H2020-INFRAEOSC-2018-3

NI4OS-Europe

National Initiatives for Open Science in Europe

Deliverable D4.2

Data repository integration and ORDM/FAIR compliance guidelines

Lead beneficiary(s): UOB

Author(s): Branko Marović, Eleni Toli, Electra Sifakaki, Elli Papadopoulou, Vladimir Otašević, Milica Ševkušić, Biljana Kosanović

Status –Version: Final – e

Date: March 4, 2020

Dissemination Level: Public

Abstract: This document delivers the results of the analysis of guidelines for open science – more specifically, a description of the approach to ORDM and FAIR for 15 NI4OS-Europe partner countries, by observing the relevant existing guidelines on data repository integration, ORDM and FAIR on the global level. The applied methodology for selection, adoption and promotion of relevant guidelines is also described.

The NI4OS-Europe project is funded by the European Commission under the Horizon 2020 European research infrastructures grant agreement no. 857645.

The NI4OS-Europe Consortium consists of:

GRNET SA	Coordinator	Greece
ATHENA RC	Beneficiary	Greece
CYI	Beneficiary	Cyprus
UCY	Beneficiary	Cyprus
IICT	Beneficiary	Bulgaria
SRCE	Beneficiary	Croatia
RBI	Beneficiary	Croatia
KIFU	Beneficiary	Hungary
DE	Beneficiary	Hungary
ICI BUCURESTI	Beneficiary	Romania
UEFISCDI	Beneficiary	Romania
ARNES	Beneficiary	Slovenia
UMUKM	Beneficiary	Slovenia
IPB	Beneficiary	Serbia
UOB	Beneficiary	Serbia
RASH	Beneficiary	Albania
UNI BL	Beneficiary	Bosnia-Herzegovina
UKIM	Beneficiary	North Macedonia
UOM	Beneficiary	Montenegro
RENAM	Beneficiary	Moldova (Republic of)
IIAP NAS RA	Beneficiary	Armenia
GRENA	Beneficiary	Georgia

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
10.12.2019	a	Branko Marović	Initial ToC defined
22.03.2020	b	Branko Marović, Vladimir Otašević	Bulk initials content provided
25.03.2020	c	Branko Marović, Eleni Toli, Electra Sifakaki, Elli Papadopoulou,	Additions (mostly from WP4 Tools Catalogue), clarifications, comments and corrections, alignment with D4.3
03.03.2020	d	Branko Marović, Milica Ševkušić, Eleni Toli, Elli Papadopoulou, Biljana Kosanović, Mihajlo Savić (reviewer),	Minor additions and clarifications, comments and corrections
03.03.2020	e	Branko Marović	Final check and formatting

Table of contents

1. Introduction 10

2. Methodology 12

2.1. Overall governance of WP4 instruments 12
2.2. Desk research on guidelines and tools 13
2.3. Categorisation of instruments 13
2.4. Findings from NI4OS-Europe landscaping survey 15
2.5. Data protection 16
2.6. Maturity models 17
2.7. Instrument customisation and development 17

3. Selected guidelines and tools 19

3.1. Overall policies and approaches 20
3.1.1. Instruments: Policy and approach 21
3.1.2. Instruments: Registries 22
3.1.3. Instruments: Standards, APIs and specifications 23
3.2. FAIR 23
3.2.1. Instruments: FAIR 25
3.2.2. Instruments: PIDs 25
3.3. Open 25
3.3.1. Instruments: Open 26
3.4. Overall ORDM and DMPs 27
3.4.1. Instruments: ORDM and DMPs 29
3.5. Access control 30
3.5.1. Instruments: Access control 31
3.6. IPR 31
3.6.1. Instruments: IPR 32
3.7. Management of repositories and data sets and certification 32
3.7.1. Instruments: Management of repositories and data sets and certification 33
3.8. Specialised repositories 33
3.8.1. Instruments: Specialized repositories 34
3.9. Machine readability and metadata 34
3.9.1. Instruments: Machine readability and metadata 35
3.10. Monitoring, metrics and quality indicators 36
3.10.1. Instruments: Monitoring, metrics and quality indicators 37

4. Conclusions 39

References

[1] Project NI4OS-Europe-857645 – Annex I – Description of the Action, (n.d.).

[2] Final Report and Action Plan from the European Commission Expert Group on FAIR Data: Turning FAIR into reality, http://doi.org/10.2777/1524, https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf, https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en

[3] Deming Cycle (PDCA) https://link.springer.com/referenceworkentry/10.1007%2F1-4020-0612-8_229

[4] European Open Science Cloud for Research pilot project (EOSCpilot) Deliverable 3.2: EOSC Open Science Monitor specifications, https://eoscpilot.eu/content/d32-eosc-open-science-monitor-specifications

[5] European Open Science Cloud for Research pilot project (EOSCpilot) Deliverable 3.5: Open Science Policy Toolkit, https://eoscpilot.eu/content/d35-open-science-policy-toolkit

[6] European Open Science Cloud for Research pilot project (EOSCpilot) Deliverable 3.7: Updates to Policy-Supporting Services, https://eoscpilot.eu/content/d37-updates-policy-supporting-services

[7] European Open Science Cloud for Research pilot project (EOSCpilot) Deliverable 3.6: Final Policy Recommendations, https://eoscpilot.eu/content/d36-final-policy-recommendations

[8] Data Maturity Model WG, https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

[9] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18, https://www.nature.com/articles/sdata201618,https://www.go-fair.org/fair-principles, https://www.force11.org/group/fairgroup/fairprinciples

[10] Open Archives Initiative Protocol for Metadata Harvesting, https://www.openarchives.org/pmh/.

[11] Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak, https://wellcome.ac.uk/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-covid-19-outbreak,

List of Tables

TABLE 1: CHARACTERISTICS OF FAIR AND OPEN DATA..... 26

List of Acronyms

AAI Authentication and Authorization Infrastructure

API Application Program Interface

CRIS Current Research Information System

CTS CoreTrustSeal, a core level certification for data repositories based on the DSA-WDS Core Trustworthy Data Repositories Requirements catalogue and procedures

D2.1 NI4OS-Europe Deliverable 2.1: Stakeholder map, inventory, policy matrix

D4.1 NI4OS-Europe Deliverable 4.1: Incentives for supporting ORDM and FAIR

D4.2 NI4OS-Europe Deliverable 4.2: Data repository integration and ORDM/FAIR compliance guidelines

D4.3 NI4OS-Europe Deliverable 4.3: Mapping of legal, technical and procedural tools

D4.4 NI4OS-Europe Deliverable 4.4: Delivery of data management and certification tools (type Demo)

D4.5 NI4OS-Europe Deliverable 4.5: Delivery of legal, technical and procedural tools (type Demo)

D4.6 NI4OS-Europe Deliverable 4.6: Second set of data management and certification tools (type Demo)

D4.7 NI4OS-Europe Deliverable 4.7: Second set legal, technical and procedural tools (type Demo)

D4.8 NI4OS-Europe Deliverable 4.8: Final report on WP4 tools

D5.1 NI4OS-Europe Deliverable D5.1: Provider landscape analysis and provider categorization

DMP Data Management Plan, a formal document that outlines how data are to be handled both during a research project and after the project is completed

EC European Commission

EOSC European Open Science Cloud

EU European Union

FAIR Findable, Accessible, Interoperable and Reusable

GDPR EU General Data Protection Regulation

IPR Intellectual Property Rights

M4.1 NI4OS-Europe Milestone 4.1: Data repository integration and ORDM/FAIR compliance guidelines available

M4.2 NI4OS-Europe Milestone 4.2: First set of data management and certification tools in place

NDA Non-disclosure Agreement

OA Open Access

ORDM Open Research Data Management

OS Open Science

PDCA Plan, Do, Check, Act cycle; also known as the Deming circle

PID Persistent Identifier

RDM Research Data Management

RI Research Infrastructure

RFO Research Funding Organisation

RPO Research-performing Organization

SLA Service Level Agreement

T3.1 Harmonization of best practices for on-boarding of generic and thematic service providers

T4.1 NI4OS-Europe Task 4.1: Developing rewards and incentives mechanisms for ORDM and FAIR

T4.2 NI4OS-Europe Task 4.2: Policies and guidelines for incorporation and implementation of FAIR data practices

T4.3 NI4OS-Europe Task 4.3: Open and FAIR data in practice: tools for RDM following the data lifecycle

T4.4 NI4OS-Europe Task 4.4: Certification schemes for a "free flow of data"

TF Task Force

ToS Terms of Service

URL Uniform Resource Locator

WG Work Group

WP2 NI4OS-Europe Work Package 2: EOSC national initiatives and policy support

WP3 NI4OS-Europe Work Package 3: Federation and on-boarding operational mechanisms and tools

WP4 NI4OS-Europe Work Package 4: ORDM standards, processes, tools and certification schemes

WP5 NI4OS-Europe Work Package 5: Service provider integration, onboarding and service maintenance

Executive summary

What is the focus of this Deliverable?

The purpose of this deliverable is to identify available technical guidelines and technical, organisational and legal tools supporting the integration of data repositories and compliance with ORDM and FAIR requirements while aiming to apply these instruments in NI4OS-Europe partner countries.

What is next in the process to deliver the NI4OS-Europe results?

D4.1 defined the set of effective incentives and rewards for the application of FAIR and open research data management standards, processes, tools and certifications. Different instruments for the adoption of ORDM and FAIR are identified, selected and mapped in D4.2 (this document and associated milestone M4.1) and forthcoming (M7) D4.3 "Mapping of legal, technical and procedural tools". They will be followed by the implementation of demonstrable data management and certification (D4.4 and M4.2, M10) and legal, technical and procedural (D4.5, M14) tools. These tools will help the onboarding of services through WP3 and WP5, and the second iteration of demonstrations (D4.6 in M25, and D4.7 in M26). The final report on the application of delivered instruments will be provided in D4.8 (M36).

What are the deliverable contents?

This deliverable sets the basis for the management of FAIR, ORDM and repository integration guidelines and tools, also pointing out the key tools that should be applied in the management of NI4OS-Europe repositories. Its content is based on the information from partner institutions, the landscaping study conducted as part of WP2 and the exploration of the existing instruments used in the technical implementation of open science and FAIR repositories and data sets.

As the line between guidelines and tools is often blurred, they will be jointly managed. The focus in this document is on the elaboration of the context, methodology and areas that are to be covered, while the particular instruments collected in the framework of the NI4OS-Europe WP4 activities will be detailed in the forthcoming D4.3 "Mapping of legal, technical and procedural tools".

Conclusions and recommendations

This deliverable offers a detailed analysis and categorization of the existing technical solutions and models addressing issues related to FAIR and ORDM, as well as practical concerns of NI4OS-Europe repositories integration. This snapshot will be useful in bridging the gap between specific user groups' needs and existing tools and models. Thus, this work will help us to spot inconsistencies in the current landscape, identify user requirements, and most importantly, design fit-for-purpose tools for both EOSC stakeholders and the EOSC-Core architecture.

1. Introduction

This deliverable is part of the Work Package 4 (WP4) of the NI4OS-Europe project [1] and is only one segment of its activities that aim to provide clear pathways, as well as new products that would make the transition to national Open Science ecosystems easier. Hence, WP4 deals with the development of guidelines, tools, mechanisms and processes that assist in the formulation of Research Data Management (RDM) policies; facilitate Data Management Plans (DMP) management; ensure technical compliance of infrastructures and services with the FAIR principles; assess data FAIRness; support decision-making; define legal and ethical actions that enable to open sharing of scientific data; help managing license compatibility, etc. All activities are undertaken in support of and in coordination with WP2, which is tasked with setting up national open science initiatives and supporting the establishment of governance policies.

As T4.1 has defined incentives for supporting ORDM and FAIR in D4.1, this deliverable (D4.2) is focused on providing a solid approach to the adoption of guidelines, but also related tools and other instruments managed within WP4. It is an output from T4.2, oriented on the incorporation and implementation of FAIR data practices. As T4.3 deals with the practical implementation and delivery of RDM and FAIR tools and T4.4 with certification schemes, it has been decided to clearly separate the scope of D4.2 and D4.3 (produced by T4.3), so that D4.2 will refine the methodological approach for all guidelines and other instruments that support ORDM and FAIR, while D4.3 will be focused on actual categorisation and elaboration of specific guidelines, models and tools used for decision-making, certification and support. This clear delineation between the two deliverables and corresponding tasks helps in taking a coherent approach for all WP4 tasks and contributes to the achievement of its objectives. Further down the line, T3.1 deals with procedures and policies for the onboarding of the services to the EOSC service catalogue. Guidelines, tools and certificates are instruments that should support and prepare the services for the onboarding process, while the actual onboarding is a strict and formal process that is achieved only after all requirements are met. The joint work of WP4 and WP3 should result in technical policies and interoperability with all EOSC projects, OpenAIRE and other core initiatives, while iteratively improving in a PDCA cycle.

In preparation of this document, the strategic report and action plan from the European Commission Expert Group on FAIR Data [2] was used as a reference in the identification of the areas that should be considered in the analysis. At the same time, the local perspectives and goals of NI4OS-Europe are taken into consideration in order to facilitate data practices across national and regional communities. Similarly, domain-specific and generic approaches and concerns for federated services need to be considered, also bearing in mind the fact that the long tail of science in repositories and services is a particularly challenging topic.

In the FAIR and ORDM realms, many guidelines have been developed by a number of projects and initiatives, while numerous working groups are discussing and developing new ones. The guidelines that are most relevant for NI4OS-Europe should be selected; some guidelines, tools or templates may need to be customised; some gaps could be filled. Hence, this document outlines an overall methodology for the management of repositories, as well as FAIR and ORDM instruments. The work of WP4 will synergise with WP3 procedures before the validation through actual service onboarding and delivery and use of integration tools in WP5. This interaction will help in providing and refining practical

and realistic guidelines but also in the approach to the integration of repositories and data assessment.

The application of the approach described in this deliverable, along with the application of the guidelines and tools and establishment and enforcement of related policies should result in:

- Selection and specification of common rules, guidelines and standards for harmonising the delivery and adoption of the data-oriented services which are practical and applicable in local contexts;
- Establishing and mainstreaming certification schemes for ORDM and the service management of data repositories and the GDPR compliance;
- Harmonization with standards, tools and mechanisms that are widely used in ORDM;
- Establishment of high-quality ORDM practices;
- Facilitating the delivery and adoption of the EOSC services at national and institutional levels, while integrating NI4OS-Europe repositories and services into the wider EOSC infrastructure;
- Alignment with international initiatives on research data sharing;
- Contribution to the harmonization of related policies in Europe.

2. Methodology

2.1. Overall governance of WP4 instruments

The overall process for the management of WP4 instruments (guidelines, tools, models, certification schemes and standards) is the common Deming improvement cycle (PDCA) [3]:

- Plan – Prepare: elaborate needs, instruments that are needed, assess their application areas, set expectations, define actions (use existing, adapt or develop)
- Do – Core work: test, adopt and apply instruments
- Check (Study) – Evaluate: identify discrepancies, gaps, new needs, relevant developments, and environmental changes
- Act (Adjust) – Adjust plans and methodology, enhance process and tools for dealing with instruments, change instruments

This process will be repeated for all individual stages of WP4, including the development of guidelines mapping of tools, and the development of tool demonstrations.

For example, for the mapping of tools, the PDCA cycle looks like this:

- Plan – Preparatory stage
    - Capture the open & FAIR landscape (desk research)
    - Identify strategic areas and goals
    - Identify stakeholder groups
    - Explore the needs of specific user groups
    - Initial mapping of tools (desktop research and input from the landscaping survey)
- Do – Core work
    - Analysis to see if these tools are up-to-date
    - Detailed evaluation
    - Detailed categorisation
- Check – Evaluate
    - Identify what is missing in the list of instruments
    - Check if there are some conflicts or overlaps in guidelines and tools
- Act – Post-evaluation
    - Spin-off collaboration and adaptation where possible or development where justified
    - The produced customisations, implementations or developments will be described and adopted within NI4OS-Europe, promoted and offered to relevant communities and the plans for their future tracking and support will be defined and put in place.

2.2. Desk research on guidelines and tools

D4.2 aims at providing an overview of existing sources for open and FAIR RDM to assist stakeholders in complying with relevant mandates at European and national level. Hence, the list of instruments is comprised of guidelines and tools that are essential building blocks of research ecosystems, touching more upon the technical side of open and FAIR implementation while also including models and workflows that support policymakers, RPOs and RFOs in adopting coinciding policies in their area of influence. Besides, the list contains tools and methods of immediate use by researchers to be embedded in their everyday work thus enhancing current research practices.

The tools catalogue was populated mostly with desk research which was later complemented by findings of the D2.1 survey. Moreover, landscape review took stock of preliminary work undertaken in the EOSCpilot project's WP3 policy work, namely the open science monitor [4] and the policy toolkit [5, 6], to ensure the continuation of work in the context of EOSC. The open science monitor provides a comparison of open and FAIR principles highlighting differences and commonalities between them, thus giving great input to WP4 activities focusing on Open and FAIR RDM. The policy toolkit lists sources that are necessary for the effective operation of the EOSCpilot policy supporting services (the open science monitor being one of them) and ensures successful implementation and adoption of the EOSCpilot policy framework [7] by its stakeholders. To meet the scope and objectives of NI4OS-Europe, the EOSCpilot toolkit was cleared up, updated, re-categorised and enriched with new sources found either directly from searching the internet or indirectly from attending talks and presentations on OS, FAIR and ORDM. On certain occasions, teleconferences between the WP4 team and tool developers were conducted to increase comprehension on how specific tools could be used by the WP4 team and by EOSC and NI4OS-Europe stakeholders.

Finally, the tools catalogue looks into services (and sometimes tools offered as services) that are crucial for EOSC to properly function taking into consideration the Rules of Participation and the EOSC architecture. Therefore, the work performed in this work package should not be confused with the one on onboarding of services (as sometimes tools are offered as services). On the contrary, it should be seen as complementary to the enrichment of EOSC-Core with necessary services that are yet to be developed or with tools and mechanisms that allow the NI4OS-Europe stakeholders to smoothly transition to more open science ecosystems and code of conduct.

2.3. Categorisation of instruments

In general, it is extremely difficult to establish a model that would differentiate various supporting artifacts, services or products related to open science. A straightforward approach would be to try to classify them by a set of clean and hopefully orthogonal dimensions such as type of the instrument and its method of application. With this approach we could have the following types:

- Guidelines -– Documents to be applied and consulted, such as guidelines, policies, specifications, standards, including collections of several related documents of different types but on the same primary topic.

- Models – Templates or sets of related templates, such as models for IPR, SLA, ToS, or DMP documents that are applicable in a specific set of situations but need to be tailored before use. Collections of exemplary documents of one specific type are also included.
- Tools – Interactive or self-guided tools that help their users perform some part of the work in an easier, usually semi-automated way; some of these tools require prior customization.

Another generalist classification is based on the function with the examples being:

- Support in execution of a task where the user is in the driver's seat
- Decision support, where the user is guided by the instrument through the decision-making process
- Certification, a formal scheme, procedure or tool used to certify or assess the user, service, repository, or data set

Whatever approach we adopt, some instruments may resist unequivocal classification. Training materials may be delivered as static documents, as applications, or as databases structured in a particular way. A multi-options template may be wrapped in a web application which, after guiding the user through several questions or a decision tree, delivers a custom document or recommendation. Also, different people understand and classify things in different ways. It is the same with some popular repositories of resources; for example, FAIRsharing uses four types of records: standards, databases, policies and collections. In that regard, a tool may be a database, but does not have to; a registry may be a collection or a database. It is not always possible to establish the mapping between such different paradigms and categorisations and provide the deterministic mapping. Finally, some instruments are genuine hybrids, constructed to satisfy several related goals and types of users.

Finally, intended users may be classified in different ways. But we can usually easily recognize some typical roles, such as researchers, research managers, librarians, data stewards, repository managers, service developers or providers, other operators or supporters, managers of research performing organisations (RPOs) or research infrastructures (RIs), policymakers, funders or ministries. One option would be to use the types of stakeholders identified in D2.1 and providers that will be listed in D5.1.

The task of classifying what is appropriate for whom and how it works is daunting and will be detailed in D4.3, while this deliverable aims to enumerate some key areas that are covered by various instruments, describe these areas and list some typical and popular instruments primarily associated with those areas.

While in Section 3. we provide one simple thematic structure for the classification of instruments, other views may also be needed. For certain, the thematic and functional domains of some instruments need to be simultaneously classified into several categories. The future needs may require the mapping between mutually related guidelines and tools, between similar or complementary guidelines, between instruments and steward organisations, or mappings to alternative classifications of thematic areas. Finally, the supporting experts, reference users and their experiences may need to be tracked throughout the use of recommended tools.

2.4. Findings from NI4OS-Europe landscaping survey

The landscaping survey for partners in the NI4OS-Europe collected 575 complete answers between October 21st and December 10th 2019. It provided an overview of open science in partner countries, with a preliminary analysis in D2.1 and D4.1. The charts, values and conclusions from D4.1 analysis were also used during the preparation of this document in order to check how the different thematic areas of guidelines and tools and related ORDM and FAIR related concepts fare in fifteen NI4OS-Europe partner countries.

For example, the number of institutional OA policies on research data is significantly lagging behind the number of policies on publications, reflecting the fact that the EU data mandate was introduced much later than most OA mandates for publications. When it comes to parameters used in the evaluation of researchers, the traditional parameters (researcher publications, participation in projects, attraction of funding, etc.) were highly rated, while the parameters related to OS activities received the lowest scores (within the range 1 – 5): data (average score – 3.6), software (3.2), OS and OA (3.4), and social outreach, knowledge transfer, citizen science (3.5). This indicates that these suppressed aspects need to be particularly supported by guidelines and tools.

When it comes to support and training provided by respondents, they mostly provide training in intellectual property rights and copyright (47%) and repositories (40%), followed by training in open practices (methodologies, peer review, metrics, citations, etc.) (38%), open education resources (36%) and licenses (31%), while only 26% provided training in RDM (publishing of open data, FAIR, RDM plans, data protection, data curation, long-term preservation). In addition, of the 353 respondents who currently do not provide training in RDM, more than half (184) plan to provide it in the future, so this area should be also covered by guidelines and tools.

Among 21 offered concepts of importance for open data and OS, the respondents in the 'support' role (typically librarians, repository and service providers) had a high valuation of guidelines and best practices as well as national, funder and institutional policies (65% respondents rated them as 'very important', fourth and fifth place out of 21). FAIRness assessment and repository monitoring tools were less popular (57%, 9/21), while data validation, quantitative indicators and metrics, decision-making tools and quantitative indicators and metrics were ranked very low (46%, 45%, 43% and 41%, positions 17-20). Therefore, NI4OS-Europe should use policies and guidelines as an instrument of change and promote ones that emphasise and advocate the less popular concepts.

Additionally, the landscaping survey contained 12 open-ended questions. In answers to those questions, the respondents often mentioned various documents, products, tools and services. Textual input was analysed and prominent terms, phrases and links were extracted. After removing obviously irrelevant items, the remaining ones were described (if recognised), searched for, inspected and classified. Many well-known and relevant names and resources were found, but quite a few were of limited or personal significance or were mentioned due to a particular perspective and insufficient awareness on the topic. There were also URLs that pointed to a different version of the same document or several elements of the same collection. The final purged list consists of 440 items that were grouped into the following working categories: Guidelines (63), Specifications (9), Certifications (7), Services (40), Software (15), Project or organisations (24), PIDs and related service and organisations (16), Concepts (24), Institutional repositories (20), Local resources and solutions (28), Other related items (10), Potentially relevant URLs (184).

Some items had to be additionally studied in order to identify the most suitable category. Tools were classified as Services or Software, depending on how they were designed to be used. The last three categories have been only coarsely classified, as they include URLs of various resources in local languages and without sufficient information or external references, or particular solutions developed by respondents. All these items were not of importance for this deliverable but could be revisited and reclassified as needed.

The resulting list provides a baseline in terms of the knowledge and expectations of survey participants. Some of the items from the survey coincide with the instruments identified through the independent desk review and search of catalogues, registries and outputs of other related projects. Other items listed in this document that would not have been found if they had not been pointed out by the participants of the survey are the UK DRAMBORA and TRAC, DSA–WDS catalogue of procedures for certification of repositories, GEDE-RDA report on PIDs Consolidated assertions, RDA & CODATA legal interoperability guidelines, ERC guidelines on implementation of OA in projects, EIFL resources, Companion to the Enabling FAIR Data author guidelines and Research Object explanations on RO-Crate specifications.

2.5. Data protection

WP4 itself will not collect any sensitive information. In general, WP4 selects, curates, adopts and harmonises technical solutions and procedures to be used or deployed by the community and within the infrastructure. The FAIR aspect does not imply opening any data, including personal data, but it allows identifying personal data placeholders within metadata schemes and personal data within data objects. The Open aspect, which is more related to actual (personal) data protection, will be addressed by corresponding legal and policy tools.

However, some processes described in the guidelines on the data management lifecycle or uses of tools and certification schemes may lead to the processing of sensitive data. The tools that will be endorsed, adapted or developed will primarily process already shared data and metadata or those that are about to be shared. The vast majority of FAIR and ORDM instruments, by design, does not capture and process any sensitive information. A possible exception to this could be the externally hosted tools assisting in data anonymisation by actually accessing the data that is to be anonymised.

In some instances, tools could be used by more restrictive or closed groups by instantiating them in settings controlled by these groups. For example, some operational tools could be used by WP5 or service providers to assist the transition and integration of repositories and services into the wider open data infrastructure and their registration in core registries. Even internally used tools should not require any excess privileges and should be open-source so that users could check and track how they process potentially sensitive data and settings.

If a tool or service collects personal data, which may be the case with certification platforms, any potentially suspicious and unjustified personal data processing will be assessed by WP4 and result in the elaboration of identified privacy or security concerns or exclusion of the tool from the WP4 catalogue, or a record about the resolution of the concern. WP4 will base its conclusion and recommendation on the insight into data that is being collected, tool’s documentation and its data privacy notice. Processing of personal

information will be scrutinised if the tool or service is hosted in a non-secure country in terms of GDPR.

2.6. Maturity models

Maturity models are widely applied managerial instruments used for the improvement of practices and processes and may be applied to organisations, projects or services. They identify the objectives relevant for domains of their application and measure the performance of their target entities concerning the objectives. A maturity model usually serves to identify the key elements that help to conduct the work, mark out the improvement directions and provide a method to evaluate the maturity.

The RDA FAIR Data Maturity Model WG is developing its maturity model [8] that is primarily intended not for the evaluation of repositories but the FAIRness of data resources. When this set of indicators was offered for testing and comments, the NI4OS-Europe WP4 team assessed and commented in the context of the University of Belgrade 15 publication repositories. It was subsequently invited to subscribe to the WG and contribute; a WP4 representative joined it. The draft indicators proposed by the model help in the reflection on important aspects of repository maintenance, operation and use, and provide a platform for discussing them between service providers, customers, technical teams, data creators and consumers. However, direct application of the model indicators to publication repositories is difficult, as they are associated with a slightly different terminology and conceptual framework. Offering an adaptation and how-to guide for such a large group of repositories would be extremely beneficial. This communication will continue, particularly as this model and its indicators in their current form are not supposed to be directly applied on data sets, but to serve a lingua-franca across emerging evaluation approaches and tools so that their results could be compared. Its use should be adapted to the needs, capabilities and terminology of specific communities, possibly introducing explicit maturity levels and objective-specific improvement recommendations for achieving them.

Other more specific FAIR and ORDM related maturity models may be developed soon. Also, elements of OpenAIRE compliance could be also mapped. Given the popularity and practicality of the maturity model concept, NI4OS-Europe should continue to observe this field and contribute where possible.

2.7. Instrument customisation and development

The selection and analysis of tools may reveal an acute or emergent need not covered by already existing tools may require the development of a novel tool. Work in this deliverable will, thus, ultimately support the development of tools as part of the activities in T4.3 and T4.4. For example, some manual steps or requirements of repository configuration and onboarding could be partially automated, verified or debugged with a scripted tool based on preexisting components. Such a tool could check how a repository is presenting itself, whether it meets harvesting requirements (like those additionally imposed by OpenAIRE), its availability and performance or if it is present in the relevant registries. Additional development directions of the NI4OS-Europe tools may include the focus on FAIR

repository assessment (rather than data), or the delivery of a tool that supports legal and IPR clearance.

For the development of NI4OS-Europe tools, the following criteria will be considered:

- The addressed gap and related practical need should be evident and clear.
- The solution should provide clear benefits to NI4OS-Europe researchers, repository managers or operators of integrative services.
- The solution should have a clear potential to be relevant and useful to other open science incentives or infrastructures.
- The required effort and expertise should be realistic and achievable.
- Ideally, the endeavour could be implemented in incremental steps and with quick wins.

3. Selected guidelines and tools

To classify the selected instruments, the guidelines, models and tools are grouped into thematic subsections, each containing a summarising description of its topic and, quite often, suggestions for the NI4OS-Europe audience. The lists that follow can be split further into more specific subsets as needed. This organisation helps in identifying where to look when searching for a relevant guideline or tool, but also in deciding where to place a newly added instrument and in matching and comparing it with the related ones.

Here presented structure is based on:

- The highlights and strategic recommendations report from the EC expert group report and plan [2];
- The conducted desk research;
- Information obtained from the NI4OS-Europe community through the landscaping survey conducted by WP2;
- Preliminary thematic descriptors assigned to the instruments captured in the WP4 Tools Catalogue;
- Iterative regrouping of the chosen instruments in order to maximise congruence.

The offered enumeration is not taxonomic but practical; in a hierarchical approach, the overall policy and approach would incorporate FAIR, which, within the accessibility principle, would include the use of metadata, standard protocols and access control. On the other hand, the provided flat structure is mostly aligned with subjects and levels of individual guidelines. Individual instruments are therefore placed within the most specific and closely matchings subsection.

As ORDM, FAIR, OS and the related concerns are interconnected in different ways, this organisation could be realigned with evolving needs and perception, but it has been extremely instrumental in the preparation of this document.

These topics are of common interest for all projects supporting EOSC setup and governance. They are, thus, addressed in the cross-project collaboration established and included in the Joint Activity Plan drafted and signed by all INFRAEOSC-5 projects. Two of the Task Forces, the FAIR TF and the Service Onboarding TF, have included in their work plans activities such as:

- Review and standardization of FAIR policies to enable FAIR-aligned services and repositories;
- Analysis of FAIR guidelines and standards to enable adoption in the light of the local and disciplinary context;
- Valuation and development of FAIR technical aspects including metadata frameworks, interoperability, service infrastructure, FAIR software;
- Models for interoperability for the regional/thematic service catalogues into the EOSC central one (i.e. the catalogue hosted by the EOSC portal);
- Impact on validation models, the different aspects and how this can be applied in a distributed environment.

3.1. Overall policies and approaches

At the research and institutional levels, the key interventions are the enhancement of skills for open research data science, open research data management (ORDM) and data stewardship. This also includes engagement with specific communities. The advancement at the infrastructural level includes the increased provision and professionalisation of data stewardship, data repositories and data services, which can be supported through the training of staff and certification of repositories, services and processes. These three are supported by various work packages of NI4OS-Europe, by various guidelines, tools and other instruments referred to or provided by WP4, with a special focus on addressing these needs at the national level.

Many general guidelines and policy-related tools are available in the OS community. However, it is worth the effort to select and recommend those that support a strategic approach but can also quickly benefit the policies and implementation of FAIR, practices, repositories and services.

The effort on FAIR and ORDM can be easily articulated as actionable and practical, but it should also support and assist stewardship, long-term commitment and establishment of sustainable funding and governance. This is done by offering a clear value proposition for research and organisations and by covering both data producers and providers. It is good to prioritize initiatives and developments that bring strong benefits to particular communities or moderate benefits to larger groups, such as application of practices and standards that address common needs and situations or facilitate interdisciplinary research. In a quickly developing and unsaturated field such as open science, strong benefits can be often achieved without a major investment or high level of technical expertise.

From the investment perspective, the early question about the selection of the most useful services quickly turns into the question about the next intervention that could bring the most value for the spent money and effort.

Small and low-effort actions may bring immediate benefits and visibility and strongly enhance the participation and interest. For example, while the adoption of the national OS policy has speed-up the establishment of institutional publication repositories in Serbia, the adjustments towards better harvesting by Google Scholar have improved visibility and motivated researchers to upload their research outputs into institutional repositories.

Skills and capacity building can be evoked by various incentives, like those described in D4.1. Different roles require different expertise. Researchers' data skills are the source which creates data, elevates expectations and creates the pressure to advance open science. At the same time, those with a research background, information professionals and librarians are the base for recruitment of data science and information experts and data stewards. But a clear path for a transition to data professionals should be offered to them. Both formal and informal learning paths should be offered, including university curricula, staff exchanges, fellowships, etc. On the other hand, there is a gap between researchers, their communities and curators, and IT staff, infrastructure professionals and service providers, which could be tackled by embedding professionals into projects, establishing dedicated groups within data service providers or supporting entities, or by creating centres of excellence. This should be additionally supported by the establishment of policies at the founder level, and the inclusion of RDM in research assessment.

In NI4OS-Europe context, the national RDA Nodes in EU countries could act as anchor points in these developments. Another set of coordinating bodies are national open science initiatives, once they are established in most countries of the region.

Costs can be optimised by sharing efforts and resources. This can be done by working towards the establishment and reuse of shared or replicated common components. With multiple instances of similar components, such as repositories, this can be done by using containers, virtualisation and cloud computing, sharing storage, and automating tasks that are repeated across instances (administration, platform and software management, data curation). On a large scale, optimisation is achieved by federating infrastructural services and registries. Reuse and sharing of resources are also supported by the creation or use of registries on top of the FAIR ecosystem, but also with catalogues specialised for various types of services.

Adoption of shared and federated basic services also reduces the effort and increases reuse. Examples for this are academic and GÉANT network connectivity and authentication and authorization by federated national or international AAI services.

Sharing, reuse, interoperability, automation and scaling within and across communities and infrastructures are made possible by the application of common and open standards, specifications and APIs. They are best established, agreed upon and maintained in collaboration and by using consensus-making mechanisms for coordination.

In the following, we present a non-exhaustive list of instruments supporting OS policy development and implementation. The listed items have been included based on their "policy enforcement potential" (e.g. EC generated instruments) or their adoption and appreciation level by scientific communities.

3.1.1. Instruments: Policy and approach

- H2020 Online Manual: Data management, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- FAIRsharing Policies – Catalogue of data preservation, management and sharing policies from international funding agencies, regulators and journals, https://fairsharing.org/policies/
- Guidelines on Implementation of Open Access to Scientific Publications and Research Data in Projects (ERC – European Research Council), https://erc.europa.eu/content/guidelines-implementation-open-access-scientific-publications-and-research-data-projects
- HowOpenIsIt? Guide to Research Funder Policies helping funders to establish criteria for the level of OA required for their policies and mandates, http://www.orfg.org/policy-development-guide
- LEARN Toolkit of Best Practice for Research Data Management – Set of case studies on best practices, advocacy strategies, policy development, and implementation issues, and a model policy template for RDM, http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf
- LIBER Open Science Roadmap – LIBER community translated the OSPP (Open Science Policy Platform) recommendations for libraries combined with

suggestions drawn from expertise and experiences, https://zenodo.org/record/1303002

- OpenDOAR Policy Tool helps repository administrators to formulate and/or present their repository's policies based on current industry standards, http://sherpa.ac.uk/policytool/
- PASTEUR4OA – Essential materials to policymakers on basic OA concepts, http://pasteur4oa.eu/resources
- Rainbow of Open Science Practices – presentation of 17 OS practices throughout the whole research workflow, focusing on the openness of workflows, https://zenodo.org/record/1147025
- RDA Practical Policy WG Policy Templates – Collected and categorised policies of data production systems, https://www.rd-alliance.org/practicalpolicyoutcomespolicytemplates-v2.html
- RISE – The Research Infrastructure Self-Evaluation framework is a benchmarking tool facilitating RDM service planning and development at the institutional level, http://www.dcc.ac.uk/resources/how-guides/RISE
- Toolkit on Public Engagement with Science – resources which help users to understand different aspects of public engagement, https://toolkit.pe2020.eu/
- To watch: Output from the EOSC RoP (The Rules Of Participation) WG: Draft Rules of Participation, https://repository.eoscsecretariat.eu/index.php/s/QWd7tZ7xSWJsesn
- Transparency and Openness Promotion Guidelines aiming to promote the adoption of journal policies furthering transparency, open sharing, and reproducibility, https://cos.io/top/
- Turning FAIR into Reality – EC expert group report on FAIR data with recommendations, http://doi.org/10.2777/1524, [2]
- WHO/ITU National eHealth Strategy Toolkit elaborates many general approaches and concepts that could be applied in the development of national OS strategies, https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-E_HEALTH.05-2012-PDF-E.pdf

3.1.2. Instruments: Registries

- FAIRsharing Databases – Catalogue of databases, described according to the BioDBcore guidelines, along with the standards used within them, https://fairsharing.org/databases
- FOSTER Open Science Resources – Collection of resources and information about OS structured by the FOSTER taxonomy, https://www.fosteropenscience.eu/resources
- OpenDOAR – quality-assured global directory OA repositories and their policies, /https://v2.sherpa.ac.uk/opendoar/
- re3data – Global registry of research data repositories from a diverse range of academic disciplines; it can be used to identify the most appropriate data repository, https://www.re3data.org/search

3.1.3. Instruments: Standards, APIs and specifications

- ADA-M Automatable Discovery and Access Matrix – Standardized way to unambiguously represent the conditions related to data discovery and access, https://github.com/ga4gh/ADA-M
- CERIF Description of an OA Policy – Canonical representation of OA strategies and policies at the national level, https://www.sciencedirect.com/science/article/pii/S1877050917303022
- Draft OpenAIRE Guidelines for Other Research Products, https://guidelines-other-products.readthedocs.io/en/latest/
- Draft OpenAIRE Guidelines for Software Repository Manager, https://software-guidelines.readthedocs.io/en/latest/
- FAIRsharing Standards – The standards in FAIRsharing are manually curated from a variety of sources, including BioPortal, MIBBI and the Equator Network, https://fairsharing.org/standards/
- RDA Metadata Standards Catalog provides researchers, research support staff and developers with access to accurate information about metadata standards, https://rdamsc.dcc.ac.uk/
- Research Object – Explanations on how to use RO-Crate specifications, http://www.researchobject.org/specifications/
- SmartAPI – Community-based extension of the OpenAPI specification, which aims to improve the FAIRness of APIs, http://smart-api.info/

3.2. FAIR

FAIR guiding principles [9] outline guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets. They formulate what is needed to allow computational systems to handle data with none or minimal human intervention, but also the attributes needed to enhance reuse by humans. The goal of FAIR is to unlock the potential of data reuse by analysis by humans and machines and integration across a distributed and federated infrastructure. The conformance with FAIR principles is often referred to as FAIRness.

Research outputs and objects that can be FAIR also include code/software, documentation, images, workflows, protocols, models and other digital research objects.

The principles primarily apply to data, but should also extend to the related services. The FAIR ecosystem consists of key data services needed to support FAIR. These services provide persistent identifiers, metadata specifications, repositories, stewardship and actionable policies and data management plans (DMPs). Data services should also be FAIR but they should meet additional criteria supporting organisational capacity to deliver and, in particular, sustain services. Ideally, it should be possible to establish automated workflows between component services.

Publications and journal articles are a special type of data due to well-established history, established channels, the review process and publishing funding models. Scientific papers

and the underlying data are increasingly expected or required to be FAIR and mutually linked.

FAIR is increasingly adopted in traditional disciplines, typically from the initial adoption at sub-disciplinary levels. It is of particular importance in interdisciplinary research areas – as it fosters interdisciplinary interoperability, common standards and brokering across disciplines.

Many standards and practices support FAIR. The essential components of its implementation include policies, DMPs, identifiers, standards and repositories. There are also supporting data-sharing framework (principles and practices, community-agreed data formats, exchange protocols, metadata standards, tools, data infrastructures. Data is preserved in a structured form using common and, if possible, open formats and exchange protocols. FAIR data objects must be accompanied by persistent identifiers, metadata and contextual documentation to enable discovery, citation and reuse. Persistent Identifiers (PIDs) are assigned to data, software, institutions, researchers, funders, projects and instruments. The presence of metadata enables finding, using and citing data. Metadata standards and vocabularies are adopted by relevant research communities; very often, provenance data, soft annotations or textual descriptions are also needed where metadata specifications are not elaborated and precise. Good documentation and contextual information enable other researchers to correctly use and understand shared data. Software, algorithms, and data are also supplied with machine-actionable statements about dependencies and licencing.

There is ongoing work to provide universal mechanisms that would allow implementations of FAIR principles across various scientific domains. However, the particulars of various research domains, communities, and data types cannot be avoided. Interoperability frameworks define community practices that are aligned with the objectives and attitudes of different communities while supporting FAIR across discipline boundaries. The development, refinement and adoption of shared vocabularies, ontologies, metadata specifications and standards for interoperability and reuse at scale are ongoing activities of the highest priority.

Assessability (the ability to judge data reliability, accuracy, quality, competence of creators and whether it meets the needs) is supported by metadata. The presence of provenance information, certificates and records of periodic assessments (for services and data sources) in particular support trust and leads to reusability and actual reuse.

There may be different priorities and rules related to sharing of different categories of data due to scientific, methodological, ethical or economic reasons and norms of particular disciplines. Therefore, clear guidelines on what to prioritise, appraise, select for publishing and preservation, and make FAIR should be established. Therefore, practical guidelines for selection, stewardship, assessability, legal interoperability, timeliness of sharing are of particular interest.

Increased visibility, citation and recognition due to sharing is not a part of FAIR but can be fostered with metadata and DMPs. FAIR does not imply that data and services should be open and freely available. FAIR principles also apply to data that is restricted and internal, as they increase its usability. They can be accommodated even when some level of closeness is necessary; therefore, any data should be as FAIR as possible – this is technically almost always the case, but is not always feasible.

3.2.1. Instruments: FAIR

- Australian Research Data Commons' FAIR data self-assessment tool, https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/
- Data Fairport – Interoperability platform and set of tools allowing data owners to publish their (meta)data and allows data users to search for and access data (subject to licenses), https://www.dtls.nl/fair-data/find-fair-data-tools/, https://www.datafairport.org/
- FAIR principles for findable, accessible, interoperable and re-usable data, https://www.go-fair.org/fair-principles/, https://www.force11.org/fairprinciples, [9]
- FAIRshake – Evaluates the FAIRness of Digital Objects, https://fairshake.cloud/
- RDA Sharing Rewards and Credit (SHARC) Interest Group templates for FAIRness evaluation criteria, https://docs.google.com/spreadsheets/d/1vloqbekIGlqiDwzE9jqZzoaoDCbwYQlxOWbZzIxIYbI/edit#gid=448406479

3.2.2. Instruments: PIDs

- FREYA project outputs on PIDs, including collaboration with RDA on PID requirements, standards and protocols for interoperability, https://www.project-freya.eu/en/resources/project-output
- Group of European Data Experts in RDA (GEDE-RDA): Persistent identifiers: Consolidated assertions, https://www.rd-alliance.org/system/files/PID-report_v6.1_2017-12-13_final.pdf
- To watch: The EOSC FAIR WG and EOSC Architecture WG PID Task Force: draft Persistent Identifier Policy for EOSC, https://doi.org/10.5281/zenodo.3574203

3.3. Open

Open access to data is an extension of the shared research culture that has gradually developed in scientific research for centuries and is now highlighted and endorsed by the new technological, political and economic drivers. It is also promoted by the EC principle "As open as possible, as closed as necessary".

The relationship between FAIR and open and between managed (in terms of RDM) and unmanaged data can be summarised in Table 1, where FAIR, regardless of the selected level of openness, is always preferred:

Data	Not FAIR, may be managed	FAIR, always managed
Closed	Difficult to obtain and use	*Internal or restricted but clear easy to validate or apply downstream*
Open	Free but difficult to use	*Efficient and transparent open science with strong community benefits*

Table 1: Characteristics of FAIR and open data

It should be noted that some disciplines have historically been open to a different degree, often for a good reason. This has been influenced by ethical or commercial concerns and the possible impact of making the information public on research subjects, such as individuals, populations, sites and other sensitive finds and findings, due to security issues, or commercialization potentials or plans.

The reluctance to share may be caused by the cultural context and preferences caused by volatile collective history (as in NI4OS-Europe region), negative experiences from inappropriate data management or common advancement patterns within academic and power hierarchies. It is reflected in phrases such as "security through obscurity", "control of information and access is the leverage of power" and "openness is weakness".

Regardless of the decision whether data should be open, the information about its existence should always be open, as it allows discovery and negotiation on use and prevents pointless replication of work.

With the increasing benefits of open access information reuse and costs and burdens associated with restricting information sharing. Open access is increasingly adopted by researchers and requested by funders. From the research angle, it is crucial to find a balance between giving data creators the time to produce results and obtain the initial benefits (which are biggest early), and those of sharing with the community (which increase with time).

Data can be expensive to produce but is increasingly inexpensive to share, making reuse more feasible and desirable. Charge to end-users should be avoided, although open does not necessarily mean that the access is free (as in "free speech" and not as in "free beer"). In other words, a proportional fee needed to provide access may be required, but it should not impede it. Collection of such fees is technically achievable but with current payment technologies, it may cost more then what is needed to provide access and share.

It is possible to open data in a controlled way while respecting restrictions. Methods to achieve this that may be appropriate for some situations include getting the consent, defining embargo periods, anonymisation, pseudonymisation, data aggregation, data bucketing, blurring and masking, removal of sensitive attributes or controlling access to them, or by making protective arrangements through data-sharing agreements or using safe data havens. It is sometimes better to control and restrict access to data than to reduce its usability, informative content or resolution.

3.3.1. Instruments: Open

- Companion to the Enabling FAIR Data – Commitment Statement and Author Guidelines of signatory publishers, http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/

- FORCE11 Decision Trees – Support framework for decisions on how to comply with different OS principles and requirements, https://www.force11.org/group/scholarly-commons-working-group/wp3decision-trees
- H2020 Online Manual: Open access, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
- HowOpenIsIt? A Guide for Evaluating the Openness of Journals with OA terminology, https://sparcopen.org/our-work/howopenisit/
- Open Access Toolkit – The University of Western Australia step-by-step guide for researchers and research managers in navigating the OA procedures and requirements, http://guides.library.uwa.edu.au/openaccesstoolkit
- Open Data Certificate – Free Open Data Institute online tool to assess and recognise the sustainable publication of quality open data, http://certificates.theodi.org/en/
- Pathways to Open Access – Report analysing various approaches for achieving OA (Green, Gold-APC, Gold-non-APC), and the actionable strategies that exist to implement each approach, https://libraries.universityofcalifornia.edu/about/initiatives/scholarly-communication
- SHERPA FACT checks if compliance with funder OA policies can be achieved with a particular journal, http://sherpa.ac.uk/fact/
- SHERPA Juliet lets researchers and librarians see funders' conditions for OA, publication and data archiving, https://v2.sherpa.ac.uk/juliet/
- SHERPA REF – Service helping authors and institutions decide whether a journal allows compliance with the UK Research Excellence Framework policy for OA, https://ref.sherpa.ac.uk/
- SHERPA RoMEO enables researchers and librarians to see publishers' per journal conditions on OA and self-archiving, http://sherpa.ac.uk/romeo/index.php
- Wiley Author Compliance helps authors, research managers, and librarians assess which Wiley journals comply with the OA policies of different funders and/or institutions, https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/author-compliance-tool.html

3.4. Overall ORDM and DMPs

Scientific data should be managed throughout the research data lifecycle. If this management is done in an ad-hoc manner, the outcome of this process will be inconsistent, unreliable and of low quality. Instead, it should be decided in advance by defining a DMP at the research, project, organisation or programme level. Data management activities and related documents include the following elements:

- Data management planning
- Creating and capturing

- Storing for immediate or short-term use
- Documenting with contextual information needed for interpretation and use, structure, formats, limitations and provenance
- Backup
- Accessing, viewing, analysing, processing or using otherwise
- Selecting and deciding what to keep
- Determining licencing, citation and distribution rules and conditions
- Sharing with others and making it easily reusable
- Storing, describing and referencing in digital repositories
- Re-using by internal or external users
- Preservation and archiving in a long-term storage
- Data culling, deletion or even data sanitation

The data management plan (DMP) is a document covering the key concerns of FAIR and ORDM that outlines what data will be created and how, how it will be managed and plans for data sharing and preservation. It covers all outputs and instructs researchers how to make informed decisions, anticipate and avoid problems and duplication, data loss and security breaches. DMPs increase quality and save time and effort by helping in documenting the research, tracking provenance, assuring validation and supporting reproducibility. Their application ensures data is accurate, complete, reliable, secure and suitable for further analysis and secondary uses.

Five common themes

- Description of data to be collected or created in terms of collection method, content, type, format, volume and temporal or spatial resolution
- Required documentation and metadata, applied standards, formats, naming rules and conventions
- Ethical concerns and ways to address them, IPR and restrictions on data sharing due to privacy, security, data subjects or research objects protection, commercial and other goals
- Plans for data sharing and access in terms of technical means, dynamics and recipients
- Strategy for long-term preservation or handover

If appropriate for the target audience, DMPs may elaborate in greater detail on:

- Dealing with scripts, logs, software, images, and workflows
- Early curation and sharing of research artifacts
- Use of exchange protocols
- Detailed instructions or rules on free use, modification and sharing for various purposes
- Reporting, accountability, transparency, retention and other regulatory requirements

- Responsibilities for different aspects of data management or phases of the data lifecycle

DMPs should be short and simple but specific, built through consultation and collaboration process, based on available skills and support, and feasible. The plans and procedures should be established early on for consistency, but should also change and evolve as needed.

Since researchers often do not possess the inclination and time to write and maintain DMPs and prowess associated with their varied aspects, development and composition of DMPs can be greatly assisted with templates, often supported with guides and tools and that integrate explanatory materials, simplify decision making and include checklists.

Finally, FAIR, creation of DMPs, research data management and long-term data preservation require work and resources and create costs. These resources, work and costs must be addressed in the research plan. Although FAIR does not deal with long-term preservation, DMPs should ensure sustainability and access to the preserved data. At least, the metadata of the data that is not preserved should be retained, as well as an indication of that decision.

3.4.1. Instruments: ORDM and DMPs

- ARGOS – Service that simplifies the management, validation, monitoring and maintenance of DMPs, https://opendmp.eu/home
- B2SAFE Data Policy Manager tool – Provides DPM functionality by letting data managers define data policies, store and share them with data service providers, service integrators and service administrators, https://eudat.eu/news/a-new-feature-for-b2safe-the-data-policy-manager-dpm-tool
- DMPOnline – Service for creating, reviewing and sharing data management plans that meet institutional and funder requirements, https://dmponline.dcc.ac.uk/
- DMP OPIDoR guides through the drafting and implementation into the practice of data or software management plans, https://opidor-preprod.inist.fr/
- easyDMP – Tool that enables researchers with minimal experience in data management a simple way of creating a data management plan, https://easydmp.eudat.eu/, https://easydmp.sigma2.no/
- Evaluate your RDM Offering aims to help institutions to develop a strategy for improved RDM policy and service infrastructure, https://sparceurope.org/evaluate-your-rdm-offering/
- Guidelines on Data Management in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Horizon 2020 Online Manual: Open Access, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm]
- Open access policies and requirements in Horizon 2020, https://ec.europa.eu/research/mariecurieactions/sites/mariecurie2/files/05_open_access_itn_2018.pdf

- Practical Guide to the International Alignment of Research Data Management https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdfAccess control
- Practical Guide to the International Alignment of Research Data Management – Support for drafting, structuring and using DMPs, https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf
- RDA DMP Common Standard for Machine-actionable Data Management Plans, https://www.rd-alliance.org/system/files/RDA_DMP_Common_Standard_Recommendation_20191202.pdf
- RDMO – Research Data Management Organiser supports planning, implementation and organization of RDM, https://rdmorganiser.github.io/en/
- Research Data Management Toolkit – University of Western Australia institutional step-by-step guide for researchers and research managers on RDM procedures and requirements, http://guides.library.uwa.edu.au/RDMtoolkit
- The realities of Research Data Management – Four-part series exploring how four research universities in the US, UK, Netherlands and Australia are addressing the RDM throughout the research lifecycle, https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html

3.5. Access control

Access to data may need to be limited depending on the consent of data subjects, intra-institutional arrangements, identification of accessors and their institutional or project affiliation, country of data consumer, as in case of GDPR cross-country stipulations, or conditioned by NDAs. In some situations, limitation of access can be avoided by applying anonymisation and related techniques mentioned in Section 3.3 or by preparing for a specific unrestricted use.

Reliable identity and authorization are technical prerequisites of access control and are crucial even in open data scenarios, as individuals and their credentials must be tracked during data creation and modification and to log usage. They are required to enforce policies and bring reliability into the tracing of data creators and consumers, as well as for those in various management roles – service operators, administrators, professional data managers and service and software developers. With them, users also need assurance that they got the resource they requested from the authoritative/targeted source.

The identification and authorization of subjects are increasingly delegated to Authentication and Authorization Infrastructures (AAI) through services and solutions such as those developed and provided by the GÉANT, AARC and its AARC Engagement Group for Infrastructures (AEGIS). All NI4OS-Europe partner countries except Bosnia and Herzegovina are members of GÉANT, and therefore may rely on the service provided by it. While network-related services are relatively transparent for the application layer in the focus of NI4OS-Europe (as long they provide adequate performance), other GÉANT services, particularly those in the domain of trust and identity (such as eduGAIN,

eduTEAMS and other members of the eduGAIN family) may have a significant impact on the usage and accessibility of services provided within NI4OS-Europe. Seamless access to federated identity management and fine-grained attestation of participation in teams or specific organisational and project roles would greatly streamline the management and use of repositories, tools and other services.

The solutions and services such as eduGAIN are being increasingly adopted by end-users and service providers, and are therefore an additional potential contributor to adoption and accessibility of NI4OS-Europe services. In other words, NI4OS-Europe should integrate the existing federated identity management into its regional ecosystem wherever possible, by following the related technical guidelines and implementing all specifications and formal preconditions required for the AAI integration and interoperability.

3.5.1. Instruments: Access control

- AARC Guidelines, https://aarc-project.eu/guidelines/
- eduGAIN guidelines and instructions, https://wiki.geant.org/display/eduGAIN/Guides+and+Instructions
- EGI CheckIn, https://www.egi.eu/services/check-in/

3.6. IPR

Besides technical, the IPR interoperability is also required for reuse. To use an existing data set, it is necessary to know the license conditions and limitations of use, who holds the copyright and how to obtain the needed permissions. It is better to learn this early on in the research process to plan and design the research according to the already available inputs and thus avoid wrong assumptions or need for subsequent IPR handling.

The produced data set should be made open as possible and in line with provisions set or indicated in the DMP. Selecting the most permissive open license is particularly important when the data set is combined with other data sets. In such cases, the most restrictive license has to be applied to the final data set to achieve licensing interoperability. For example, just one of the used sources may prohibit commercial exploitation or require the inheritance of its license for the entire derived work, effectively restricting its openness or applicability. This may be an intended consequence for enforcing downstream sharing, or a side effect limiting actual use.

The most popular open licenses that enable the free distribution of copyrighted work are Creative Commons (CC) licenses. They offer various options for sharing and reuse of the licensed work.

- BY (attribution) clause requires licensees to give the author or licensor the credits. It is present in all CC licenses since version 2.0, requiring all direct or indirect derivative work to acknowledge all contributors, which leads to attribution stacking.
- SA (Share-alike, copyleft) requires distribution of derivative works under a license identical that is not more restrictive to the one that governs the original work. This prevents the licensed data from being combined with data released under a more restrictive or a different copyleft licence.

- NC (Non-commercial) licenses limit reuse and derivative work to non-commercial uses only and opens an ambiguity in interpretation whether a use is commercial or not.
- ND (Non-derivative) clauses limit shared reuse to verbatim copies of the work. But since CC version 4.0, it is possible to produce derivative works without sharing.

As some of these options are cumulative and transitive in their limitations, the least restrictive CC BY and CC0 licenses are usually recommended for data. Besides, research data is sometimes combined with the open data from government or public institutions which are under licenses such as the open government licenses for Germany, UK or Canada.

Automation and automated handling of IPR can be enhanced by the incorporation of machine-readable IPR information. This also facilitates validation and assessment of repositories and data sets.

3.6.1. Instruments: IPR

- Creative Commons – provides Creative Common licence and public domain tool, https://creativecommons.org/licenses/
- Data licence Germany – attribution – version 2.0, https://www.govdata.de/dl-de/by-2-0
- Data licence Germany – Zero – Version 2.0, https://www.govdata.de/dl-de/zero-2-0
- EIFL resources on licensing, copyright for libraries, OA, library consortium development, free and open-source software (FOSS) for libraries, and public library innovation, https://www.eifl.net/resources
- Open Government Licence – Canada, https://open.canada.ca/en/open-government-licence-canada
- RDA & CODATA Legal Interoperability Of Research Data: Principles And Implementation Guidelines, https://www.rd-alliance.org/rda-codata-legal-interoperability-research-data-principles-and-implementation-guidelines-now
- UK Open Government Licence, https://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/uk-government-licensing-framework/open-government-licence/

3.7. Management of repositories and data sets and certification

Repositories and data sets need to be aligned with the DMP, common practices within the scientific discipline and expectations of users. Besides addressing the needs of data creators and consumers, guidelines and tools should ease technical management, interoperability and integration with other service and e-Infrastructure providers. CoreTrustSeal (CTS), maturity models, checklists and other frameworks to help to self-assess or decide on next actions while setting up or maintaining repositories or data collections. Certified repositories, data services and other components will have better

visibility, relevance and integration at EU and global levels. All these will also help WP3 and WP5 in ORDM and FAIR implementation within repositories.

It is good to incentivise or enforce the use of these specifications and standards where possible. This is particularly the case with data formats, structure, types, vocabularies and metadata for the most prolific scientific domains. Interoperable data and metadata that are provided in line with specifications and standards are essential to data interoperability and reuse. Also, their machine-friendly formats and content make it possible to index, discover, reuse, cite, and use the shared scientific data across domains and on the scale, as such interoperability allows not only researchers but also machines to easily find, access and reuse the data.

3.7.1. Instruments: Management of repositories and data sets and certification

- Amnesia – OpenAIRE data anonymization tool, https://amnesia.openaire.eu/
- CoreTrustSeal offers data repositories a core-level certification based on the DSA-WDS Core Trustworthy Data Repositories Requirements catalogue and procedures, https://www.coretrustseal.org/
- Data Stewardship Wizard (DSW) for data stewards based on the Data Stewardship Knowledge Model covering all stages of the data lifecycle, https://opidor-preprod.inist.fr/
- DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) – Tool for the management of risks associated with digital objects by the digital curator, http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora
- DSA–WDS Partnership WG on Repository Audit and Certification: Catalogue of Common Procedures for certification of repositories, https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html
- OpenAIRE Repository Validator helps a repository manager to validate a data or publication repository, journal, archive or aggregator and register it into the OpenAIRE network, https://www.openaire.eu/validator/
- To watch: Output from EOSC FAIR WG Metrics and Certification Task Force on repository certification, https://repository.eoscsecretariat.eu/index.php/s/zCnHTcytBHaLjRp
- TRAC – Trustworthy Repositories Audit & Certification, http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories

3.8. Specialised repositories

NI4OS-Europe is particularly focused on these three areas:

- Life sciences
- Climate science
- Digital cultural heritage

If the researchers use specialised disciplinary and certified repositories (if they exist), then they can greatly benefit from the expertise of the associated curators and community and domain-aligned stewardship. Besides, individual disciplines have different conventions, expectations and modes of work and sharing, which should be, in general, followed.

3.8.1. Instruments: Specialized repositories

- CLARITY (Integrated Climate Adaptation Service Tools for Improving Resilience Measure Efficiency) DMP, https://zenodo.org/record/1491532
- FAIR-TLC: Metrics to Assess Value of Biomedical Digital Repositories: Response to RFI NOT-OD-16-133 – On the application of the FAIR principles to the evaluation of the value, utility, and impact of biomedical digital repositories, https://zenodo.org/record/203295
- Framework for Discipline-specific Research Data Management – Generic framework and guidance for the implementation of discipline-specific research DMPs and domain data protocols, http://www.scienceeurope.org/media/nsxdyvqn/se_guidance_document_rdmps.pdf
- Materials from a workshop on interoperability of metadata standards in cross-discipline data integration and analysis, considering social, health and environmental sciences and generic metadata standards, https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/433553433/Interoperability+of+Metadata+Standards+in+Cross-Domain+Science+Health+and+Social+Science+Applications
- Science Europe Guidance Document: Presenting a Framework for Discipline-specific Research Data Management, http://www.scienceeurope.org/media/nsxdyvqn/se_guidance_document_rdmps.pdf

3.9. Machine readability and metadata

Structured data formats allow data to bear clear semantics and be adequately reused. Applied formats should be in line with existing standards and suitable specifications, wherever they are available and applicable, which often depends on the discipline. The gaps can be covered through grassroots initiatives such as informal agreements on common data models and shared service APIs. Such developments may emerge as side-effects of research projects, but also from discussions at community conferences and events. Whatever the format is used, it should be associated with the corresponding human and machine-interpretable specification. This may include clear labelling of data items and units for quantitative values or classifiers, but also elaborate schemes for flexible data and data types specifications. If several appropriate specifications are in use, the one that is predominantly used by the corresponding community or research cluster (or is gaining the momentum towards majority) should be used. If a researchers' cluster uses a minority specification, an effort to establish a mapping (if possible) or initiate conversion or convergence process should be made.

It is important to select or define formats and machine-readable common data objects and metadata early, also planning their maintenance and updating, if needed, and curation. Early metadata entry is also an opportunity for the researcher to reflect on the produced object from a different angle and avoid pitfalls that might occur later. A delayed entry is also more likely to result in lost parts of data or metadata, uncaptured subtleties and jeopardised reproducibility. At the end of the research lifecycle, the work on early data may be out of interest or constrained or hastened due to the lack of resources or time. With late data publishing, all benefits of the reuse that would be gained in the meantime are lost.

Both model metadata (for experimental results) and confounder metadata (needed for replication and reproduction of experimental results) should be provided. Information about repositories and data sets' metadata should also be in a machine-readable format so that it could be harvested and interpreted by automated means. It should also include information on allowed purposes, licensing and waivers. This can be extended with machine-actionable DMPs such as DMPonline that enable information exchange across the FAIR ecosystem by providing structured options and choices for projects.

To ensure harvesting, exposure and interoperability of repositories with EOSC and (as the NI4OS-Europe repositories will be onboarded through the OpenAIRE OpenAIRE Content Provider Dashboard web service), they have to follow other OpenAIRE guidelines on metadata for literature repositories, data archives, CRIS managers, software repository managers and other research products. They stipulate that metadata should be exposed via OAI-PMH v2.0 [10] protocol and use DataCite metadata format, assorted properties from Dublin Core, DataCite and OpenAIRE schemes, and adhere to specified mandatory or recommended properties, data encoding and vocabularies. Literature repositories and data repositories should explicitly link the corresponding publications and data sets.

Controlled vocabularies and ontologies for metadata or data that are recommended for use in specific fields should be applied. The existing vocabularies and ontologies should be used whenever possible and their extension or modification should be managed within the corresponding community. Whenever they are applied, their proper use should be validated by corresponding support tools, such as lookup services. Use of specific applied vocabularies and ontologies, metadata specifications and persistent identifiers should be described in corresponding data management plans.

The development of novel and independent vocabularies and ontologies or forking of the existing ones by individual research groups without coordination with a wider community of researchers will not be encouraged, supported or endorsed by NI4OS-Europe. Researchers should be discouraged from taking this path, which is likely to produce significant negative effects in terms of the effort required to maintain these semantic structures, as well as accessibility and acceptance of the data sets that apply them.

Automated metadata harvesting by different service providers over standard protocols allow easier searching for useful data, creation of specialised catalogues and groups, but also use tools for data cleaning and checking of data quality.

3.9.1. Instruments: Machine readability and metadata

- Data Catalog Vocabulary (DCAT) Version 2 – DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogues on the Web, https://www.w3.org/TR/vocab-dcat/

- OpenAIRE Guidelines for CRIS Managers, https://guidelines.openaire.eu/en/latest/cris/index.html
- OpenAIRE Guidelines for Data Archives, https://guidelines.openaire.eu/en/latest/data/index.html
- OpenAIRE Guidelines for Literature Repositories, https://guidelines.openaire.eu/en/latest/literature/index.html
- RDA Data Type Registries WG output, https://zenodo.org/record/1406127 and its prototype Data Type Registry, http://typeregistry.org
- To watch: Draft OpenAIRE Guidelines for Other Research Products, https://guidelines-other-products.readthedocs.io/en/latest/
- To watch: Draft OpenAIRE Guidelines for Software Repository Managers, https://software-guidelines.readthedocs.io/en/latest/

3.10. Monitoring, metrics and quality indicators

Measurement and monitoring enable tracking of different aspects of services or their content in order to enhance quality and optimise delivery. They allow the identification of weak areas so that remedial actions can be taken. The used metrics and indicators must be easy to understand, simple to measure and actionable. Practical and relevant metrics and quality indicators help in improving what is being measured. The key metrics should be continuously monitored to support effective and timely decision-making at the operational and management level.

In OS, the metrics of primary interest for researchers and data managers are oriented on the quantity and quality of content and metadata, FAIRness of data or services, costs or article publication time. When available at the personal level, these metrics can serve as compelling incentives for OS and FAIR. Funders of research are often oriented on the overall openness – how many data sets or publications are open and which licenses are used, trends in OA adoption, the number of funder policies on OA and open data, policies and outputs of supported journals and researchers' behaviour. Metrics can be generic and related to common openness and FAIR criteria or be community-specific ones in order to support interoperability and collaboration within individual research disciplines.

Metrics and measurements could be also used in the regular assessment of citations, reuse and impact and trustworthiness of repositories and data. But such metrics are much harder to produce as they are based on inherent and directly accessible characteristics of observed objects, but on their usage, relationship with other objects and appraisal by the community. Regular peer-reviewed self-assessments are a quite effective way to achieve this without appraisal of external references or employment of a heavyweight certification scheme; on the other hand, having an independent mechanism for measurement of the related metrics is another element supporting trustworthiness, as it makes it much more difficult for those who are running the repository to fabricate the values. These reviews and measurements could be also used to justify costs and direct the future work, ensuring continuous and increasing technical alignment with the national and thematic landscape and thus leading to long-term access.

At the service level, the collection of data extends from simple technology and processes-related monitoring and measurements towards service metrics, which provide end-to-end

measurements related to availability, reliability and performance. In OS-related documents, these metrics are often secondary in comparison to these related to content and metadata.

The purpose of monitoring is to supervise and to continually or periodically check and critically observe a critical metric or indicator to determine its status. Monitoring and measurement in short time intervals are primarily significant in the evaluation of the operational health of specific components of the ecosystem. However, it includes determining what the required quality of service is, as it may heavily depend on the context. Measurements over a specific period could be used as an element of a larger assessment framework for certification of FAIR services.

But even when it comes to data and its quality, the measurement and monitoring should be performed at several levels: whether the repositories are up and running; whether they are harvested regularly by all relevant services; is the number of entries harvested in regular intervals in line with expectation and consistent and, for all tracked sources, coherent across catalogues at different levels; is the amount and quality of metadata consistent within communities and across sources of the same type. The last step may include checking of the validity and consistency of person, project, funder and publication identifiers in metadata and within publications as well as identification of metadata attributes that are frequently missing or have unusual or unexpected values (which may not be wrong, but a consequence of inadequate policy or awareness). When it comes CORE, OpenAIRE and BASE are working on metrics that should be helpful in measuring the utilization of repositories and their content.

It should be not that measurement and monitoring detect the issues and raise alerts; the corrective actions require different guidelines and tools, like those described in Section 3.7 that support detailed infrastructural diagnostics or validation and management of data, metadata and repositories.

Continuous tracking of changes in the FAIRness of datasets or repositories over longer periods can greatly enhance uptake if it is used by funders and policymakers to determine outcomes of their interventions and the next actions. To be useful, such tracking requires collation of statistics, visualisation of trends and advanced analytical capabilities.

While monitoring and measurement enable the gathering of information and foster knowledge, passing this information on to diverse stakeholders requires an effective reporting mechanism to be in place. For reporting purposes, data must be collated, summarised and translated into meaningful language. Information presented in this way will support decision-making and be usable for target groups beyond technical staff. Analysis and reporting should provide suitable information presented in a smart and consistent format, in a consistent and timely fashion in order to allow successful identification of trends and exploration of events or glitches.

It is always good to complement automated and quantitative measures with domain-specific counterparts and subjective observations. This helps in making sure that metrics remain useful and fit for purpose and in avoiding turning them into self-serving goals.

3.10.1. Instruments: Monitoring, metrics and quality indicators

- EC Open Science Monitor – Open Science Monitor – Trends and statistics on OA to publications, open research data, and collaborative research, covering EU member states and selected associated countries, with trends across various

disciplines, https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en

- FAIR Metrics Group work on a FAIR framework with maturity indicators, compliance tests and evaluator tool, https://doi.org/10.1038/s41597-019-0184-5, https://github.com/FAIRMetrics/Metrics
- German Open Access Monitor, https://open-access-monitor.de/#/publications
- How Open is Your Research? A Checklist for Institutions, designed to enable research institutions to assess quickly the openness of their research and teaching outputs, https://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-research-checklist-institutions/
- Metrics Toolkit – Provides an overview and assists users in the selection of appropriate metrics to assess research impact claims, https://www.metrics-toolkit.org/
- Monitor Local – Cloud-based, customised solution for institutions to record and report data relating to the publication of OA outputs by their academics, https://monitor.jisc.ac.uk/local/about/
- Monitor UK – Service to help UK institutions benchmark their spend on article processing charges at a national level, https://monitor.jisc.ac.uk/local/
- The Danish Open Access Indicator monitors how the Danish universities fulfil the targets of the National Strategy for OA, https://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-access/Publications/open-access-barometer
- To watch: OpenAIRE related work in progress – Comprehensive Open Metrics for Repositories, https://pub.uni-bielefeld.de/record/2934581
- To watch: Output from FAIR working group – Metrics for FAIR data, https://repository.eoscsecretariat.eu/index.php/s/C3a5WkpsFHL6GD3
- To watch: Outputs from the forthcoming "FAIRification of Nordic and Baltic data repositories" workshop, https://www.eosc-nordic.eu/events/fairification-of-nordic-and-baltic-data-repositories/

4. Conclusions

As the guidelines and the needs addressed by them are emergent and evolving, it is necessary to define an overall approach for dealing with them. This document describes one such approach but also lists the existing guidelines on the integration of repositories and ORDM/FAIR compliance along with examples of related tools. As such, it is a guide on open science guidelines and tools.

Here listed instruments will be further selected and elaborated in D4.3 in terms of their evaluation, mapping and classification, and cataloguing. The following step will be the subsequent delivery of the selected data management and certification (D4.4, M4.2, and D4.6) and legal, technical and procedural (D4.5 and D4.7) tools. The promotion and implementation of tools, together with work in WP3 and WP5, will establish a supportive environment where the work (that will be also backed by incentives identified in D4.1) will result in value co-creation for researches, research organisations and service and infrastructure providers.

By analysing, reusing, delivering and using various ORDM/FAIR instruments, NI4OS-Europe supports: (1) researchers' and stewards' transition to FAIR; (2) data science as the ability to handle, process and analyse data and learn from it; and (3) data stewardship as the set of skill that are needed for managing, sharing and preserving data throughout research lifecycle and beyond. The provided set of guidelines and tools will be part of the platform for a wider regional community that should bring together data experts, domain scientists, interdisciplinary researchers and industry and facilitate dialogue between the relevant stakeholders at various levels.

The existing and emerging guidelines significantly affect relationships between researchers and publishers as well as requirements on data producers and services. While the context of NI4OS-Europe requires the available guidelines and tools to be used to support the expansion of open science in the region, we must also establish an approach that will be able to support the local and emergent needs. Current guidelines are primarily focused on the static and formal aspects of scientific information, but they must increasingly deal with temporal phenomena and public ramifications of research.

For example, the scientific community is currently using the open access and FAIR to support and accelerate the ongoing research on the COVID-19 virus and at the same time using the related health emergency as a device to promote open science and as leverage to overcome some of the established habits and existing barriers. Many publishers, funders, scientific societies and other organisations signed a joint statement [11], requesting the rapid sharing of research data and findings relevant to the novel coronavirus. They call for peer-reviewed research papers relevant to the outbreak to be immediately made available in open access or freely available at least for the duration of the epidemic, and appeal for the clear statements regarding the availability of underlying data, as well as protocols and standards used to collect them.

Today, it is not enough to ensure that the available results are at the highest standard and quickly and effectively shared with the world. Works in progress should be treated as such and not confused with established facts. The papers, data and other results should be traceable, quickly and easily accessible to researchers, adequately scrutinized for reliability and presented in a way that defends the public from unrealistic expectations, false conclusions and misinterpretation by conspiracy theorists and other bad actors.

New guidelines need to be established and existing ones updated in order to help in adapting to specific situations by providing an effective balance between the timely information and highly accurate but late data. They will be based on the reflection on whether the alerts were delayed or shunned and on how to assess, understand and combat the new emergencies while reducing the noise in public space. The guidelines need to assist in protecting from intentional distortions, bias induced by 'scientific influencers', spreading of falsehoods, and other maladies that currently obstruct other areas of human communication and collaboration. Although the scientific method and application of FAIR principles protect the scientific work more than it is the case in other fields of human communication and information use, some historic examples can be easily drawn from various fields of research. Some examples are difficulties associated with findings and initiation of actions affecting corporate or political interests are those related to climate change and health effects of tobacco, leaded gasoline or air pollution, where the traditional approach required decades of combat with powerful stakeholders before the related conclusions and policies could become mainstream. Similar actual phenomena influenced by external drivers include unfiltered and low-quality scientific overproduction, plagiarism and the lasting echoes of the past controversies, such as the one about the MMR vaccine and autism.

Therefore, the guidelines for open science must go beyond their current limits, purists' pickiness or ethical ruminations and become more accessible, practical and actionable then they are now. The scientific community needs new protocols that support fast-track sharing and screening, effective and meaningful interpretation and use of results by domain experts, researchers from other areas and the general public. We must learn and improve from the application of the existing guidelines but also from the current problems.

NI4OS-Europe can contribute to this effort of the scientific community by trying to locally (in South-East Europe) achieve the results of a wider significance. When it comes to open science, what is at stake is not only the participation of the researchers and organisations from NI4OS-Europe countries in the first-class scientific research and collaborations. More importantly, research, data and knowledge must be freed, integrated across different specialisations and aligned with the new technological and societal landscape. Otherwise, at stake are support to the funding of scientific research, the general attitude towards science and relevance of facts and expertise for decision making.

