A Practical Workflow for an Open Scientific Lifecycle Project: EcoNAOS

Annalisa Minelli[1(✉)], Alessandro Sarretta[1], Alessandro Oggioni[2], Caterina Bergami[3], and Alessandra Pugnetti[1]

[1] CNR-ISMAR Venezia, Arsenale - Tesa 104, Castello 2737/F, 30122 Venezia, Italy
annalisa.minelli@ve.ismar.cnr.it
[2] CNR-IREA, Via Bassini, 15, 20133 Milano, Italy
[3] CNR-ISMAR Bologna, Via Gobetti 101, 40129 Bologna, Italy
http://www.ismar.cnr.it/
http://www.irea.cnr.it/

Abstract. This paper represents a review of the practical application, work done and near-future perspectives of an open scientific lifecycle model. The EcoNAOS (Ecological North Adriatic Open Science Observatory System) project is an example of the application of Open Science principles to long term marine research. For long term marine research we intend here all the marine research projects based on Long Term Ecological Data. In the paper, the structure of the lifecycle, modeled over Open Science principles, will be presented. The project develops through some fundamental steps: database correction and harmonization, metadata collection, data exploitation by publication on a web infrastructure and planning of dissemination moments. The project also foresees the setting up of a data citation and versioning model (adapted to dynamic databases) and a final guidelines production, illustrating the whole process in detail. The advancement state of these steps will be reviewed. Results achieved and expected outcomes will be explained with a particular focus on the upcoming work.

Keywords: Open science · Open data · Data citation and versioning

1 Introduction

The main aim of this paper is to describe an application of Open Science principles to marine research.

During the last years, the whole scientific community has expressed an increasing interest in Open Science. This is maybe due to the wider accessibility of scientific research products when shared using Open Access instruments [1–4]. For scientific research products we intend here a various range of outcomes (and inputs) from a research project such as, for example, scientific data, research ideas, metadata, experimental results, etc. The EcoNAOS project [5] (which stays for Ecological North Adriatic Open Science Observatory System) is based

P. Manghi et al. (Eds.): IRCDL 2019, CCIS 988, pp. 209–221, 2019.
https://doi.org/10.1007/978-3-030-11226-4_17

on the testing and application of Open Science principles in marine long term ecological research. The principles we want to put the accent on are: free access to scientific research, research reproducibility, use of all available knowledge at early stage, sharing knowledge as early as possible [6] which could lead to social and economical benefits (ideas producers are also users, enabling new working models, new social relationships, saving money and time [7,8]), other than increase transparency of research and make cooperation between researchers as easy as possible. These principles are known and accepted at international level, and more specifically by the European Union, which defines the guidelines for Open Science application in scientific research and the FAIR principles applied to research data management [9]: data must be Findable, Accessible, Interoperable and Reusable (FAIR). Long Term Ecological Research (LTER) is a branch of ecological research aiming at collecting ecological data, from decadal to centennial scale, to understand environmental changes across the globe detecting trends, preventing and solving environmental and socio-ecological problems through question and problem-driven research. This data is collected and shared, when possible, by international network (ILTER) composed by ecological sites over four continents. Sites are then grouped and managed both at continent scale (eLTER) and national scale (LTER-Italy). LTER-Italy counts 79 research sites including terrestrial, freshwater and marine ecosystems distributed throughout the country, with a marked trans-ecodomain approach.

In this work we applied Open Science principles and the EcoNAOS model to the LTER macrosite "Northern Adriatic Sea", registered into the Dynamic Ecological Information Management System - Site and dataset registry (DEIMS-SDR), created to collect information about geographical macroregions (https://data.lter-europe.net/deims/92fd6fad-99cd-4972-93bd-c491f0be1301).

1.1 EcoNAOS as a Workflow

When dealing with the usual research lifecycle, we think to a linear process involving: research project design, data collection, data analysis, data preservation and curation, presentation of results. More recently, the "shape" of this process has been evolved from linear to circular [10,11] involving a research review phase which allows the initial project to be corrected and to perform all the other phases subsequently.

The EcoNAOS project has been conceived as a spiral-shaped workflow, which represents the Open Research project lifecycle (Fig. 1). This lifeycle is composed by many different phases of sharing research products: inputs (research ideas), data, metadata, results, procedures, code, and reviews in order to establish a flux of knowledge between the project and the scientific world. In particular, an open peer review process allows the reviewers to be identified and the review accounted as research act [12].

The EcoNAOS project was funded by the RITMARE (Italian Research for the Sea) flagship project of the Italian Ministry of Education and Research (http://www.ritmare.it/) which had, among others, the aim to create an homogeneous and coordinate marine observatory system. The North Adriatic Sea

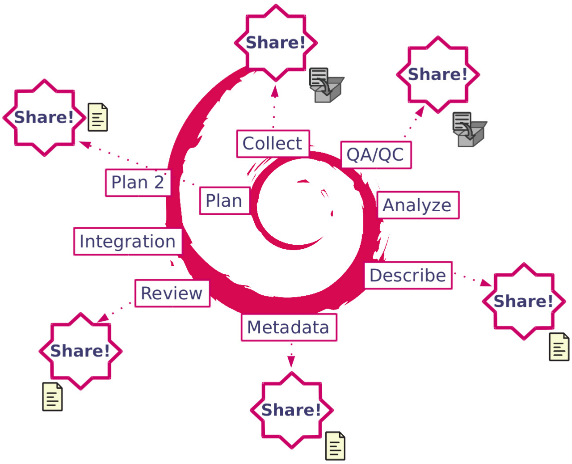

Fig. 1. The Open research project lifecycle.

(NAS) represents the perfect area for a marine observatory because of its transnational characteristics, the numerous interests involved (touristic, commercial, economical) from public and private actors and the presence of marine protected areas. Moreover, data from this zone is diverse in typology (single point observations, time series, samples collected in field or analyzed in laboratory), spatial and temporal coverage. So, the more heterogeneous data is, the more it is complex to manage. In this way, we obtain a wide range of possible situations that could occur in a marine observatory.

CRUISE	STATION	mon/day/yr	hh:mm	Long	Lat	Depth	TEMP	SAL	DENS	pH	Alkal Tot	Oxyg	Ox%
PP/1	B	04/12/1965	9:33	12.68	45.33	0.5	13.122	29.610	22.220	8.422	2.810	6.880	115.82
PP/1	B	04/12/1965	9:33	12.68	45.33	5.0	12.345	35.66	27.040	8.330	2.740	6.390	110.42
PP/1	B	04/12/1965	9:33	12.68	45.33	10.0	12.450	35.43	26.850	8.329	2.710	6.590	113.91
PP/1	B	04/12/1965	9:33	12.68	45.33	20.0	12.136	38.01	28.920	8.333	2.680	6.410	112.15
PP/1	C	04/12/1965	12:20	12.86	45.28	0.5	12.251	35.44	26.890	8.241	2.730	6.290	108.28
PP/1	C	04/12/1965	12:20	12.86	45.28	5.0	12.237	35.46	26.930	8.331	2.690	6.320	108.80
PP/1	C	04/12/1965	12:20	12.86	45.28	10.0	11.155	37.79	28.950	8.312	2.730	6.430	110.03
PP/1	C	04/12/1965	12:20	12.86	45.28	20.0	12.301	37.92	28.810	8.335	2.740	6.490	113.85
PP/2	A	04/28/1965	6:42	12.48	45.40	0.5	12.266	33.0	25.030	8.197	2.919	6.070	102.80
PP/2	A	04/28/1965	6:42	12.48	45.40	1.0	12.370	33.39	25.300	8.316	2.892	6.110	103.94
PP/2	A	04/28/1965	6:42	12.48	45.40	5.0	12.437	35.39	26.980	8.255	2.779	6.120	105.89
PP/2	A	04/28/1965	6:42	12.48	45.40	10.0	12.225	37.3	28.350	8.277	2.753	5.780	99.76
PP/2	B	04/28/1965	9:10	12.68	45.33	0.5	12.485	32.9	24.890	8.275	3.014	6.510	110.64
PP/2	B	04/28/1965	9:10	12.68	45.33	5.0	12.432	33.78	25.600	8.225	2.907	6.490	110.84
PP/2	B	04/28/1965	9:10	12.68	45.33	10.0	11.917	37.21	28.340	8.221	2.734	6.200	107.35
PP/2	B	04/28/1965	9:10	12.68	45.33	20.0	10.502	37.72	29.000	8.217	2.758	6.330	106.75

Fig. 2. Example of observations stored in the database with specification of cruise and station code, date, hour, coordinates, water depth where the observation was collected, and parameters.

1.2 LTER Marine Data in the Northern Adriatic Sea

The database on which we applied the EcoNAOS model is a collection of LTER data in the LTER_EU_IT_012 (NAS) site. Observations are both abiotic measurements and plankton (phyto- and zoo-) counting, collected during oceanographic cruises (data in a specific time range and heterogeneously distributed over a geographical zone) or by sensors (continuous data in a fixed geographical position). The time range of these observations spans from 1965 to 2015 and the first ones came from on board journals, until 1990, when data was passed in an unique spreadsheet. Now the database counts about 36000 observations of 22 parameters, intended as a separate datasets, reporting date and time of the observation, name of the sampling station (where present) and geographical coordinates. An extract of the database can be seen in Fig. 2.

The database is dynamic, because it is always growing due to new observations from cruises, sensors or laboratory analysis.

2 Implementation

The application of Open Science principles to the LTER marine database of the NAS passes through some fundamental steps:

Task 1: data harmonization since the database is heterogeneous it must be harmonized before being exploited for any analysis or being published for sharing;

Task 2: metadata collection since the database covers a time range of 50 years, methods and instruments changed over time. It is necessary to collect all available information regarding data, methods and tools in order to evaluate reliability of data for any possible purpose;

Task 3: data exploitation a first exploitation of data is accomplished through the upload of data into GET-IT (Geoinformation Enable Toolkit StarterkIT [17]), an interoperable suite software for enabling researchers to share data and metadata in a SDI (Spatial Data Infrastructure). GET-IT allows the management, sharing and visualization of observational data;

Task 4: sharing and dissemination in coherence with the Open Science framing, it is fundamental to plan sharing moments, intended here not only as publication of research inputs and outputs, but also the dissemination of new concepts and methods (e.g. by participating in meetings and conference) and the promotion of exchanges with other researchers, potentially interested into the same topic, which could pick new ideas for their work and, conversely, improve EcoNAOS project;

Task 5: data citation since our database is dynamic (e.g new observations are added by different people), data citation can not be intended in a traditional way. Anyway we could perform queries on data and cite only a portion of the database. It is therefore fundamental to set up versioning methods for data citation;

Task 6: guidelines preparation preparation of guidelines could be useful for summarizing the EcoNAOS experience, contributing to the improvement of

Openness of scientific workflows and FAIRness of research data in future research projects.

In the next paragraphs we will go deep into the work done with particular focus on tasks to address or currently in progress in order to better understand how we can face specific issues and define a fair starting point for near-future work.

2.1 Completed Tasks

Task 1: Data Harmonization. Probably due to some errors in transcribing old data from on board journals to spreadsheet and to some inconsistencies in the name of sampling stations, the database was quite heterogeneous. First of all, the names of the stations were homogenized maintaining the most recently used station codes. Then we corrected the position of the points wrongly located on land into their right position at sea by matching the sample station names. These procedures have been executed in GRASS GIS [13] and vector layers of sampling stations and 3D layer of observations have been produced. The whole procedure has been automatized by means of a Python code, released under GNU GPL v.3 license and available on GitHub at the following link: https://github.com/CNR-ISMAR/econaos/tree/master.

Task 2: Metadata Collection. With a database covering a 50 years period, methods and instruments changed over time, in terms of unit of measures, standard errors associated to observations and so on, which could lead to database inconsistency when treated as a unique element. With the aim to prevent this data inconsistency, we collected all possible metadata associated to instruments and data. We started from the historical cruises examining instruments on board of research vessels: we produced some technical reports [14] listing all the instruments, units of measures and other elements qualifying data reliability in relation to the period/method of the observation. We then examined the whole database including more recent data and, for each parameter, we listed the instruments/methods used in time. All this information is stored into the database and associated to each single observation. However, no parameter values were manipulated and recalculated due to changes in the unit of measure, in order to preserve original data following the principle to "Keep raw data raw" suggested in best practices works [15,16]. A summary of parameters and sensors is reported in Table 1. These methods/instruments have been also published in GET-IT infrastructure and described using SensorML language.

Task 3: Data Exploitation. We decided to upload a first set of data in the GET-IT platform, in order to share them in an interoperable and standard way, compliant with INSPIRE Directive (https://inspire.ec.europa.eu/about-inspire/563) and Open Geospatial Consortium - OGC (http://www.opengeospatial.org/about). GET-IT has been developed for the broader RITMARE project and it is a distributed and interoperable SDI that allows at

Table 1. Review of methods/sensors used in time for each parameter of the database.

Parameter	Nr. of samples	Temporal coverage	Current method or sensor	Unit of measure	Previous method or sensor
Transparency	7840	1965–2015	Secchi Disk	m	-
Temperature	35594	1965–2015	CTD	C	Tilting thermometer
Salinity	35639	1965–2015	CTD	PSU	Autosal salinometer; Salinometer Bisset-Berman; Mohr-Knudsen Titration
Density	28253	1965–2015	Derivation parameter	Kg/mc	-
pH	12833	1965–2011	CTD	-	Strickland and Parsons; Grasshoff 1999; Beckman Zeromatic
Alkalinity	2948	1965–2002	Titrino titration	meq/l	-
Oxygen	12176	1965–2012	CTD	cc/l	Winkler titration; Methron titration; Titrino titration
N-NH3	10401	1965–2015	Easy Chem Plus	mol/l	Strickland and Parsons; Grasshoff 1983; Grashoff 1999
N-NO2	10478	1965–2015	Easy Chem Plus	mol/l	Strickland and Parsons; Grasshoff 1983; Grashoff 1999
N-NO3	10545	1965–2015	Easy Chem Plus	mol/l	Strickland and Parsons; Grasshoff 1983; Grashoff 1999
P-PO4	10448	1965–2015	Easy Chem Plus	mol/l	Strickland and Parsons; Grasshoff 1983; Grashoff 1999
Si-SiO4	10669	1965–2015	Easy Chem Plus	mol/l	Strickland and Parsons; Grasshoff 1983; Grashoff 1999
Chlorophyll-a	11121	1965–2015	Spettrofluorimeter Holm Hansen	g/l	Spettrophotometer Perkin Elmer SCOR
Pheopygments	5960	1979–2015	Spettrofluorimeter Holm Hansen	mg/l	Spettrophotometer Perkin Elmer SCOR
Phytoplankton	2949	1977–2015	Inverted microscope	Cell./l	-
Diatoms	2949	1977–2015	Inverted microscope	Cell./l	-
Dinoflagellates	2949	1977–2015	Inverted microscope	Cell./l	-
Coccolithophores	2949	1977–2015	Inverted microscope	Cell./l	-

sharing data and instruments. Using GET-IT, we can visualize at the same time different types of spatial data referring to one or more geographical areas or create graphs of available data for specific sampling station and sensors (Fig. 3). It is possible to experience GET-IT capabilities following this link: http://demo2.get-it.it/.

Task 4: Sharing and Dissemination As showed in Fig. 1, almost all the steps of a research lifecycle can be shared or participated. At the present time we took some initiatives in order to share our work:

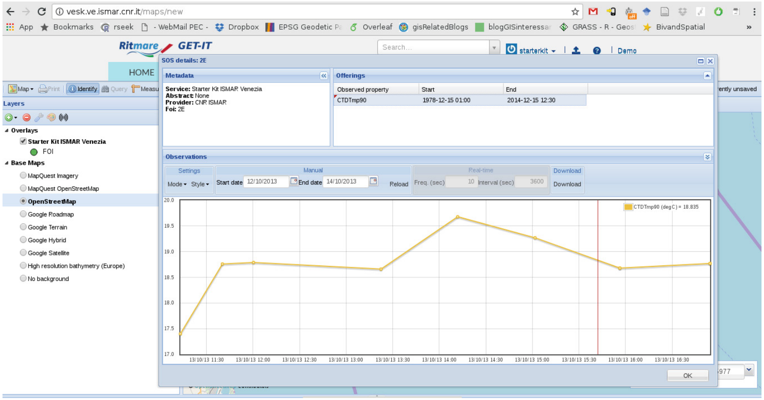

Fig. 3. A graph created by querying temperature data on "2E" sampling station.

- We published a "research idea" [5] as statement of our project in RIO journal (Research Ideas and Outcomes). We have chosen this specific journal for many reasons:
  - it is Open Access;
  - it includes a wide range of publication types that cover all the research life-cycle;
  - it implements an open peer review process where reviews are open, allowing research to be transparent and anyone from outside the process to comment;
  - reviews have a DOI, so that the review work is recognized and directly citable (e.g. [18,19]).
- We shared the code used to harmonize the database. A data sample, pseudocode, the Python code and a brief description, are available under GNU GPL v.3 license and in the release 0 of the code [20].
- We shared a subset of data by publishing in GET-IT the superficial (at the sea surface) value for each observation and for each parameter in the database. Metadata of the sensors are also shared in GET-IT, using Sensor Metadata Language (SensorML) edited by using the EDI interface [21,22], which is a powerful metadata editor, developed for RITMARE project as well. These metadata are accessible following this link: http://vesk.ve.ismar.cnr.it/sensors/.
- We organised a workshop involving other researchers from the NAS LTER_EU_IT_012 macrosite in order to present our work and to understand their feelings towards Open Science and Open Access. There is a real interest in these themes but there's still some fear in losing intellectual property of data. Anyway, researchers manifested their interest in sharing a portion of their data in order to analyze the variation of an index (TRIX, [23]) along Mediterranean Sea. This experiment is currently ongoing and it is conceived in order to show how Open Data and Open Science can speed up analyses

on wide geographical zones and boost cooperation between different research groups.
– We presented our work at national and international conferences: XXVII Conference of Italian Society of Ecology (SItE) publishing a contribution on congressional records [24]; RDA Ninth Plenary Meeting, presenting a poster and attending many interesting meetings about scientific data handling; EurOcean conference on ocean observation systems, presenting a poster and we wrote an article for a book about Responsible Research and Innovation [25].

A schema of these dissemination actions is reported in Fig. 4. There, it is possible to evidence that some general steps reported in Fig. 1 changed in reason of the specific research project.

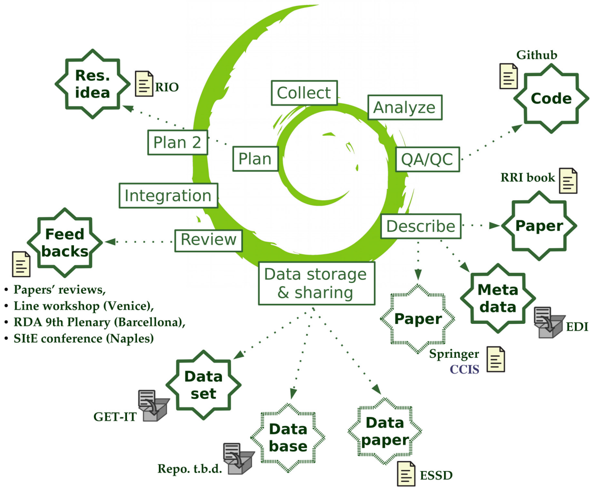

Fig. 4. The Open Research Project Lifecycle related to LTER data in the Northern Adriatic Sea macrosite. Filled-line callouts identify dissemination actions already taken, dotted-line callouts identify dissemination actions in progress.

2.2 Ongoing Works

Task 4: Sharing and Dissemination We are currently writing a datapaper and planning to release the entire database in Open Access mode. The database released will be as compliant as possible with the FAIR data principles. For this purpose, we chose Earth System Science Data, which is an Open Access journal allowing scientific data publication in repositories with a persistent identifier, with an open license (e.g. CC-BY: anyone must be free to copy, distribute, transmit, and adapt the data sets as long as he/she gives credit to the original authors), accessible over the Web, and long-term availability.

Task 5: Data Citation The study of data citation comes from library science, lately it has been extended to the more general data science.

Our database is a dynamic database since it changes as long as we collect new observations recorded by sensors and during cruises. Despite the different available technological solutions, it is relatively easy to share a standard database, conversely dynamic databases rises some new questions:

– how to make data available for citation? if the database changes, citation must change accordingly;
– how to cite a portion of a dataset and how to cite aggregated data?
– how to make a citation persistent if the original database changes?
– how to update the reference to the database, if data changed?
– what is the threshold to define a "substantial" change in the database, to the point that its reference must change accordingly?

Facing these questions is a relatively recent issue, in fact data citation itself is a quite recent theme, even if the importance of data publishing with particular attention to data reference is fairly recognized [26,27]. A review on the evolution in time of data citation is formulated by Altman and Crosas [28] and some fundamental concepts are expressed by Silvello [29].

Our database had two main characteristics: it is dynamic and contains oceanographic data.

Dynamic database citation case has been recently afforded from RDA working group on dynamic database citation which first created a set of 14 rules [30] in order to create an automated mechanism able to:

– identify and cite databases or portions of it;
– cite and retrieve data also in past versions of database;
– be interoperable.

These rules contain recommendations on how to prepare data and store queries, how to persistently identify datasets, how to retrieve portions of database in a client-server architecture and what to focus on if database infrastructure is modified. Another strength point of RDA WG on data citation rules is that their application is in continuous testing phase on real evolving databases. The monitoring of rules applications on these pilot projects will lead to rules improvement and adaptation based on specific and generic requirements. A framework for citing evolving data, by quantification of the changes in the database as long as new data are added, is formulated in [31,32]. Citation of aggregated data has been also afforded by Baker, Amsi and Uytvanck [33].

Oceanographic data citation is matter of a literature work collecting all the pilot projects on this theme [34]. Urban et al. not only identify projects but also Open Access journals which follow specific rules for database release or Open Access Servers and repositories (e.g. Earth System Science Data journal, PANGEA repository and the Woods Hole Open Access Server). The Ocean Link project [35] focuses on data discovery instead and it is a web platform which identifies "links between data centers, digital repositories, and professional societies to enhance discovery, enable collaboration, and begin to assess research contribution" on oceanographic projects. Particularly interesting is the use of semantic-based systems for data discovery. A very specific study on Long Term Oceanographic data practices is conducted by Baker and Chandler [36]. This work focuses on growth of information infrastructures that support multi-scale sampling, data repositories, and data integration, comparing two projects. The work also recognize the importance of a global information management strategy (long and short term) for any long term oceanographic research project.

Guidelines Preparation. Another task to tackle is the preparation of guidelines, which could both resume and detail our experience. In fact, the scope of these guidelines is both to facilitate (easily explain) and guide (detailing phases) researchers through an application of Open Science principles to their research projects. In general, we suppose we could categorize actions in two groups:

- acting generically on process: all the actions aimed at making project more "Open Science compliant". Examples are the release of source code, drafting of data policy, choice of journals with open peer review, etc.
- acting specifically on data: actions aimed at making data as FAIR as possible. For example operations on database structure, semantic harmonization, efforts on interoperability, etc.

Obviously the guidelines preparation can only be faced when all the other tasks are completed, in order to oversee the entire process when the application of Open Science principles is "complete".

3 Conclusions and Future Perspectives

The EcoNAOS project represents an attempt to apply Open Science principles not only to specific and sporadic research products but to a whole research project and it proposes a complete workflow from the formulation of a research idea to the review of the whole project by the scientific community. It is specifically conceived around marine science and LTER data but the application of these principles can be extended to projects in any research topic and involving any kind of data. Moreover, the projects envisages a number of moments to share ideas, material and revisions with other researchers and the scientific community: this allowed us not only to facilitate exchanges with researchers working on the same topics, but also to deepen into explicit and implicit barriers to a complete application of Open Science principles in science. Results about this theme are in an analysis phase and soon to come. EcoNAOS does not want to overturn previous researchers' habits, but since Open Science must be helpful and facilitate research, our work wants to provide an example of Open Science application at everyday research work. This model might not be exhaustive or complying with all the possible necessities, so it might be completed and adapted according to specific requirements.

After having collected above cited literature information and experiences on dynamic database citation and versioning, we plan to focus on data citation and versioning task in the near future [37–39].

For sure, the interest on deepen this argument is exploring new perspectives of data citation of complex data in format and substance. For example, for geographical data, there are not many studies involving data citation and OGC standards. From a data analysis point of view, it could be interesting to study citation methods of complex data which must be ordered, extracted and aggregated with the aim to be distributed via web following up an OGC standard request. These requests often involve not only one type of data, but multiple types of data (numeric, geographical, metadata, semantic).

Moreover, aggregated data citation is still an open research topic and there are not standard and well-defined procedures. In our specific case, since we want to redistribute data (or portions of data) following up a request from a client, the definition of a URI (Uniform Resource Identifier) is a central point. Then, once the URI is defined, the data attribution is also a central element. Another interesting point is that, while data citation science is exploring new and always growing perspectives in other fields (bibliometrics, economy, music, biology), there are still few significant experiences and applications in long term series of spatially distributed oceanographic and ecological data. Consequently, there are not standards to cite oceanographic datasets. This could represent a good starting point for a new and different research topic. Since this seems to be a wide and still open theme, we propose, for this specific task, to enlarge the discussion and to deepen in parallel the argument of oceanographic and ecological data citation.

References

1. Stevan, H., Brody, T.: Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-lib Mag. **10**(6) (2004). https://doi.org/10.1045/june2004-harnad
2. Hajjem, C., Harnad, S., Gingras, Y.: Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. arXiv preprint cs/0606079 (2006)
3. Gargouri, Y., et al.: Self-selected or mandated, open access increases citation impact for higher quality research. PloS one **5**(10), 13636 (2010). https://doi.org/10.1371/journal.pone.0013636
4. McKiernan, E.C., et al.: How open science helps researchers succeed. eLife **5** (2016). https://doi.org/10.7554/elife.16800
5. Minelli, A., et al.: The project EcoNAOS: vision and practice towards an open approach in the Northern Adriatic Sea ecological observatory. Res. Ideas Outcomes **4** (2018). https://doi.org/10.3897/rio.4.e24224
6. European Commission: Open innovation, open science, open to the world - a vision for Europe. (2016) RTD-PUBLICATIONS@ec.europa.eu [ISBN 978-92-79-57346-0] 10.2777/061652

7. David, P.A.: The economic logic of "open science" and the balance between private property rights and the public domain in scientific data and information: a primer. The Role of the Public Domain in Scientific and Technical Data and Information 19–34 (2003)
8. Peters, M.A.: Open education and the open science economy. Yearb. Nat. Soc. Study Educ. **108**(2), 203–225 (2009)
9. European Commission: H2020 Programme - Guidelines on FAIR Data Management in Horizon 2020 (2016). http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
10. Pennock, M.: Digital curation: a life-cycle approach to managing and preserving usable digital information. Libr. Arch. **1**, 34–45 (2007)
11. Ingram, C.: How and why you should manage your research data: a guide for researchers. An introduction to engaging with research data management processes, JISC (2016)
12. Tennant, J.P., et al.: A multi-disciplinary perspective on emergent and future innovations in peer review. F1000Research **6**, 1151 (2017). https://doi.org/10.12688/f1000research.12037.3
13. GRASS Development Team: Geographic Resources Analysis Support System (GRASS) Software, Version 7.2. Open Source Geospatial Foundation (2017). http://grass.osgeo.org
14. Scovacricchi, T.: Technical recovery of material and methods applied in the cruises of 1965, 1966, 1978/79 as described by P. Franco (1970, 1972, 1982). Internal Report, CNR-ISMAR Venezia (Italy), (2017)
15. Borer, E.T., Seabloom, E.W., Jones, M.B., Schildhauer, M.: Some simple guidelines for effective data management. Bull. Ecol. Soc. Am. **90**(2), 205–214 (2009). https://doi.org/10.1890/0012-9623-90.2.205
16. Hart, E.M., et al.: Ten simple rules for digital data storage. PLoS Comput. Biol. **12**(10), e1005097 (2016). https://doi.org/10.1371/journal.pcbi.1005097
17. Oggioni, A., et al.: Interoperability in marine sensor networks through SWE services: the RITMARE experience, pp. 200–223 (2017) https://doi.org/10.4018/978-1-5225-0700-0.ch009
18. Marchesini, I.: Review of: the project EcoNAOS: vision and practice towards an open approach in the Northern Adriatic Sea ecological observatory. Res. Ideas Outcomes **4**, e24224 (2018). https://doi.org/10.3897/rio.4.e24224.r72863
19. Peterseil, J.: Review of: the project EcoNAOS: vision and practice towards an open approach in the Northern Adriatic Sea ecological observatory. Res. Ideas Outcomes **4**, e24224 (2018). https://doi.org/10.3897/rio.4.e24224.r72862
20. Minelli A: Data Harmonisation for EcoNAOS project (Version 0). https://doi.org/10.5281/zenodo.1416102
21. Pavesi, F., et al.: EDI - a template-driven metadata editor for research data. J. Open Res. Software **4**(1) (2016). https://doi.org/10.5334/jors.106
22. Tagliolato, P., Oggioni, A., Fugazza, C., Pepe, M., Carrara, P.: Sensor metadata blueprints and computer-aided editing for disciplined SensorML. IOP Conf. Ser. Earth Environ. Sci. **34**(1), 012036–012036 (2016). https://doi.org/10.1088/1755-1315/34/1/012036
23. Vollenweider, R.A., Giovanardi, F., Montanari, G., Rinaldi, A.: Characterization of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity and generalized water quality index. Environ. Official J. Int. Environ. Soc. **9**(3), 329–357 (1998)

24. Minelli, A., et al.: Handling Long Term Ecological marine data: an Open Science approach. La ricerca ecologica in un mondo che cambia. Libro degli Abstract, XXVII Congresso Nazionale della Società Italiana di Ecologia (2017)
25. L'Astorina A., Di Fiore M.: Scienziati in affanno? Ricerca e Innovazione Responsabili (RRI) in teoria e nelle pratiche. 180 p. Cnr Edizioni. ISBN 978 88 8080 250 1, (2018). https://doi.org/10.26324/2018RRICNRBOOK
26. Wood, J., et al.: Riding the wave: How Europe can gain from the rising tide of scientific data. European Union (2010)
27. Ball, A., Duke, M.: How to Cite Datasets and Link to Publications. DCC How-to Guides, Edinburgh: Digital Curation Centre (2015). http://www.dcc.ac.uk/resources/how-guides
28. Altman, M., Crosas, M.: The evolution of data citation: from principles to implementation. IAssist Quarterly **37**, 62–70 (2013)
29. Silvello, G.: Theory and practice of data citation. J. Assoc. Inf. Sci. Technol. **69**(1), 6–20 (2018)
30. Rauber, A., et al.: Data citation of evolving data: recommendations of the working group on data citation (WGDC). Result of the RDA Data Citation WG **20** (2015)
31. Proll, S., Rauber, A.: Scalable data citation in dynamic, large databases: model and reference implementation. In: 2013 IEEE International Conference on Big Data. IEEE (2013)
32. Proell, S., Rauber, A.: A scalable framework for dynamic data citation of arbitrary structured data. In: 3rd International Conference on Data Management Technologies and Applications (DATA2014) (2014). https://doi.org/10.5220/0004991802230230
33. Rauber, A., Amsi, A.: Uytvanck, Dv., Pröll, S.: Identification of Reproducible Subsets for Data Citation. Sharing and Re-Use. Bulletin of IEEE Technical Committee on Digital Libraries, Special Issue on Data Citation, 6–15 (2016)
34. Urban, E., et al.: Pilot projects for publishing and citing ocean data. Eos, Trans. Am. Geophys. Union **93**(43), 425–426 (2012)
35. Narock, T., et al.: The oceanlink project. In: 2014 IEEE International Conference on Big Data (Big Data). IEEE (2014)
36. Baker, K.S., Chandler, C.L.: Enabling long-term oceanographic research: changing data practices, information management strategies and informatics. Deep Sea Res. Part II: Topical Stud. Oceanograp. **55**(18–19), 2132–2142 (2008)
37. Angelini, M., Fazzini, V., Ferro, N., Santucci, G., Silvello, G.: CLAIRE: a combinatorial visual analytics system for information retrieval evaluation. Inf. Process. Manage. **54**(6), 1077–1100 (2018)
38. Alawini, A., Davidson, S.B., Silvello, G., Tannen, V., Wu, Y.: Data citation: a new provenance challenge. IEEE Data Eng. Bull. **41**(1), 27–38 (2018)
39. Agosti, M., Ferro, N., Silvello, G.: Digital libraries: from digital resources to challenges in scientific data sharing and re-use. In: Flesca, S., Greco, S., Masciari, E., Saccà, D. (eds.) A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years. SBD, vol. 31, pp. 27–41. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61893-7_2