EOSCsecretariat.eu
Setup and management of the EOSC Secretariat supporting the EOSC Governance

Final report of Co-creation Activity

Id: **0063**

Title of Activity: **Alignment of EOSC Strategic Implementation Plan with US NIH Strategic Plan for Data Science**

Applicant: **Carlos Luis Parra-Calderón**

Date: February 5, 2021

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644 European a is an initiative of the European Union, financed by the European Union's Connecting Europe Facility and European Union Member States (https://pro.europeana.eu/ and https://www.europeana.eu)".

Final Activity Summary.

Description of the activity

The NIH Strategic Plan for Data Science describes NIH's Overarching Goals, Strategic Objectives, and Implementation Tactics for modernizing the NIH-funded biomedical data-resource ecosystem.
Adherence to FAIR principles is key to this strategy.
In this respect, this plan has important similarities with the European Union's EOSC Strategy.
The NIH Plan develops eight strategic objectives. The idea is to identify the similarities and differences between the different levels of this Plan (Overarching Goals, Strategic Objectives, and Implementation Tactics) and the EOSC Strategic Implementation Plan.
Once the similarities and differences have been identified, the details of these will be explained, and a set of actions will be proposed for collaboration and strengthening of the plans.
The alignment in time is fundamental since both plans are being developed close in time, but with differences, therefore the proposed actions have to be viable throughout the year 2020 since even in the very short term there will be difficulties for the synchronization of the actions in the temporal logic of each plan.

Development of the planned tasks.

The work has been developed generally according to plan, and this development is detailed below, except for the participation in the AMIA Symposium 2020 Panel. Also, during the task's development, it has detected fundamental collaboration opportunities with the NIH Office of Data Science Strategy. An effective communication channel has been established, with regular meetings with Director Susan Gregurick and her team. In this sense, it has

materialized in a concrete action, which participates in the HL7 FHIR4FAIR Implementation Guide.
In any case, the collaboration is established, and the communication channel is open with the Data Science Office, and we maintain joint tasks. Now we are evaluating the preparation of a joint paper to be presented at the next AMIA Symposium 2021 in November.

Benefit for EOSC achieved.

This activity has had several beneficial impacts for EOSC:

- First-hand knowledge of the EOSC initiative itself by the Director of the NIH Office of Data Science Strategy, as well as key documents.
- Identification of relevant documents from NIH or related documents that substantiate a potential collaboration with EOSC to develop FAIR data policies.
- Recommendations for collaboration with EOSC on RFIs for the "Strategic Opportunities and Challenges for the National Library of Medicine, National Institutes of Health" (Annex I) and the "NIH-Wide Strategic Plan for COVID-19 Research" (Annex IV).
- Effective collaboration in the development of the HL7 FHIR4FAIR Implementation Guide currently underway.

Details of the tasks performed.

Task 1: Hours of NIH and EOSC documentation, exchange information, analysis and document editing.

Task 1 has been completed. As for the proposed methodology, it has been finally adapted to the relevant terms proposed by the EOSC Board and new relevant documents have been referenced throughout the development of the activity, being in any case classified and studied according to the foreseen method:

- Classification as primary or secondary document.
- Annotation of typology, authorship of the document, term of validity and force of implementation.
- Identification of the frequency with which the following concepts/keywords are added in each of the documents appears and the concepts already identified so far:
    - Machine-Readability
    - Data-visiting
    - Distributed Data Mining DDM
    - Open Science
    - Open Data
    - Landscape
    - Rules of Participation
    - Architecture
    - FAIR
    - Sustainability
    - International Collaboration

With respect to this list of concepts, it has been detected that the first 3 have not appeared at all, and it has been observed that in some cases similar concepts have been used, but in any case, of very low relevance.

Analyzed primary documents.

NIH-Wide Strategic Plan for COVID-19 Research
https://www.nih.gov/sites/default/files/research-training/initiatives/covid-19-strategic-plan/coronavirus-strategic-plan-20200713.pdf Last reviewed on July 13, 2020.

NIH Strategic Plan for Data Science (SP4DS)
https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf Last reviewed on August 7, 2019

National Health IT Priorities for Research: A Policy and Development Agenda
https://www.healthit.gov/sites/default/files/page/2020-01/PolicyandDevelopmentAgenda.pdf. Last reviewed on February 24, 2020

EOSC Strategic Implementation Plan (EOSC SIP)
https://ec.europa.eu/info/publications/european-open-science-cloud-eosc-strategic-implementation-plan_en Published on July 24, 2019

A Platform for Biomedical Discovery and Data-Powered Health
National Library of Medicine Strategic Plan 2017–2027 - Report of the NLM Board of Regents – December 2017
https://www.nlm.nih.gov/pubs/plan/lrp17/NLM_StrategicReport2017_2027.pdf

Analyzed secondary documents.

DRAFT NIH Policy for Data Management and Sharing
https://osp.od.nih.gov/wp-content/uploads/Draft_NIH_Policy_Data_Management_and_Sharing.pdf

Compiled Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance
https://osp.od.nih.gov/wp-content/uploads/RFI_Final_Report_Feb2020.pdf

Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan
https://osp.od.nih.gov/wp-content/uploads/Supplemental_DRAFT_Guidance_Elements_NIH_Data_Management_and_Sharing_Plan.pdf

NIH Data Sharing Policies
https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html

EOSC Work Plan 2020
https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/ae215698-af7b-11ea-bb7a-01aa75ed71a1 Published on June 15, 2020

EOSC Interoperability Framework (v1.0) - Draft for community consultation https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf Published on May 3, 2020

National Academies of Sciences, Engineering, and Medicine. 2020. Life Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs. Washington, DC: The National Academies Press. https://doi.org/10.17226/25639.

Recommendations on FAIR Metrics for EOSC https://op.europa.eu/en/publication-detail/-/publication/ced147c9-53c0-11eb-b59f-01aa75ed71a1/language-en Published on January, 2021

The deliverable "Analysis of NIH and EOSC documentation" final version is submitted as an annex at the end of this report (Annex II).

Task 2: "Travel expenses to USA, panel participation in AMIA Symposium 2020 includes registration fee."

This task has already been canceled (see the previous follow-up report). This expense concept is canceled.

Task 3: Hours for the development of the alignment plan and the communication plan.

The alignment plan established with the Director of the NIH Office of Data Science Strategy has been agreed on the collaboration for the development of the FAIRness for FHIR Implementation Guide in the HL7 Standardization Development Organization (Annex IV).

The communication channel with the NIH Office is still open, and future collaborations have been established, both in the framework of the RDA working groups and in possible immediate scientific collaborations for conferences in the field of medical informatics (Medinfo 2021 and AMIA Symposium 2021).

Task 4: Hours of follow-up and management of the proposed actions. Reports to EOSC Board.

Four follow-up reports (including this one) have been submitted to the EOSC Board of Directors.

Work has been carried out on two specific issues under the direction of the EOSC Board, which have materialized in the inclusion of new concepts to be analyzed in the documentation studied (Annex II), and in an analysis of the NIH funding programs for ELSI issues (Annex III).

A presentation of this action was developed in the co-creation session at the EOSC Symposium 2020, last October 21. The recording and the presentation are available [1][2]

[1] https://youtu.be/tRgu54J5i0E

[2] https://repository.eoscsecretariat.eu/index.php/s/oEJtGJMB9gBYJnK/download?path=%2FDay%203%20(21%20October%202020)%20-%20EOSC%20implementation%20challenges%2FCo-

creation%20session&files=EOSC%20Symposium_NIH_EOSC%20Alignement%20-%20Carlos%20Luis%20Parra%20Calderon.pdf&downloadStartSecret=zv8prc5z6fm

Annex I

Request for Information (RFI): Strategic Opportunities and Challenges for the National Library of Medicine, National Institutes of Health.

Final Version
October 6, 2020

Major opportunities or challenges that have emerged over the last five years and that have implications for the future of NLM in the area of:

a) Science (including clinical health sciences, biomedical science, information science, informatics, data analytics, data science, etc.)

Within the scope of Objective 1.2, "Advance research and development in biomedical informatics and data science," I propose to collaborate with the European Open Science Cloud (EOSC) to define a common FAIR Digital Object for Health Research (De Smedt, K.; Koureas, D.; Wittenburg, P. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications 2020, 8, 21.).

I also propose to work to make research questions computable (Vazquez M, Valencia A. Patient Dossier: Healthcare queries over distributed resources. PLoS Comput Biol. 2019 Oct 17;15(10):e1007291. doi: 10.1371/journal.pcbi.1007291. PMID: 31622330; PMCID: PMC6797086.) and generate evidence on SARS-COV-2 and Covid19 automatically from what is published (https://www.gps.health/covid19_knowledge_accelerator.html).

b) Technology (including biotechnology, platforms, hardware, software, algorithms, processes, systems, etc.)

Within the scope of objective 1.4, "Create a sustainable institutional, physical, and computational infrastructure," I propose to collaborate with EOSC in technical challenges such as authentication and authorization, assuring compliance with permissions required for data access and re-use, and security of data sets and session permission. EOSC seeks to integrate and consolidate electronic infrastructure platforms for research to federate existing research infrastructures and scientific clouds and to support the development of cloud-based services for Open Science.

f) Perspectives, practices, and policies (including those related to open science, the need for diversity, equity, and inclusion in research, algorithmic bias, expectations of reproducibility of research, etc.)

Within the scope of objective 1.3, "Foster open science policies and practices", I propose to collaborate with Research Data Alliance in the definition of open science policies at a global level. Specifically, with the Practical Policy WG (https://www.rd-alliance.org/groups/practical-policy-wg.html) y con el Global Open Research Commons IG (https://www.rd-alliance.org/groups/global-open-research-commons-ig).

Major opportunities or challenges that have emerged in the last five years and that have implications for the future of NLM in other areas or areas not well captured above.

The opportunity to collaborate with the European Open Science Cloud initiative.

Opportunities or challenges on the horizon over the next five years that fall within the purview of the NLM's mission.

The effective implementation of infrastructures that allow the management of FAIR Digital Objects for Health Research, even at the international level, must be a challenge on the horizon of the next five years.

Annex II

Analysis of NIH and EOSC documentation.

Final Version
February 5, 2021

This annex includes the third version of the results of the analysis of the primary and secondary documents identified so far. It includes the frequency of occurrence of the key concepts in these documents.
These results are represented in the following two tables (in green the rows of the new incorporated document).

Document name	Typology	Authorship	Term of validity	Force of implementation
NIH-Wide Strategic Plan for COVID-19 Research	Primary	NIH Office of the Director (OD)	Draft pending update from public comment compilation	High
A Platform for Biomedical Discovery and Data-Powered Health	Primary	NLM Board of Regents	Current document in the process of being updated	High
NIH Strategic Plan for Data Science	Primary	NIH Office of the Director (OD)	Current document but not recently updated.	High
National Health IT Priorities for Research: A Policy and Development Agenda	Primary	The Office of the National Coordinator for Health Information Technology (ONC)	Current and recently updated.	High
EOSC Strategic Implementation Plan	Primary	EOSC Executive Board	Current document but not recently updated.	High
Draft NIH Policy for Data Management and Sharing	Secondary	NIH Office of Science Policy	November 2019 draft pending update from public comment compilation	High
Compiled Public Comments on a Draft NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance	Secondary	Public Comments	Comments completed on January 10, 2020.	Medium
Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan	Secondary	NIH Office of Science Policy	November 2019 draft pending update from public comment compilation	High

NIH Data Sharing Policies	Secondary	Trans-NIH BioMedical Informatics Coordinating Committee (BMIC)	Last Reviewed: August 9, 2019	Medium
EOSC Work Plan 2020	Secondary	EOSC Executive Board	Current and recently publication.	High
EOSC Interoperability Framework	Secondary	Interoperability TF of the EOSC FAIR WG, with participation from the EOSC Architecture WG.	Draft for consultation and recently publication.	Medium
Life Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs	Secondary	US National Academies of Sciences, Engineering, and Medicine.	Prepublication Copy—Subject to Further Editorial Correction	Medium
Recommendations on FAIR Metrics for EOSC	Secondary	EOSC Executive Board	Current and recently publication.	High

Document name	Frequency of occurrence of relevant concepts										
	Machine-Readability	Data-visiting	Distributed Data Mining DDM	Open Science	Open Data	Landscape	Rules of Participation	Architecture	FAIR	Sustainability	International Collaboration
NIH-Wide Strategic Plan for COVID-19 Research	0	0	0	0	0	0	0	0	1	0	1
A Platform for Biomedical Discovery and Data-Powered Health	0	0	0	25	0	1	0	0	3	3 (sustainable)	0
NIH Strategic Plan for Data Science	0	0	0	0	3	2	0	3	17	7	10 (with the concept of international agencies)
National Health IT Priorities for Research: A Policy and Development Agenda	0	0	0	0	2 (excluding references).	5	0	17 (excluding references, table of contents and annexes)	8 (excluding references)	1	0

EOSC Strategic Implementation Plan	0	0	0	8 (excluding references, table of contents and annexes)	1 (excluding references, table of contents and annexes)	10 (excluding references, table of contents and annexes)	22	13 (excluding references, table of contents and annexes)	64 (excluding references, table of contents and annexes)	17 (excluding references, table of contents and annexes)	1 (in annex)
Draft NIH Policy for Data Management and Sharing	0	0	0	0	0	0	0	0	1	0	0
Compiled Public Comments on a Draft NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance	0	0	0	28 (excluding references)	20 (excluding references)	6	0	6	>100 (excluding references and table of contents)	14	1
Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan	0	0	0	0	0	0	0	0	0	0	0
NIH Data Sharing Policies	0	0	0	0	0	0	0	0	0	0	0
EOSC Work Plan 2020	0	0	0	8 (excluding references, table of contents and annexes)	0	14	8	18	22 (excluding references)	13	11 (uses the concept of internationalization in a variety of ways)

EOSC Interoperability Framework	1 7 (machine readable)	0	0	6	0	0	3	5 (excluding references)	39 (excluding references)	2	1 (international level)
Life Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs	0	0	0	3 (excluding references, table of contents, bios and annexes)	2 (excluding references)	7 (excluding table of contents)	0	4 (excluding bios)	21 ((excluding references, table of contents, bios and acronyms)	4 (excluding acronyms)	0
Recommendations on FAIR Metrics for EOSC	0	0	0	0	1	0 (excluding references).	0	0 (excluding the name of the EOSC Architecture WG)	>100	1 (excluding references).	1

The frequency with which the concepts appear in the various documents already gives an overview of the issues with a broad view with the highest potential for strategic alignment.

The critical documents by or linking to NIH that demonstrate alignment on the fundamentals in the FAIR data policy are as follows:

- NIH Strategic Plan for Data Science
- National Health IT Priorities for Research: A Policy and Development Agenda
- Compiled Public Comments on a Draft NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance

Annex III

NIH Funding Opportunities on ELSI for Health Research.

Final Version
October 6, 2020

There are three types of funding opportunities open for research projects at ELSI that differ in duration, type of study and magnitude of funding. All calls are focused on human genome research.

For small projects, such as those involving single investigators, focused conceptual or analytical studies, or secondary data analyses, applicants may wish to consider **PAR-20-257, the ELSI Small Grant (R03) FOA**, which provides a total of up to $50,000 in direct costs a year for two years. For small projects that are primarily exploratory or designed to generate pilot data in preparation for a larger study, applicants should consider **PAR-20-255, the ELSI Exploratory/ Developmental Research (R21) FOA**, which provides a total of up to $275,000 in direct costs over two years. For larger multi-disciplinary studies building on preliminary data and requiring funding beyond two years, applicants may wish to consider **PAR-20-254, the ELSI Research Project Grant (R01) FOA**, which provides funding for up to five years. Note that not all Institutes participating in this R21 FOA participate in the R03 FOA.

PAR-20-254, the ELSI Research Project Grant (R01) FOA.
This Funding Opportunity Announcement (FOA) invites Research Project Grant (R01) applications that propose to study the ethical, legal, and social implications (ELSI) of human genome research. Applications may present studies using either single or mixed methods. Proposed approaches may include but are not limited to data-generating qualitative and quantitative approaches, legal, economic and normative analyses, and other types of analytical and conceptual research methodologies, such as those involving the direct engagement of stakeholders.

PAR-20-255, the ELSI Exploratory/ Developmental Research (R21) FOA.
This Funding Opportunity Announcement (FOA) invites Exploratory/Developmental Research Grant (R21) applications that propose to study the ethical, legal, and social implications (ELSI) of human genome research. These applications should offer single or mixed methods studies that break new ground, extend previous discoveries in new directions, or develop preliminary data to prepare larger studies. Of particular interest are studies that explore the implications of new or emerging genomic technologies or novel uses of genomic information.

PAR-20-257, the ELSI Small Grant (R03) FOA.
This Funding Opportunity Announcement (FOA) invites Small Research Grant (R03) applications to study the ethical, legal, and social implications (ELSI) of human genome research. These applications should be for small, self-contained research projects, such as those that involve single investigators. Of particular interest are projects that propose normative or conceptual analyses, including focused legal, economic, philosophical, anthropological, or historical studies of new or emerging issues. This mechanism can also be used to collect preliminary data and the secondary analysis of existing data.
The increased integration of personal genomic information into many aspects of modern life raises new and urgent research questions. Genomic data may begin to influence behavior and our concepts of health and disease, responsibility and justice, family, identity, and community.

Annex IV

Comments and Suggestions on the NIH-Wide Strategic Plan for COVID-19 Research.

Final Version
December 8, 2020

From this activity, we have participated in the call "Request for Information (RFI): Inviting Comments and Suggestions on the NIH-Wide Strategic Plan for COVID-19 Research." (NOT-OD-21-018).

The Strategic Plan provides a framework that describes how NIH is accelerating the development of therapeutic interventions, vaccines, and diagnostics in response to the COVID-19 pandemic. The Strategic Plan outlines how NIH will implement five Priorities guided by three Crosscutting Strategies:

Priorities

- Improve Fundamental Knowledge of SARS-CoV-2 and COVID-19
- Advance Detection and Diagnosis of COVID-19
- Advance the Treatment of COVID-19
- Improve Prevention of SARS-CoV-2 Infection
- Prevent and Redress Poor COVID-19 Outcomes in Health Disparity and Vulnerable Populations

Crosscutting Strategies

- Partnering to promote collaborative science
- Supporting the research workforce and infrastructure
- Investing in data science

In this activity, comments and suggestions have been proposed to the priorities of advancing treatment and preventing and reducing health disparities in vulnerable populations. An ambitious proposal has also been made to promote a memorandum of understanding to develop a Joint Research Center for Data Science in SARS-COV-2 and Covid19, within the crosscutting strategy of partnership for collaborative science.

The comments and suggestions are detailed below.

Priority 3. Advance the Treatment of COVID-19

Operation Warp Speed must take into account the traceability of citizens who refuse to be vaccinated, predicting the impact on herd immunity that this may cause.

Priority 5. Prevent and Redress Poor COVID-19 Outcomes in Health Disparity and Vulnerable Populations

Regarding objective 5.4 "Address global health research needs from COVID-19", it is proposed to develop connectors for the interoperability of data with Europe that will be findable, accessible,

interoperable, and reusable through the European COVID-19 Data Platform (https://www.covid19dataportal.org/).

Crosscutting Strategies

Promote a Memorandum of Understanding between the US and the European Union to create a US-European Joint Research Center in Data Science at SARS-COV-2 and Covid19, giving the leading role NIH Data Science Office and the European Open Science Cloud. The goal of such a center should be to advance FAIR data for biomedical and health research from the US and Europe. FAIR data will include all types of data, from chemical structures to clinical trial results. This JRC should attend to the rapid delivery of guidelines and data management protocols to meet these objectives.

Additional Comments

The scientific response to the Covid-19 pandemic requires a deepening of the need for international collaboration made effective by making FAIR the data to research.
For this research to be effective, syntactic and semantic interoperability standards must be agreed upon, and a legal and ethical context defined to facilitate collaboration and access to data.

Annex V

Proposal for NIH collaboration in the development of the FAIRness for FHIR Implementation Guide in the HL7 Standardization Development Organization.

Final Version
February 5, 2021

As part of the development of the FAIR4Health[3] project (EOSC-related), a FAIRness for FHIR Implementation Guide (FHIR4FAIR)[4] is being developed in the HL7 Standards Development Organization. The person responsible for this co-creation activity participates in the guideline working group, as well as other colleagues from the FAIR4Health project, the Research Data Alliance and industry.

HL7's FHIR is the world's most widely accepted and pervasive health information standard in the healthcare domain. This guide to implementing FAIR principles in health datasets in FHIR schemes will be a natural and useful resource for exponential FAIRification of health data in the world.

From this activity, the Director and team of the NIH Office of Data Science Strategy were invited to participate in this guide's development. This proposal has been very positively evaluated, and a member from the office has already joined the HL7 Working Group.

The guide is currently under development, with intense activity. The aim is to have a balloting version in HL7 in May 2021.

The effective collaboration in the development of this standard by the NIH Strategic Office of Data Science, the RDA, and the leadership from the FAIR4Health project EOSC-related, is evidence of the success of this activity on NIH and EOSC alignment. This standard guarantees the sustainability of this collaboration and provides a relevant impact on the facilitation of the application of FAIR principles in Health Research.

[3] https://www.fair4health.eu/
[4] https://confluence.hl7.org/pages/viewpage.action?pageId=91991234