of derivative for Open Data is not as mature as it is for publications, where the concept of copyright and the notion of intellectual property as well as creative work are well understood both at the technical and the legal level. Indeed, in many cases Data cannot be classified as a creative work, and the intellectual property of Data does not yet have a commonly accepted technical and legal definition. Furthermore the publication of data must comply with legal provisions on privacy and individual protection.

netCommons deems in any case that data collected and used for scientific research (specially if receiving public funding), must be made available to the scientific community for validation and falsification of results and theories and to the public community at large for transparency and control. We identifies Zenodo as the primary repository where to publish our data, without excluding the possibility of publishing also on institutional repositories as well as other platforms, including github or similar. The selection of the license to define the correct scope of use of the data proved particularly challenging. Indeed, the lack of proper understanding, from a legal point of view, of what is data, and the lack of licenses that satisfactorily address all the facets of data publication resulted in the impossibility of setting a common license suitable for all the data produced by the project. For instance, the classical open licenses, like Creative Commons, clearly refer to creative works of the human mind, and not to data, leaving many aspects of data usage out of the licensing scope. On the other hand, specific licenses like the Open Database License version 1.0, refers only to the structure and collection work of the database and not to the data itself, called there Content, specifically requiring an additional license to set the distribution scope of the data. In some cases, e.g., a database of images, this double licensing scheme can be quite clear and easy to implement, but in others, like a dataset of answers from a survey, just to take an example that interests netCommons, this is far less clear and indeed probably impossible to apply, as single items of the

dataset are normally semantically void (e.g., the answer to a YES/NO question), and taken in isolation are not subject to any licensing scheme.

Thus finally, netCommons decided to consider which license to use to publish Open Research Data on a case-by-case basis to guarantee that any item has the most appropriate licensing scheme. In any case all Data is published at least on Zenodo, guaranteeing that it is correctly indexed by the OpenAIRE platform.

Bibliography

- [1] C. Fuchs, F. Giovanella, R. Guidolin, R. Lo Cigno, and L. Maccari, “Data Management Plan (v1),” netCommons Deliverable D7.1, Jul. 2016. <http://netcommons.eu/?q=content/data-management-plan-v1-0>
- [2] The European Commission, “Participants Portal on-line Manual: Open Access & Data Management,” http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm, Accessed in June 2016.
- [3] European Commission – Directorate-General for Research & Innovation, “Guidelines on Data Management in Horizon 2020 – Version 2.1,” http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, Feb. 15, 2016.
- [4] OpenAIRE Consortium, “What is the Open Research Data Pilot?” <https://www.openaire.eu/opendatapilot>, Accessed in June 2016.
- [5] D. Lazer and et al., “Computational Social Science,” *Science*, vol. 323, p. 721–723, Feb. 6, 2009.
- [6] The Network of Excellence in Internet Science Consortium, “Project Web Site,” <http://www.internet-science.eu/network-excellence-internet-science>, Accessed in June 2016.
- [7] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, “The Open Archives Initiative Protocol for Metadata Harvesting, Protocol Version 2.0,” <http://www.openarchives.org/OAI/openarchivesprotocol.html>, Jan. 8, 2015.
- [8] IEEE, “IEEE Policies,” https://www.ieee.org/content/dam/ieee-org/ieee/web/org/about/whatis/ieee_policies.pdf, Last visited June 20 2018.
- [9] —, “IEEE Publication Services and Products Board Operations Manual 2018,” <https://www.ieee.org/content/dam/ieee-org/ieee/web/org/pubs/pspb/opsmanual.pdf>, Last visited June 20 2018.

A. Appendix: Template of the Informed Consent Form

Informed Consent Form

This survey/interview is part of the EU Horizon 2020 research project “netCommons: network infrastructure as commons”: <http://www.netcommons.eu>. Scholars from the five EU-based institutions involved in the netCommons project carry out the survey research. The study does not have any commercial purposes, the involved researchers do not have any monetary benefits by conducting the study and the results will be published in the form of a report and research papers based on the survey. Furthermore, the collected data will be published in anonymous form as open data. The open data will not contain any personal identifiers, which is data that we are not interested to collect, do not ask for and do not publish. We will not ask you to provide personally sensitive data in this survey and all the answers provided will be used only in aggregate and anonymous form.

By signing this form, you confirm the following:

- I agree to the digital recording of the interview/survey
- I agree that the answers you give are stored in digital form in a database in such a way that I am not personally identifiable (anonymous or pseudonymous form)
- I have been given the opportunity to ask questions about the project
- I understand that my taking part is voluntary. I can withdraw from the study at any time during the interview/survey and I do not have to give any reasons for why I no longer want to take part.
- I understand my personal details such as my name, email, phone number and address will only be used by the researcher to contact me if necessary and will not be revealed to people outside the project. In any case such information will be completely deleted at the end of the project.
- I understand that my words may be quoted in publications, reports, web pages, and other research outputs in anonymous or pseudonymous form only (no name or other personal identifiable data will be mentioned).

The person responsible for the treatment of the data used in this survey/interview is: Prof.

University of

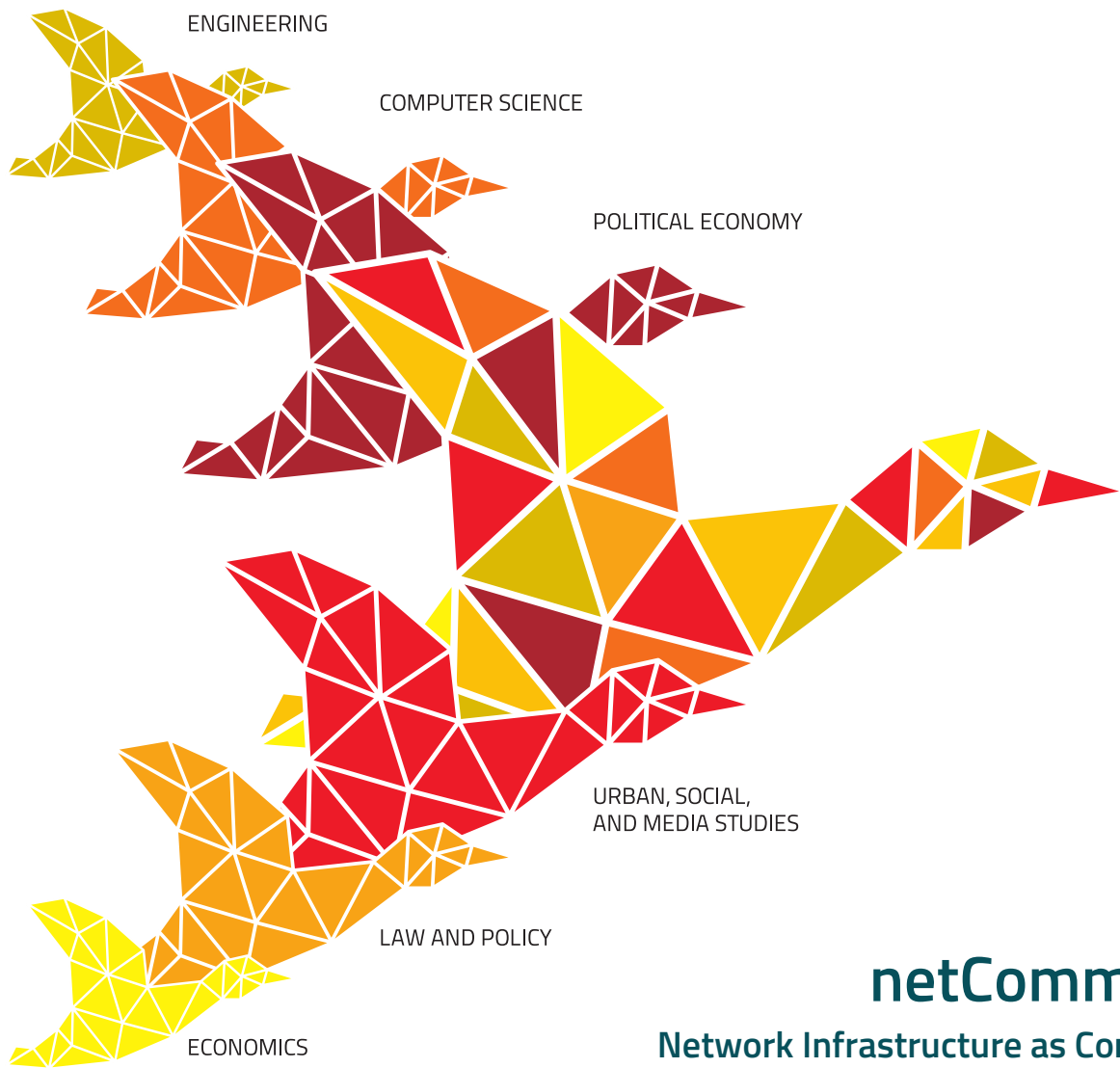
E-mail

Phone

If you you have any questions, don't hesitate to contact him/her. I agree to these terms and want to participate in the interview/survey.

Yes

No



netCommons
Network Infrastructure as Commons

Deliverable 7.3 Data Management Plan (v2)

Deliverable Number D7.3
Version 1.0
June 24, 2018



This work is licensed under a Creative Commons "Attribution-ShareAlike 3.0 Unported" license.

