The Principles of Data Reuse
in Research Infrastructures

Nikolay Skvortsov[0000-0003-3207-4955]

Institute of Informatics Problems of Federal Research Center "Computer Science and Control"
of the Russian Academy of Sciences, Moscow, Russia
nskv@mail.ru

Abstract. The principles known by FAIR abbreviation have been applied for different kinds of data management technologies to support data reuse. In particular, they are important for investigations and development in research infrastructures but applied in significantly different ways. These principles are recognized as prospective since, according to them, data in the context of reuse should be readable and actionable by both humans and machines. The review of solutions for data interoperability and reuse in research infrastructures is presented in the paper. It is shown that conceptual modeling based on formal domain specifications still has good potential for data reuse in research infrastructures. It allows to relate data, methods, and other resources semantically, classify and identify them in the domain, integrate and verify the correctness of data reuse. Infrastructures based on formal domain modeling can make heterogeneous data management and research significantly more effective and automated.

Keywords[1]: FAIR data principles, research infrastructures, data reuse.

1   Introduction

The significant time in most research projects is still spent on detecting the semantics of the data from various sources and solving the problems of their heterogeneity. However, with the current volume and variety of data in science and all the directions of their necessary analysis, it is no longer possible to solve these problems manually. Therefore, today the focus is on the development of technologies for data management and preservation achieving their interoperability and reuse in the research community.

Global research infrastructures are developed, and they aim at supporting the availability of research data. The heterogeneity of data also remains high in them and is resolved mainly by the efforts of researchers during solving scientific problems. Meanwhile, research infrastructure services provide researchers with the necessary information to overcome data heterogeneity. Nevertheless, the scientific community holds active research to minimize the problem of heterogeneity in research infrastructures.

This paper is devoted to a review of approaches used in research infrastructures to support data reuse, taking into account the generally accepted and frequently declared principles of its provision. The potential of conceptual modeling for solving this problem is studied.

The following section describes the principles of providing the possibility of research data reuse, their possible interpretations, and the corresponding directions in research. Section 3 provides an overview of various solutions in research infrastructures following these principles, the advantages and disadvantages of those solutions. Section 4 discusses the opportunities that conceptual modeling based on formal specifications of the research domains can bring to research infrastructures.

2 The Principles of Achieving Data Interoperability and Reuse

During the research, source data being the result of observation or previous studies are cleaned, enriched, and classified, research methods are implemented and applied for generating data containing new knowledge about the research objects. All these results are useful if they can be reproduced or reused in further studies, over different source data, by different research groups. Therefore, in data curation, aiming at managing data during their lifecycle for their optimal long-term usage, the support of accessibility and reuse of data and methods as well as the reproducibility of research results has become an important part. So for many years, these problems have been the subject of intensive research.

Taking into account these objectives, the requirements for methods of long-term preservation and access to data have been formed. It is important to describe data with metadata to determine their semantics and provenance and make it possible to search and select data relevant to the research problem. As well, the source and resulting data should be linked to the tools for their processing and analysis, and methods of the research domain for the possibility to reproduce the result. A good representative of the implementation of these requirements was the WF4Ever project [1] aimed at the reproducibility of research results. Reusable research workflows are at the center of this project. Containers called research objects are formed, in which the workflow specifications are accompanied by data, metadata, services, documentation, links, and access rules. Research communities and projects that have access to research facilities, the ability to publish and search for workflows.

Later, the data management principles known by the abbreviation FAIR were proposed [3]. It was the result of the experience of achieving data interoperability and reuse gained in the scientific domains and discussions in the scientific communities, particu-

larly in the FORCE11 community [2] aimed at electronic publications based on semantic technologies. To be reusable, the data must be uniquely identified, semantically defined using common vocabularies and ontologies, accompanied by the provenance information, must have clear access rules, comply with known protocols and standards, or have known mappings in them. The properties of findability (search capabilities), accessibility, and interoperability (the possibility of functioning together in the system) are the conditions for the main target property of data – the possibility of their reuse (see Table 1). The same requirements are applied to metadata, methods, processes related to data, which are considered as kinds of data too.

Table 1. The Guiding FAIR Data Principles

Findable data	Accessible data
• unique permanent identifiers • extensive metadata description • metadata include data identifiers • data registration or indexing for search	• standard protocol • retrieving by identifiers • authorization and authentification • metadata are accessible even if data are no longer available
Interoperable data	Reusable data
• using a formal knowledge representation language • using FAIR-compliant vocabularies • qualified references to other related data	• defining a plurality of applicability attributes • licensing • provenance information • domain community standards

The relevance of the declared goals, the memorable and intuitive abbreviation, the simplicity of the formulations, the universality and abstractness without imposing certain implementations quickly made the principles popular, and they were supported at the highest level. The principles of FAIR data have already played and continue to play an important role in activating and directing data reuse research. However, the informal definition of these principles has led to quite different interpretations of them, sometimes contradictory or limiting their goals. As a result, further research has become significantly multidirectional:

1. There are proposals for further detailing the data requirements, such as different lists that describe and complicate each of the principles, new beautiful abbreviations that introduce additional informal principles for various domains and goals.
2. The FAIR principles are often considered not as indicating the direction, but as requirements for which existing and new databases and repositories of research data should be evaluated. As a result, evaluation metrics and even certification systems for compliance with the FAIR data principles are proposed.
3. Technologies are selected and interfaces are developed, which are declared to provide implementations of the principles.

4. FAIR data is applied to specific application domains in significantly different ways based on prior experience of data management and often without major changes in approaches.

5. Conclusions are drawn on the consequences of the FAIR data principles for various aspects of data management, in particular, data citation, data management planning, research infrastructure services. In particular, these principles have become an integral basis for the development of large research infrastructures. It is discussed in more detail further in the paper.

Due to the wide variety of interpretations of the FAIR data principles, their authors sometimes had to make efforts to return them to the original intent [4]. In this regard, it is important to emphasize some points. FAIR data can be almost any resource related to data. Metadata, vocabularies and ontologies, interfaces and protocols, method implementations, workflows and programs, data integration procedures and results, citations, documents, and other types of data can be considered as objects for reuse and should meet the FAIR properties. Thus, these principles are not only and not quite about publishing datasets.

Another important aspect of the FAIR data principles is the declared ability to manage data by a machine (machine-actionability), not only by a human. Thus, the principles welcome prospect transition from processing predefined sets of metadata to ensuring the correct interpretation of data and metadata that the machine has not previously worked with. In particular, for this purpose, the importance of formal knowledge representation languages is declared, since they can support reasoning, proving the consistency of descriptions, and verifying the compliance of data semantics.

However, today activities performed on data on the initiative of the machine are hardly achieved. Even automation based on human directives can be a non-trivial problem. While the provision of metadata search and data access can be implemented without difficulties by various approaches, achieving data interoperability meets problems related primarily to the definition of data semantics and the work with heterogeneous data.

Therefore, the principles of FAIR data are recognized as prospective. It is impossible to say that a certain project fulfills the FAIR data principles or does not follow them at all. But the desire to implement them sets the investigation direction towards semantic resolving of data heterogeneity problems, automating data processing, data analysis, and research processes, simplifying reuse of resources of various nature.

Specific implementations of the FAIR data principles are not preset and not assumed as requirements. Nevertheless, certain technologies help to get closer to their implementation. In particular, the authors of the FAIR data principles proposed a set of technologies [5] based on the RDF language as a fundamental data model. Data containers (Linked Data Platform – LDP), a technology for mapping data from different data models (RDF Mapping Language – RML), and a tool for extracting data together with metadata (Triple Pattern Fragments – TPF) are used. As well, the FAIR data principles are sometimes deemed to be related to the technologies of Linked Open Data (LOD), but some similar principles between them don't mean any technological requirements or restrictions.

FAIR Data Point (FDP) [6] has become a widely applied software implementation of the FAIR data principles. It provides creation, storage, and maintenance of metadata and allows publishing and discovering metadata on available digital objects. The metadata must conform to the schemas defined by the domain communities. Such repositories and research infrastructures as EUDAT [7], Zenodo [8], BioPortal [9] use FDP to provide data and resource reuse and integration with other infrastructures that can search resources by metadata.

3 Following the FAIR Data Principles in Research Infrastructures

In research infrastructures designed to support the access to research data and accompany them with data processing and analysis tools, the principles of FAIR data have indeed become the leading ones for choosing the direction of technology development. First of all, research infrastructures provide services for preserving metadata on existing research datasets, searching for them, and solving research problems on the data. Therefore FAIR concepts are mentioned in the fundamental documents of research infrastructures, support of them is declared, and approaches to following these principles are discussed. Currently, the concepts of research infrastructures remain not firmly established, they are actively discussed and evolving. Due to different interpretations of the FAIR data principles and preferring traditional approaches and tools for data processing in certain domains, approaches to achieving the interoperability and reuse of research data may differ significantly.

3.1 Making Data Findable

The basis for providing the search for data in research infrastructures is collecting metadata describing those data and searching through the metadata to retrieve relevant data. Accompanying data with metadata for semantically meaningful search is most often implemented using a fixed set of common attributes. Search for keywords, references to vocabularies, full-text search in text descriptions, and rarely search for typed values are used. For different domains, their own sets of attributes can be chosen, in addition to universal ones. However, these approaches cannot take into account the specific properties of different types of real-world objects, needed to be found. Metadata often describe datasets as a whole and or do not fully describe the content and the structure of data.

In research infrastructures, resource registries are organized for linking metadata to the identifiers of data and retrieving relevant ones as responses to search queries for metadata. The metadata search is usually performed for further manual integration of data resources and solving problems with their involvement. Automated search with subsequent reuse of found resources by the machine is usually not offered, although research communities in some infrastructures strive to achieve such functionalities. This is not real yet primarily due to the informality of the metadata.

The DOI and URI standards are mostly used as resource identifiers. The first one guarantees better persistence of identifiers but is less flexible and intuitive. Global identification with direct access to the necessary data fragments is rarely discussed, access to the resource as a whole is more often implemented. For instance, retrieving by identifiers of data of subsets of attributes, about certain real-world objects, selection of data slices by conditions, and similar approaches to data identification are rarely implemented. Interesting identification approaches are discussed like linking data related to the same real-world entities to the same identifiers for substituting relevant resources.

3.2   Providing Data Access

Data access provision does not cause difficulties when standard Internet protocols are used, including access restriction features. High-level interfaces and protocols like Z39.50 are supported in some research domains where they are commonly used. Metadata archives are usually not provided when if the relevant data are no longer available. Technologies of long-term data preservation providing both efficient and direct access to data and metadata are not much touched upon in the discussions on FAIR data, mostly in research infrastructure architectures.

3.3   Interoperability with Heterogeneous Data

As mentioned before, the main problem of achieving data interoperability and implementing the entire concept of FAIR data is manual approaches to heterogeneous data integration. Automation of the integration of found relevant data and resources in research infrastructures can be possible if the principles of interoperable data become the rules for work with the semantically of data in research communities so that machines are ready to interoperate with data.

The unification of interfaces, high-level protocols, and data access regulations used in research infrastructures mainly solve the problem of data transport functionalities, but do not solve the problems of heterogeneity of the transferred data, or solves it only for certain domain object types. As a solution to the problem of data models heterogeneity, simple general-purpose data formats such as CSV and JSON are often supported. However, even in this case, semantic differences in data as the main cause of the data heterogeneity have been neglected and left to the users of the data for their manual work to understand data. When creating research infrastructures, the problems of data model transformations are almost ignored [10].

Standardization of data representation defining their semantics can be based on different principles in different domains. For example, it can be the standardization of a vocabulary with described domain term semantics and formats of values for them so that any values in the domain are homogeneous. The support of domain-specific data presentation formats and standards allows working with well-defined data semantics known and accepted by the domain communities, but at the same time, the types of entities expressed by them can be severely limited.

It was mentioned above that the FAIR data concepts are often associated with LOD technologies based on the RDF language and that there are some similar LOD requirements. In particular, this is the requirement to create references to other relevant data, which could be borrowed as one of the FAIR data principles. However, in research infrastructures, this principle is understood and implemented not as linking the same of similar data but the data used together, in particular, by creating containers of related resources. Research infrastructures often use the idea of creating containers collecting data, metadata, links to external resources, programs, workflows, documentation, and other resources used together with the data and necessary for their processing and analysis.

To achieve the automation of data processing and data management by machines, the approaches are mostly used that make the autonomous operations but indeed humans predetermine them. Among such approaches there are workflow specifications for step-by-step data processing, initiating activities by events, creating data management plans for the proactive definition of such activities as publishing data, deleting irrelevant data, and others. However, within the framework of implementing the FAIR data principles, there is a desire to implement automation based on the data semantics understanding. For example, the initiative of the Internet FAIR Data and Services (IFDS) makes efforts in this direction [11].

The key principles for achieving the interoperability of heterogeneous data are the use of formal knowledge representation languages and domain vocabularies. Seemingly, it would be natural to use languages of formal ontologies and automatic reasoning over them to implement these principles. However, ontologies are often not even used in some research infrastructures or used only for direct referencing to concepts. What semantic problems could be solved using ontologies to provide data reuse in research infrastructures is discussed in Section 4.

It is necessary to give examples of different approaches to solving problems of data semantics in some research infrastructures and repositories. Thus, the most known astronomical data infrastructure CDS [12] solves the problems of data heterogeneity with the requirement for astronomical catalogs being published to use the same standardized attributes with defined domain semantics whenever possible. Using the same attributes in different catalogs, the infrastructure provides specialized interfaces and query languages that can work with several catalogs simultaneously. It often turns out, however, that the authors use common attributes differently and do not describe it in the catalog documentation. Metadata describing catalog structures have a textual, partially structured form. The semantics of attributes are described by UCD descriptors, accepted by the community, but permitting significant ambiguity. In a recently developed research infrastructure, DADI services [12] provide access to existing astronomical data repositories including CDS and related services. They declare adherence to the FAIR data principles but do not bring significant development towards semantic approaches. The practice of applying existing services to process astronomical data shows that the semantic heterogeneity of data remains significant, and solving the problems of their heterogeneity requires significant efforts in almost every research project.

In the biomedicine domain, BioPortal is a generally recognized resource [9]. It has been combining many ontologies for a long time. The domain ontologies describe different aspects of the domain and intersect with one other. The need to agree on these ontologies was realized. The bulk of ontologies remain unaffected, and they are just available in the BioPortal repository. Nevertheless, with the experience of solving different research problems in the domain, the backbone of ontologies has emerged and they were reconciled with each other for those problem solutions. These ontologies unite laboratory studies and data processing in research experiments. Annotation of data and resources in terms of these ontologies is widely used in biomedicine. The services of BioPortal simplify the compilation of semantic annotations. Annotations allow not only references to concepts but expressions in their terms too, it allows to describe more accurately the meaning of the annotated entities. There is a search for relevant concepts through the ontologies in the portal. Thus, it is a good basis for creating services that achieve data interoperability.

3.4   Achieving the Data Reuse

First of all, achieving data reuse in research infrastructures depends on ensuring the search, accessibility, and interoperability of the data discussed above. Beyond that, following the FAIR data principles, the standards of the domain communities should be supported and attributes describing the non-functional properties of the data should be evaluated. The metadata needed to verify the data reusability, rather than to search for data may include a description of the data provenance and the license terms.

In fact, data is rarely provided with their detailed provenance metadata and it is not trivial to get them in research infrastructures. The domain standards, of course, have usually been supported. However, they are often poorly integrated into FAIR data processing tools and add manual work on their integration and adaptation by data consumers.

3.5   The state of the art of data reuse

It is possible to draw general conclusions about implementations of the FAIR data principles and, accordingly, ensuring the data reuse in research infrastructures. The difference between the preferred approaches to data access in different disciplines is obvious. Traditional ways data integration in them have the greatest impact on research infrastructure solutions. Research infrastructures mainly provide the necessary documentation together with the data but solving the problems of data heterogeneity and integrating relevant resources for their analysis takes a significant part of the researchers' efforts. Machine-actionable approaches to data processing and analysis based on their semantics are almost not implemented. However, close attention is paid to all these problems in the investigations of research infrastructures. That allows us to hope for a future qualitative change in the practices of research data reuse.

The European Open Science Cloud (EOSC) research infrastructure is under development [13]. This initiative is tightly related to investigations of data reuse since some authors of the FAIR concepts are engaged in the development of this infrastructure. So

it would be interesting to consider the solutions proposed in this project. The implementations of EOSC services are largely based on the services of existing data infrastructures and the cloud is designed to combine the resources accessible in a set of other research and data infrastructures. First of all, the services of EGI [8] and EUDAT [7] infrastructures are integrated within the EOSC-Hub project [14]. The services implement resource registration, the definition of metadata and annotation of resources with them, the search through metadata and access to resources, transformation, publishing, and long-term preservation of data, and others.

In addition to the implemented services, EOSC is based on solutions that have been developed to implement the principles of FAIR data. In research projects using the cloud, data management plans (DMP) allowing automating project data management should be developed in advance [15] for the entire time of those projects. To link data with relevant resources of different nature, the concept of digital objects (DO) is used [16]. These are containers that are accessible by global identifiers, aggregating data, metadata, provenance information, the standards and formats used, program code, licenses, identifiers, links to external resources, and other resources. Registries are created to store and manipulate these types of resources. A special DOIP protocol has been developed for managing digital objects [17], which defines a set of operations for manipulating them. For data processing automation, it is proposed to split services for researchers and machine agents. Type-driven automatic data processing could be supported. In connection with the EOSC data infrastructure, the Internet of FAIR Data and Services (IFDS) initiative is being developed [11]. It links data with tools and calculations semantically relevant to them. Thus, in research infrastructures, data processing can be performed automatically by relevant tools, based on the data semantics. Such studies are consequences of the guiding principles of FAIR data.

4   The potential of conceptual modeling for data reuse
in research infrastructures

As shown above, the resolution of data heterogeneity and the automation of data processing rise the most serious difficulties for data reuse in research infrastructures due to the difference in data semantics. Therefore, the use of semantic technologies is seen as a natural direction of research in this area. Avoiding the detection of data semantics or leaving it to experts would hardly solve the burning problems of their data. Among the approaches to determining the data semantics are, first of all, ontological and conceptual modeling are used. They can become the basis for solving the problems of data heterogeneity and automation of data processing and analysis. In fact, the possibility and feasibility of using formal conceptual descriptions of data and automation based on conceptual approaches in data infrastructures are declared and investigated, but they are poorly used or avoided in favor of evaluation approaches based on similarity measures. The use of machine learning methods for linking relevant resources is an evaluation approach and they more likely are applicable for recommendations to experts rather than for automated reasoning of data or resource reuse.

To fully apply semantic approaches, it is necessary to formally describe the research domain knowledge: the domain ontologies, common data schemas for the most frequently used domain objects in the selected common data models. Formal ontologies of certain domains should be maintained by the research communities working in those domains.

Description of such domain resources as relevant data, methods, and services requires semantic annotation in terms of domain ontologies. Semantic annotations are the main metadata linking resource identifiers with the domain concepts defined in ontologies. These conceptual metadata are used for semantic search in research infrastructures. For the most common areas of knowledge in community domains, conceptual schemas with structures for different types of domain objects and specifications of their behavior as well as commonly used transformation rules, calculations, and other specifications applied to them are designed.

Metadata are stored in registries based on domain ontologies. Such registries are well-classified collections of domain resources being reused by the communities. The community can collect datasets, data models and formats, protocols, implemented methods, workflows, programs, and other kinds of resources. Semantic descriptions of them and their interfaces in terms of ontologies allow finding data and resources applicable to them. During the registration process, data and other resources are semantically integrated with conceptual data schemas in communities. Data model transformations, schema matching, method and service integration, entity resolution, and data fusion can be included in such a process. After integration, data and related resources can be reused repeatedly in the domain community without additional integration. The results of the integration are reused together with the integrated resources. Thus, the registration of data or other resources, including their integration into the domain of the community, can be performed once.

Research problems can be solved in the data infrastructure by using found relevant registered resources with data, methods, programs. Solving problems using common schemas in communities allows reusing the results of the integration of registered resources. If required integration hadn't been performed earlier, researchers can map the resources to needed common resources or schemas and later complete registries with new metadata. If it is necessary to publish the results of solving problems and to make them accessible in the community, researchers register them using semantic annotation in terms of community ontologies. The same is done with newly used external resources that previously were not found in the registers.

A formal description of ontologies and common conceptual schemes of domains for such a scenario of data infrastructures requires effort. However, firstly, most data-intensive research domains have already had some developments in this direction like domain-specific formats and schemas that can be commonly used and further developed. Secondly, communities working in certain domains are interested in this work. They should not only be users of data infrastructures but collaboratively develop and agree on domain specifications in their communities. Then they can demand commitment to these specifications for the members of the community. It just corresponds to

the principle of supporting the domain community standards. Third, once agreed specifications can be reused many times and be developed and changed with version fixing. It minimizes further efforts and makes research in communities efficient.

The use of ontologies and semantic annotation fully corresponds to the principles of data description with rich metadata. Sufficiently detailed analysis of the essential properties of domain objects leads to deep domain conceptualization and, with a formal approach, to domain ontology creation.

The formality of the ontology allows to classify data about almost any domain object in detail and provides search automation taking into account the semantics of the data. Ontologies allow describing data from several views simultaneously using annotations in terms of different ontologies or logical expressions of ontological concepts in annotations. Support for metadata registries based on ontologies and search in them based on ontology reasoning and query languages implement the principles of data findability.

On the other hand, the use of ontologies and conceptual schemes corresponds to the principles of using formal knowledge representation languages and vocabularies. The use of formal conceptual models allows automatic reasoning. In particular, models expressible in the OWL ontology language support the automatic reasoning for the satisfiability problem and inclusion between their set theory interpretations. Solving these problems is so important for checking the semantic correspondence between different descriptions of resources, searching for resources relevant to the query in terms of ontology, and, therefore, for providing data interoperability and their findability. Automatic reasoning over semantic specifications can become the key to ensuring machine-actionability in research infrastructures. At the same time, a detailed description of data, the semantics of methods, elements of their interfaces, and schema mapping using formal ontologies guarantees the correctness and automatic application of methods on data. Different dialects of the OWL language can be chosen as formal languages of knowledge representation and other languages can be used too. For example, SQL standards with their extensions can be a fairly formal basis for data management.

Creating conceptual data schemas complies with the principle of supporting community domain standards. Conceptual schemes, data formats, high-level protocols, data management standards used in domain communities, should unify the presentation of data and definition of data semantics. Conceptual schemas of domains can define not only the structures and constraints for data representation but the constraints of data manipulation, the behavior of the described objects, and methods related to certain types of entities.

Metadata in terms of ontologies are designed not only to describe data semantics in the domain for their classification and search. Some metadata are necessary to express non-functional aspects of the data description, such as authorship, relevance in time and place, completeness, data quality, license terms necessary for the correct data reuse. The support for specialized ontologies that reflect non-functional characteristics used in the domain communities corresponds to the principle of providing the attributes of data applicability. Such ontologies can be, in particular, the ontology of data quality, the ontology of data provenance. And data can be described in their terms in the same way that is used to describe the domain semantics of data. It is also preferable to use

relationships semantically defined in special ontologies of the domain for making references between different kinds of data used together. To support metadata, when the resource is no longer available, it is necessary to explicitly label them with time relevance information among other provenance metadata. Also, licenses terms must be clear for humans and machines, and their description can be implemented through special ontological descriptions as well.

5   Conclusion

The state of the art of the FAIR data principles implementation in research infrastructures was considered. Used approaches were considered, and formal data specification and automation based on conceptual approaches were at the center of the investigation. These approaches are used insufficiently. An ontological approach with common conceptual schemas for data representation and manipulation was described and evaluated from the view of the FAIR data principles. It allows semantic describing, classifying, and collecting data and related resources used in communities and avoids needed efforts for solving the problems of data heterogeneity. The approach allows reusing semantically integrated data, methods, and other resources as well as the results of their integration and significantly increases the efficiency of research up to the automation of work with heterogeneous data.

References

1. Belhajjame, K. et al.: Workflow-Centric Research Objects: A First Class Citizen in the Scholarly Discourse. Proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web (SePublica2012). P. 1–12 (2012).
2. Improving Future Research Communication and e-Scholarship. Bournea, P. et al. (eds.). The Future of Research Communications and e-Scholarship. 2011. https://www.force11.org/
3. Wilkinson, M.D. et al.: The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3, 160018. – (2016). https://doi.org/10.1038/sdata.2016.18
4. Mons, B. et al. Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the European Open Science Cloud. Information Services & Use 37(1). P. 49–56 (2017). https://doi.org/10.3233/isu-170824
5. Wilkinson, M.D. et al.: Interoperability and FAIRness through a novel combination of Web technologies. PeerJ Computer Science 3, e110 (2017). https://doi.org/10.7717/peerj-cs.110
6. FAIR Data Point design specification. 2019. URL: https://github.com/FAIR-DataTeam/FAIRDataPoint-Spec
7. The EUDAT Collaborative Data Infrastructure. URL: https://www.eudat.eu/
8. EGI Foundation. URL: https://www.egi.eu/
9. Noy, N. et al.: BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research 37(2). P. W170–W173 (2009). https://doi.org/10.1093/nar/gkp440

10. Stupnikov, S., Kalinichenko, L.: Extensible Unifying Data Model Design for Data Integration in FAIR Data Infrastructures. In: Manolopoulos Y., Stupnikov S. (eds) Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2018. Communications in Computer and Information Science 1003, P. 17–36 (2019). Springer, Cham. https://doi.org/10.1007/978-3-030-23584-0_2

11. Mons B. FAIR Science for Social Machines: Let's Share Metadata Knowlets in the Internet of FAIR Data and Services. Data Intelligence 1(1). P. 22–42 (2019). https://doi.org/10.1162/dint_a_00002

12. Genova, F.: Data as a research infrastructure CDS, the Virtual Observatory, astronomy, and beyond. EPJ Web of Conferences 186, 01001 (2018). https://doi.org/10.1051/epjconf/201818601001

13. European Open Science Cloud. URL: https://www.eosc-portal.eu/

14. EOSC-Hub. URL: https://eosc-hub.eu/

15. Guidelines on Data Management in Horizon 2020. Directorate-General for Research & Innovation, EC. 2016. URL: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

16. Wittenburg, P.: From persistent identifiers to digital objects to make data science more efficient. Data Intelligence 1(1). P. 6–21 (2019). https://doi.org/10.1162/dint_a_00004

17. Digital Object Interface Protocol Specification. Ver. 2.0. DONA Foundation. 2018. URL: https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf

18. Regulations of CKP "Informatics". URL: http://www.frccsc.ru/ckp