American Journal of Data Science and Machine Learning
ISSN: - applied

Architecting FAIR Digital Objects and Computational Workflows: Interoperable Metadata, Persistent Identification, and Reproducible Research Infrastructures

[1] Dr. Javier Morales
[1] Department of Information Studies University of Copenhagen Denmark

Received: 15th Oct 2025 | Received Revised Version: 27th Oct 2025 | Accepted: 30th Oct 2025 | Published: 06th Nov 2025

Volume 01 Issue 01 2025 | Crossref DOI: 10.64917/ajdsml/V01I01-002

Abstract

The exponential growth of data intensive research across life sciences, Earth sciences, and computational domains has exposed profound challenges in interoperability, reproducibility, and long term stewardship of digital research assets. The FAIR principles have provided a normative framework for making digital objects Findable, Accessible, Interoperable, and Reusable. However, the operationalization of FAIR at scale requires coordinated infrastructures that integrate persistent identifiers, semantic web standards, research object packaging, workflow provenance capture, and policy aligned governance mechanisms. This article presents a comprehensive theoretical and architectural synthesis of interoperable FAIR digital objects and computational workflows grounded in contemporary specifications and community driven implementations. Drawing on standards such as the Digital Object Interface Protocol specification, PROV O, RDF 1.1, Schema.org, OCFL, BagIt, IEEE 2791 2020, and RO Crate, as well as community platforms including myExperiment, Whole Tale, OpenAIRE, the NCI Genomic Data Commons, and EOSC interoperability frameworks, this study constructs a layered conceptual model for digital object management. The analysis examines metadata modeling through Bioschemas and Science on Schema.org, ontology reuse, machine actionable data management plans, and persistent identifier design patterns. Particular emphasis is placed on computational workflow reproducibility using engines such as Snakemake and Galaxy, software distribution ecosystems such as Bioconda, cross platform packaging in Debian, and provenance frameworks including CWLProv and Pegasus. The article critically interrogates socio technical tensions between researcher usability and stewardship compliance, drawing on debates about data management fatigue and lifestyle oriented FAIR practice. It advances a detailed interoperability architecture integrating digital object identifiers, research object crates, linked data graphs, and repository storage layouts compliant with OCFL and BagIt. Through extensive theoretical elaboration, the paper articulates governance principles for data commons, cross domain metadata harmonization, and standards based international genomic data sharing. The findings demonstrate that sustainable FAIR infrastructures emerge not from isolated tools but from coordinated ecosystems combining persistent identity, semantic richness, workflow transparency, and institutional stewardship cultures. The study concludes by outlining policy, technical, and cultural pathways toward machine actionable, reproducible, and globally interoperable research environments.

Keywords: FAIR digital objects, research objects, computational workflows, interoperability standards, metadata ontologies, persistent identifiers, data commons.

Cite This Article: Dr. Javier Morales. 2025. Architecting FAIR Digital Objects and Computational Workflows: Interoperable Metadata, Persistent Identification, and Reproducible Research Infrastructures. American Journal of Data Science and Machine Learning 1, 01, 6-12. https://doi.org/10.64917/ajdsml/V01I01-002

1. Introduction

The contemporary research landscape is increasingly defined by data intensive methodologies, distributed computational infrastructures, and cross disciplinary collaboration. In genomics, Earth system science, bioinformatics, and digital humanities, the production of data is no longer episodic but continuous, automated, and computationally mediated. This transformation has catalyzed both unprecedented scientific opportunity and systemic fragility. Data may be generated at scale, but without robust metadata, persistent identifiers, interoperable schemas, and reproducible workflows, the long term value of such data rapidly deteriorates. The FAIR principles, introduced to articulate normative goals for data stewardship, emphasize that digital research assets must be Findable, Accessible, Interoperable, and Reusable. However, translating these aspirational principles into concrete infrastructures remains a complex socio technical challenge.

The challenge is not limited to data alone. Increasingly, computational workflows, software dependencies, and provenance records are integral components of research outputs. As argued in discussions of computational reproducibility in life sciences (Gruning et al., 2018) and in the articulation of reproducible research rules (Sandve et al., 2013), scientific claims are inseparable from the code and computational processes that generate them. Yet traditional scholarly communication infrastructures were not designed to preserve executable workflows, software environments, and provenance graphs. The result is a persistent reproducibility crisis that is not merely methodological but infrastructural.

Research object frameworks attempt to address this gap by treating data, code, workflows, metadata, and contextual documentation as integrated digital entities. The RO Crate metadata specifications (Sefton et al., 2019; Sefton et al., 2021) provide a JSON LD based packaging model for aggregating digital research artifacts with machine readable metadata grounded in linked data standards. This model is complemented by tools such as ro crate js and ro crate html js, and by broader packaging approaches such as DataCrate (Sefton et al., 2018). When integrated with persistent identification systems (McMurry et al., 2017), linked data standards (Heath and Bizer, 2011), and provenance ontologies such as PROV O (Lebo et al., 2013), research objects become candidates for durable, interoperable digital objects.

The Digital Object Interface Protocol specification (D Foundation, 2018) introduces a protocol layer for resolving and interacting with digital objects through persistent identifiers. In parallel, data commons architectures emphasize service oriented access to large scale datasets, advocating for data science as a service models (Grossman et al., 2016). The NCI Genomic Data Commons illustrates how large scale biomedical infrastructures integrate metadata, policy compliance, and computational access (Jensen et al., 2017). International coordination initiatives such as GA4GH further demonstrate the necessity of harmonized standards for genomic data sharing across jurisdictions (Rehm et al., 2021).

Despite these advances, significant literature gaps remain. First, many discussions treat metadata standards, workflow reproducibility, identifier systems, and repository storage models as discrete domains. There is insufficient integrative theorization of how these layers interlock into a coherent interoperability architecture. Second, while FAIR principles have been extended to research software (Lamprecht et al., 2019) and computational workflows (Goble et al., 2019), practical guidance often remains fragmented across tools and communities. Third, sociocultural resistance to data management mandates, as reflected in critiques of bureaucratic overhead (Neylon, 2017), indicates that purely technical solutions are inadequate without attention to researcher experience.

This article addresses these gaps by constructing a comprehensive architectural synthesis of FAIR digital objects and computational workflows. Drawing exclusively from established standards, technical reports, and peer reviewed scholarship, the study integrates persistent identification, semantic modeling, workflow provenance, packaging formats, repository storage specifications, and governance frameworks into a layered interoperability model. Rather than summarizing existing tools, the analysis elaborates theoretical interdependencies, counter arguments, and design trade offs in depth. The goal is not merely to catalogue technologies but to articulate a coherent vision for sustainable, machine actionable, and reproducible research infrastructures.

2. Methodology

This research adopts a qualitative integrative synthesis methodology grounded in standards analysis, architectural modeling, and comparative theoretical interpretation. Rather than empirical experimentation, the study systematically examines formal specifications, technical standards, community guidelines, and peer reviewed literature to construct a conceptual interoperability framework. The method proceeds through several analytic stages.

First, foundational digital object and identifier standards are analyzed to clarify the conceptual model of digital objects. The Digital Object Interface Protocol specification (D Foundation, 2018) is examined as a resolution and interaction protocol, while persistent identifier design principles are studied through identifier best practice analyses (McMurry et al., 2017). The purpose of this stage is to define what constitutes a digital object in a FAIR context and how it is persistently referenced.

Second, semantic web and metadata standards are examined to determine how digital objects are described. RDF 1.1 concepts (RDF Working Group, 2014) provide the abstract graph model underlying linked data. Schema.org (Guha et al., 2015) and its domain specific extensions in Bioschemas (Gray et al., 2017) and Science on Schema.org (Jones et al., 2021) are analyzed as pragmatic web scale vocabularies. Ontology reuse theory (Katsumi and Gruninger, 2016) informs evaluation of semantic alignment strategies. Metadata roles in reproducibility are examined through patterns literature (Leipzig et al., 2021) and cautionary accounts of metadata neglect during global crises (Schriml et al., 2020).

Third, research object packaging and repository layout standards are synthesized. RO Crate specifications (Sefton et al., 2019; Sefton et al., 2021) are examined alongside DataCrate (Sefton et al., 2018), BagIt (Kunze et al., 2018), and OCFL (OCFL, 2020). Tools such as ro crate js, ro crate excel, ro crate html js, Describo, and ocfl tools are analyzed as ecosystem components enabling implementation. GitHub repository management guidance on large files (GitHub Docs, 2023) is considered in the context of scalable versioning.

Fourth, computational workflow reproducibility is analyzed through workflow engines, provenance models, and software distribution ecosystems. Snakemake (Koster and Rahmann, 2012), Galaxy, CWLProv (Khan et al., 2019), Pegasus provenance (Kim et al., 2008), Jupyter Notebooks (Kluyver et al., 2016), and cross platform workflow collaboration in Debian (Moller et al., 2017; Moller et al., 2010) are examined. Bioconda as a sustainable distribution channel (Gruning et al., 2018) and IEEE 2791 2020 standards are integrated to understand technical reproducibility.

Fifth, governance and policy frameworks are analyzed. FAIR computational workflow principles (Goble et al., 2019), EOSC interoperability framework (Kurowski et al., 2021), OpenAIRE (Rettberg and Schmidt, 2015), machine actionable data management plan principles (Miksa et al., 2019), and data stewardship models (Mons, 2018) are examined to articulate institutional dimensions.

Throughout these stages, comparative analysis identifies interdependencies, conflicts, and complementarities. Counterfactual reasoning is used to explore implications of missing layers, for example what occurs when identifiers exist without semantic metadata, or when workflows are shared without provenance. The result is a multi layered conceptual architecture linking digital object identity, semantic description, packaging, storage, execution, and governance.

3. Results

The integrative analysis yields a layered interoperability architecture composed of six interdependent strata: identity, description, aggregation, storage, execution, and governance.

At the identity layer, digital objects are defined as entities addressable through persistent identifiers resolvable via standardized protocols. The Digital Object Interface Protocol specification (D Foundation, 2018) provides mechanisms for interacting with digital objects independent of their physical storage location. Persistent identifier design best practices emphasize global uniqueness, resolvability, metadata binding, and long term governance (McMurry et al., 2017). Without such identity infrastructure, digital artifacts remain locally contextual and cannot participate in distributed ecosystems. The analysis reveals that identifier systems are not merely locators but semantic anchors linking objects to metadata graphs and provenance trails.

At the description layer, semantic richness is achieved through linked data representations grounded in RDF 1.1 (RDF Working Group, 2014). Schema.org provides a web scale vocabulary enabling search engine discoverability (Guha et al., 2015). Domain extensions such as Bioschemas extend general schemas to life science contexts (Gray et al., 2017), while Science on Schema.org harmonizes dataset and publication descriptions (Jones et al., 2021). Ontology reuse theory clarifies that semantic interoperability requires disciplined reuse rather than ad hoc vocabulary creation (Katsumi and Gruninger, 2016). Metadata functions not as supplementary documentation but as structural infrastructure enabling reproducibility, as demonstrated by analyses of metadata roles in computational research (Leipzig et al., 2021) and by the severe consequences of metadata neglect during COVID 19 data aggregation (Schriml et al., 2020).

At the aggregation layer, research object packaging unifies heterogeneous components into coherent digital entities. RO Crate metadata specification defines a JSON LD manifest that aggregates files, workflows, software, and contextual documentation (Sefton et al., 2019; Sefton et al., 2021). DataCrate conceptualizes packaging as both distribution and display mechanism (Sefton et al., 2018). The integration of BagIt packaging (Kunze et al., 2018) ensures fixity and transfer reliability, while OCFL defines repository storage layout for versioned object management (OCFL, 2020). Tools such as Describo and ro crate excel demonstrate practical implementation pathways. The result is that aggregation transforms discrete files into semantically coherent digital objects.

At the storage layer, repository architectures ensure long term preservation and version control. OCFL structures object directories with versioned states and sidecar metadata (OCFL, 2020). Git based systems require specialized handling for large files (GitHub Docs, 2023). Integration of OCFL with RO Crate manifests creates a dual model in which semantic metadata coexists with durable file layout. The analysis demonstrates that storage and metadata layers must be coordinated; semantic descriptions referencing unstable file paths undermine persistence.

At the execution layer, computational workflows operationalize research objects. Snakemake provides rule based workflow execution (Koster and Rahmann, 2012), while Galaxy and related tools enable accessible workflow management. CWLProv extends workflow provenance capture using PROV O (Khan et al., 2019; Lebo et al., 2013). Pegasus and Wings systems provide provenance trails (Kim et al., 2008). Jupyter Notebooks enable literate computational publishing (Kluyver et al., 2016). Bioconda ensures consistent software distribution (Gruning et al., 2018), and Debian based community collaboration demonstrates cross platform robustness (Moller et al., 2017). IEEE 2791 2020 formalizes bioinformatics analysis communication standards. The results show that reproducibility depends on alignment between workflow description, software environment capture, and provenance modeling.

At the governance layer, institutional frameworks coordinate technical standards. FAIR computational workflows articulate principles for workflow sharing (Goble et al., 2019). EOSC interoperability framework defines cross European coordination (Kurowski et al., 2021). OpenAIRE integrates repository metadata at continental scale (Rettberg and Schmidt, 2015). Data commons architectures demonstrate service oriented access models (Grossman et al., 2016). The NCI Genomic Data Commons illustrates policy aligned infrastructure (Jensen et al., 2017). Machine actionable data management plans transform narrative compliance into executable metadata (Miksa et al., 2019). Data stewardship frameworks emphasize cultural transformation (Mons, 2018). These governance mechanisms reveal that technical interoperability must be embedded within policy ecosystems.

4. Discussion

The layered architecture reveals several theoretical insights. First, FAIR digital objects are not singular technologies but composite constructs emerging from the alignment of identity, semantics, packaging, storage, execution, and governance. Removing any layer destabilizes the whole. For example, workflows shared without provenance metadata may be executable but not auditable. Identifiers without semantic description are resolvable but not interpretable. Storage without packaging leads to orphaned files devoid of context.

Second, semantic interoperability depends on ontology discipline. Ontology reuse theory cautions against proliferation of incompatible vocabularies (Katsumi and Gruninger, 2016). Schema.org offers pragmatic web alignment, yet domain specificity requires extensions such as Bioschemas (Gray et al., 2017). Balancing generality and specificity remains a persistent tension.

Third, researcher experience must be considered. Critiques of data management burdens highlight cultural resistance (Neylon, 2017). Sefton argues that FAIR data management is a lifestyle rather than a lifecycle, emphasizing embedded practice (Sefton, 2021). Machine actionable data management plans attempt to automate compliance (Miksa et al., 2019), yet automation without usability risks alienation.

Fourth, international data sharing, particularly in genomics, underscores geopolitical dimensions. GA4GH standards illustrate coordination across healthcare systems (Rehm et al., 2021). Persistent identifiers and metadata standards must therefore accommodate regulatory diversity.

Limitations of this study include its reliance on documentary analysis rather than empirical implementation case studies. While the architectural synthesis is comprehensive, real world deployments may encounter unforeseen integration challenges. Future research should empirically evaluate integrated OCFL and RO Crate repositories within operational data commons.

5. Conclusion

The future of reproducible, interoperable, and sustainable research depends on coherent digital object architectures. By integrating persistent identification, linked data semantics, research object packaging, repository storage standards, workflow provenance capture, and governance frameworks, research infrastructures can operationalize FAIR principles at scale. The path forward requires not only technical specification alignment but also cultural commitment to stewardship as intrinsic scientific practice. Through layered interoperability, digital objects become durable, executable, and globally shareable components of an evolving scholarly commons.

References

1. D Foundation. Digital Object Interface Protocol Specification, version 2.0, Technical Report, 2018. https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf.
2. Garcia Silva A., Gomez Perez J.M., Palma R., Krystek M., Mantovani S., Foglini F., Grande V., De Leo F., Salvi S., Trasatti E., Romaniello V., Albani M., Silvagni C., Leone R., Marelli F., Albani S., Lazzarini M., Napier H.J., Glaves H.M., Aldridge T., Meertens C., Boler F., Loescher H.W., Laney C., Genazzio M.A., Crawl D., Altintas I. Enabling FAIR research in Earth science through research objects. Future Generation Computer Systems 98 (2019), 550 to 564.
3. GitHub UTS eResearch ro crate js. Research Object Crate utilities. https://github.com/UTS-eResearch/ro-crate-js.
4. GitHub workflowhub eu galaxy2cwl. Standalone version tool to get cwl descriptions of galaxy workflows. https://github.com/workflowhub-eu/galaxy2cwl.
5. GitHub CoEDL modpdsc. https://github.com/CoEDL/modpdsc.
6. GitHub CoEDL ocfl tools. Tools to process and manipulate an OCFL tree. https://github.com/CoEDL/ocfl-tools.
7. Giving software its due. Nature Methods 16(3) (2019), 207 to 207.
8. Goble C. What Is Reproducibility? The R Brouhaha, Hannover, Germany, 2016.
9. Goble C., Cohen Boulakia S., Soiland Reyes S., Garijo D., Gil Y., Crusoe M.R., Peters K., Schober D. FAIR Computational Workflows. Data Intelligence 2(1 to 2) (2019), 108 to 121.
10. Goble C., Soiland Reyes S., Bacall F., Owen S., Williams A., Eguinoa I., Droesbeke B., Leo S., Pireddu L., Rodriguez Navas L., Fernandez J.M., Capella Gutierrez S., Menager H., Gruning B., Serrano Solano B., Ewels P., Coppens F. Implementing FAIR digital objects in the EOSC life workflow collaboratory. Zenodo (2021).
11. Goble C.A., Bhagat J., Aleksejevs S., Cruickshank D., Michaelides D., Newman D., Borkum M., Bechhofer S., Roos M., Li P., De Roure D. myExperiment: A repository and social network for the sharing of bioinformatics workflows. Nucleic Acids Research 38(Web Server issue) (2010), W677 to W682.
12. Gray A., Goble C., Jimenez R. Bioschemas: From Potato Salad to Protein Annotation. Vienna, Austria, 2017.
13. Grossman R.L., Heath A., Murphy M., Patterson M., Wells W. A case for data commons: Toward data science as a service. Computing in Science and Engineering 18(5) (2016), 10 to 20.
14. Gruning B., Chilton J., Koster J., Dale R., Soranzo N., van den Beek M., Goecks J., Backofen R., Nekrutenko A., Taylor J. Practical computational reproducibility in the life sciences. Cell Systems 6(6) (2018), 631 to 635.
15. Gruning B., Dale R., Sjodin A., Chapman B.A., Rowe J., Tomkins Tinch C.H., Valieris R., Koster J., Bioconda Team. Bioconda: Sustainable and comprehensive software distribution for the life sciences. Nature Methods 15(7) (2018), 475 to 476.
16. Guha R.V., Brickley D., Macbeth S. Schema.org: Evolution of Structured Data on the Web. Queue 13(9) (2015), 10 to 37.
17. Heath T., Bizer C. Linked Data: Evolving the Web into a Global Data Space. 2011.
18. IEEE Standard 2791 2020. IEEE Standard for Bioinformatics Analyses Generated by High Throughput Sequencing to Facilitate Communication. 2020.
19. Jensen M.A., Ferretti V., Grossman R.L., Staudt L.M. The NCI Genomic Data Commons as an engine for precision medicine. Blood 130(4) (2017), 453 to 459.
20. Jones M.B., Richard S., Vieglais D., Shepherd A., Duerr R., Fils D., McGibbney L. Science on Schema.org v1.2.0. 2021.
21. Katsumi M., Gruninger M. What is ontology reuse? Formal Ontology in Information Systems. 2016.
22. Khan F.Z., Soiland Reyes S., Sinnott R.O., Lonie A., Goble C., Crusoe M.R. Sharing interoperable workflow provenance: A review of best practices and their practical application in CWLProv. GigaScience

8(11) (2019).
23. Kim J., Deelman E., Gil Y., Mehta G., Ratnakar V. Provenance trails in the Wings Pegasus system. Concurrency and Computation: Practice and Experience 20(5) (2008), 587 to 597.
24. Kluyver T., Ragan Kelley B., Perez F., Granger B., Bussonnier M., Frederic J., Kelley K., Hamrick J., Grout J., Corlay S., Ivanov P., Avila D., Abdalla S., Willing C., Jupyter Development Team. Jupyter Notebooks a publishing format for reproducible computational workflows. 2016.
25. Kunze J., Littman J., Madden E., Scancella J., Adams C. The BagIt File Packaging Format. RFC 8493, 2018.
26. Kurowski K., Corcho O., Choirat C., Eriksson M., Coppens F., van de Sanden M., Ojstersek M. EOSC Interoperability Framework. Technical Report, 2021.
27. Lebo T., Sahoo S., McGuinness D., Belhajjame K., Cheney J., Corsar D., Garijo D., Soiland Reyes S., Zednik S., Zhao J. PROV O: The PROV Ontology. W3C Recommendation, 2013.
28. Leipzig J., Nust D., Hoyt C.T., Ram K., Greenberg J. The role of metadata in reproducible computational research. Patterns 2(9) (2021), 100322.
29. Miksa T., Jaoua M., Arfaoui G. Research object crates and machine actionable data management plans. 2020.
30. Miksa T., Simms S., Mietchen D., Jones S. Ten principles for machine actionable data management plans. PLOS Computational Biology 15(3) (2019), e1006750.
31. Mons B. Data Stewardship for Open Science. 2018.
32. Moller S., Krabbenhoft H.N., Tille A., Paleino D., Williams A., Wolstencroft K., Goble C., Holland R., Belhachemi D., Plessy C. Community driven computational biology with Debian Linux. BMC Bioinformatics 11(Suppl 12) (2010), S5.
33. Moller S., Prescott S.W., Wirzenius L., Reinholdtsen P., Chapman B., Prins P., Soiland Reyes S., Klotzl F., Bagnacani A., Kalas M., Tille A., Crusoe M.R. Robust cross platform workflows. Data Science and Engineering 2(3) (2017), 232 to 244.
34. Neylon C. As a researcher I am a bit fed up with Data Management. 2017.
35. OCFL. Oxford Common File Layout Specification. Recommendation, 2020. https://ocfl.io/1.0/spec/.
36. RDF Working Group. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, 2014.
37. Rehm H.L., Page A.J.H., Smith L., Adams J.B., Alterovitz G., Babb L.J., Barkley M.P., Baudis M., Beauvais M.J.S., Beck T., Beckmann J.S., Beltran S., Bernick D., Bernier A., Bonfield J.K., Boughtwood T.F., Bourque G., Bowers S.R., Brookes A.J., Brudno M., Brush M.H., Bujold D., Burdett T., Buske O.J., Cabili M.N., Cameron D.L., Carroll R.J., Casas Silva E., Chakravarty D., Chaudhari B.P., Chen S.H., Cherry J.M., Chung J., Cline M., Clissold H.L., Cook Deegan R.M., Courtot M., Cunningham F., Cupak M., Davies R.M., Denisko D., Doerr M.J., Dolman L.I., Dove E.S., Dursi L.J., Dyke S.O.M., Eddy J.A., Eilbeck K., Ellrott K.P., Fairley S., Fakhro K.A., Firth H.V., Fitzsimons M.S., Fiume M., Flicek P., Fore I.M., Freeberg M.A., Freimuth R.R., Fromont L.A., Fuerth J., Gaff C.L., Gan W., Ghanaim E.M., Glazer D., Green R.C., Griffith M., Griffith O.L., Grossman R.L., Groza T., Guidry Auvil J.M., Guigo R., Gupta D., Haendel M.A., Hamosh A., Hansen D.P., Hart R.K., Hartley D.M., Haussler D., Hendricks Sturrup R.M., Ho C.W.L., Hobb A.E., Hoffman M.M., Hofmann O.M., Holub P., Hsu J.S., Hubaux J.P., Hunt S.E., Husami A., Jacobsen J.O., Jamuar S.S., Janes E.L., Jeanson F., Jene A., Johns A.L., Joly Y., Jones S.J.M., Kanitz A., Kato K., Keane T.M., Kekesi Lafrance K., Kelleher J., Kerry G., Khor S.S., Knoppers B.M., Konopko M.A., Kosaki K., Kuba M., Lawson J., Leinonen R., Li S., Lin M.F., Linden M., Liu X., Liyanage I.U., Lopez J., Lucassen A.M., Lukowski M., Mann A.L., Marshall J., Mattioni M., Metke Jimenez A., Middleton A., Milne R.J., Molnar Gabor F., Mulder N., Munoz Torres M.C., Nag R., Nakagawa H., Nasir J., Navarro A., Nelson T.H., Niewielska A., Nisselle A., Niu J., Nyronen T.H., O Connor B.D., Oesterle S., Ogishima S., Ota Wang V., Paglione L.A.D., Palumbo E., Parkinson H.E., Philippakis A.A., Pizarro A.D., Prlic A., Rambla J., Rendon A., Rider R.A., Robinson P.N., Rodarmer K.W., Rodriguez L.L., Rubin A.F., Rueda M., Rushton G.A., Ryan R.S., Saunders G.I., Schuilenburg H., Schwede T., Scollen S., Senf A., Sheffield N.C., Skantharajah N., Smith A.V., Sofia H.J., Spalding D., Spurdle A.B., Stark Z., Stein L.D., Suematsu M., Tan P., Tedds J.A., Thomson A.A., Thorogood A., Tickle T.L., Tokunaga K., Tornroos J., Torrents D., Upchurch S., Valencia A., Varma S., Vears D.F., Viner C., Voisin C., Wagner A.H., Wallace S.E., Walsh B.P., Williams M.S., Winkler E.C., Wold B.J., Wood G.M., Woolley J.P., Yamasaki C., Yates A.D., Yung C.K., Zass L.J., Zaytseva K., Zhang J., Goodhand P., North K., Birney E. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genomics 1(2) (2021), 100029.
38. Rettberg N., Schmidt B. OpenAIRE. College and

Research Libraries News 76(6) (2015), 306 to 310.

39. Sandve G.K., Nekrutenko A., Taylor J., Hovig E. Ten simple rules for reproducible computational research. PLOS Computational Biology 9(10) (2013), e1003285.
40. Schriml L.M., Chuvochina M., Davies N., Eloe Fadrosh E.A., Finn R.D., Hugenholtz P., Hunter C.I., Hurwitz B.L., Kyrpides N.C., Meyer F., Mizrachi I.K., Sansone S.A., Sutton G., Tighe S., Walls R. COVID 19 pandemic reveals the peril of ignoring metadata standards. Scientific Data 7(1) (2020), 188.
41. Sefton P., Devine G., Evenhuis C., Lynch M., Wise S., Lake M., Loxton D. DataCrate: a method of packaging, distributing, displaying and archiving Research Objects. 2018.
42. Sefton P. FAIR Data Management Its a lifestyle not a lifecycle. 2021.
43. Sefton P., O Carragain E., Soiland Reyes S., Corcho O., Garijo D., Palma R., Coppens F., Goble C., Fernandez J.M., Chard K., Gomez Perez J.M., Crusoe M.R., Eguinoa I., Juty N., Holmes K., Clark J.A., Capella Gutierrez S., Gray A.J.G., Owen S., Williams A.R., Tartari G., Bacall F., Thelen T. RO Crate Metadata Specification 1.0. 2019.
44. Sefton P., O Carragain E., Soiland Reyes S., Corcho O., Garijo D., Palma R., Coppens F., Goble C., Fernandez J.M., Chard K., Gomez Perez J.M., Crusoe M.R., Eguinoa I., Juty N., Holmes K., Clark J.A., Capella Gutierrez S., Gray A.J.G., Owen S., Williams A.R., Tartari G., Bacall F., Thelen T., Menager H., Rodriguez Navas L., Walk P., Whitehead B., Wilkinson M., Groth P., Bremer E., Castro L.G., Sebby K., Kanitz A., Trisovic A., Kennedy G., Graves M., Koehorst J., Leo S., Portier M. RO Crate Metadata Specification 1.1.1. 2021.