www.ceb.uminho.pt/bbig

University of Minho
School of Engineering

CENTRE OF BIOLOGICAL ENGINEERING

AZEREDO LAB

DATASTEWARD

IN SCIENCE

SKILLS AND CHALLENGES

Sílvio Santos | silviosantos@deb.uminho.pt

FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO

- Nov -
15

FITTED - Project

FITTED
A flexible platform for generating synthetic virus to control infectious diseases

The FITTED project intends to develop a flexible platform for generating synthetic virus to control infectious diseases.

Collected and generated data

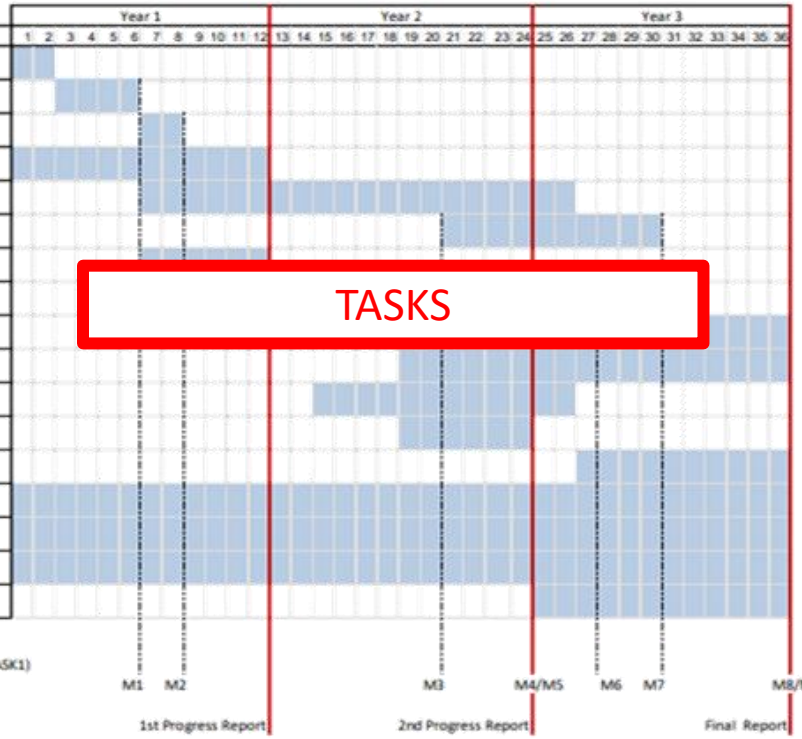
- Genomic data
- Proteomic data
- Protocols
- Images
- Code (software)
- Process
- Manuscript

FITTED - Project

Project reference: HR21-00533

Project title: FITTED - A flexible platform for generating synthetic virus to control infectious diseases

Task #	Task Denomination
1	1.1 *In vitro* characterization of native viruses
	1.2 *In vivo* selection of long-circulating viruses
	1.3 Transcriptomic profiling of viruses-bacteria infection
2	2.1 Creation of the dataset
	2.2 Predictive modelling in machine learning
	2.3 Creation of a library of BBPs
3	3.1 Construction of the viral minimal genome
	3.2 Construction of the viral chassis
	3.3 Customization of the viral chassis
	3.4 Production and purification of the synthetic viruses for *in vivo* validation and stability evaluation tests
4	4.1 Optimization of the *in vivo* model
	4.2 *In vivo* selection of the best chassis
	4.3 *In vivo* safety and performance evaluation of synthetic viruses
5	5.1 Administrative management
	5.2 Technical management
	5.3 Dissemination and communication
	5.4 IP Management



M1: Obtention of genomic sequences of viruses with optimal *in vitro* performances and high *in vivo* stability (TASK1)
M2: Generation of transcriptomic data for the design of the viral minimal genome (TASK1)
M3: D
M4: O
M5: O
M6: O
M7: Organization of an Iberian symposium about alternatives to fight infectious diseases (TASK3)
M8: Obtention of a library of 40 synthetic viruses with customized BBP for the most relevant *A. baumannii* (TASK3)
M9: Evaluation of the stability and shelf-life of 3 synthetic viruses (TASK3)
M10: *In vivo* performance evaluation of synthetic viruses (TASK4)

MILESTONES

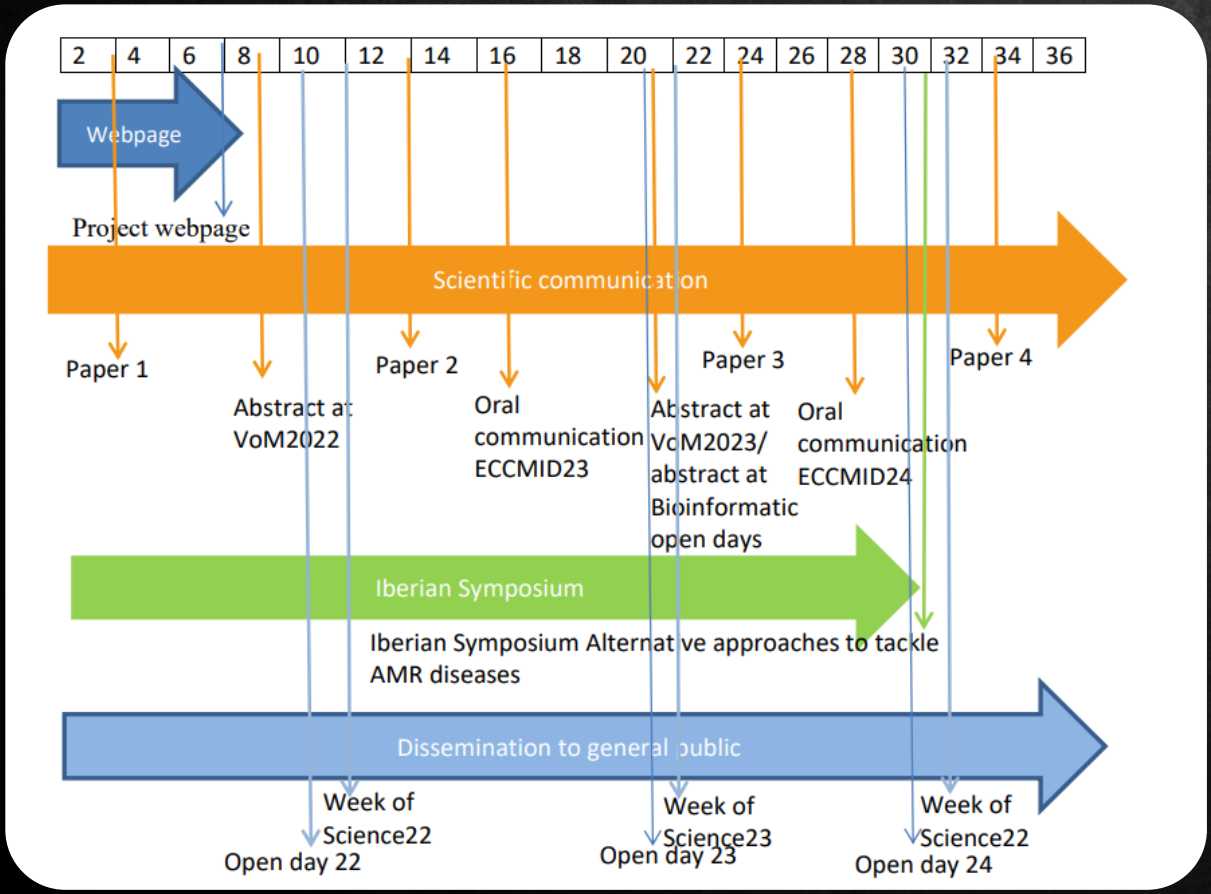
Table 2. FITTED's deliverables and due time

Deliverable name	Date (#project month)
D1.1 A report on performance evaluation of native viruses	2
D1.2 High performance and long-circulating viruses and respective genomes	6
D1.3 A report on the transcriptome profiling of viruses-bacteria interaction in blood media	8
D2.1 A report of new BPP recognizing different *A. baumannii* capsules discovered by random mutagenesis	12
D2.2 Algorithm that predicts BBP for all capsule types	26
D2.3 Software available online	30
D2.4 A library of BBPs for the 40 most relevant clinic *A. baumannii* capsule types	30
D3.1 Report on the minimal genome obtained by gene deletion	12
D3.2 Report on the *in vitro* efficacy of different viral chassis based on the minimal genome plus the insertion of cargos	24
D3	
D3	
D4 on the minimal genome plus the insertion of cargos	
D4.2 Report on the optimized *in vivo* model	24
D4.3 Report on the safety and efficacy evaluation of synthetic viruses	36
D5.1 Project webpage	3
D5.2 Communication and dissemination plan	6
D5.3 Intellectual property management plan and patent application	36
D5.4 Project outputs (article publication, participation in conferences)	-*

* To be defined in the dissemination plan

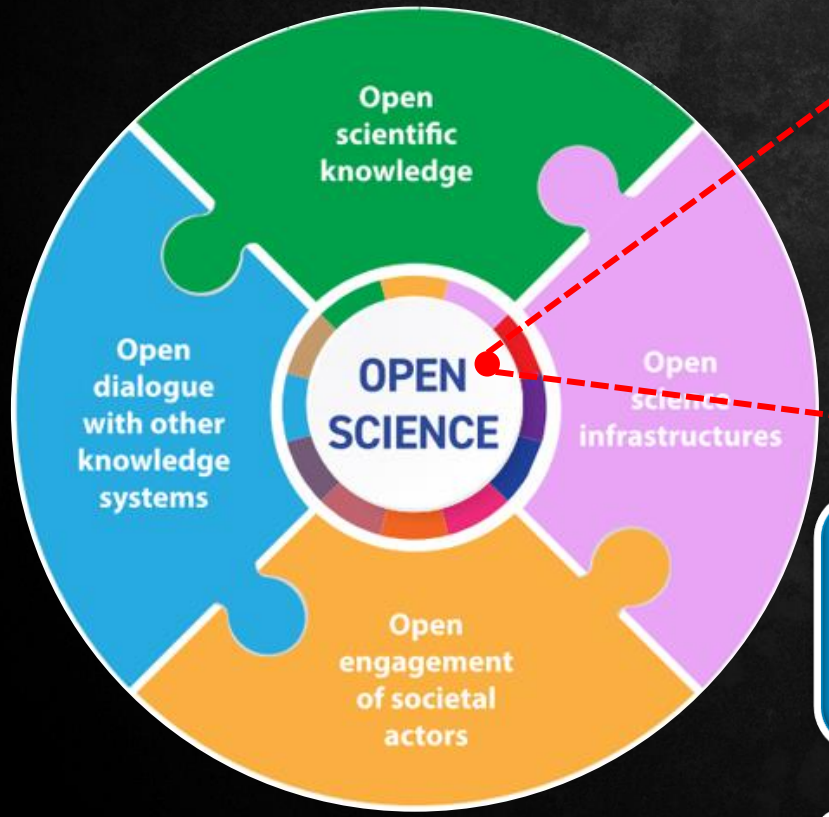
DELIVERABLES

FITTED - Project



- Defined schedule
- 2 research groups
- Many researchers
- Different roles
- DMP

OPEN SCIENCE



As open as possible
As closed as necessary

Availability of data and publications in open access.

openness of the scientific process
(scientific social responsibility)

transfer of scientific knowledge to society.

Advantages of Open Science

- Increases research efficiency
- Increases knowledge of the scientific work process
- Promotes academic rigor and increases the quality of research
- Accelerates the creation of new research themes
- Promotes the involvement of society and culture / scientific literacy
- Increases the economic and social impact of science
- Values intellectual property
- Promotes scientific return to institutions

DMP

Research data life cycle

PLAN

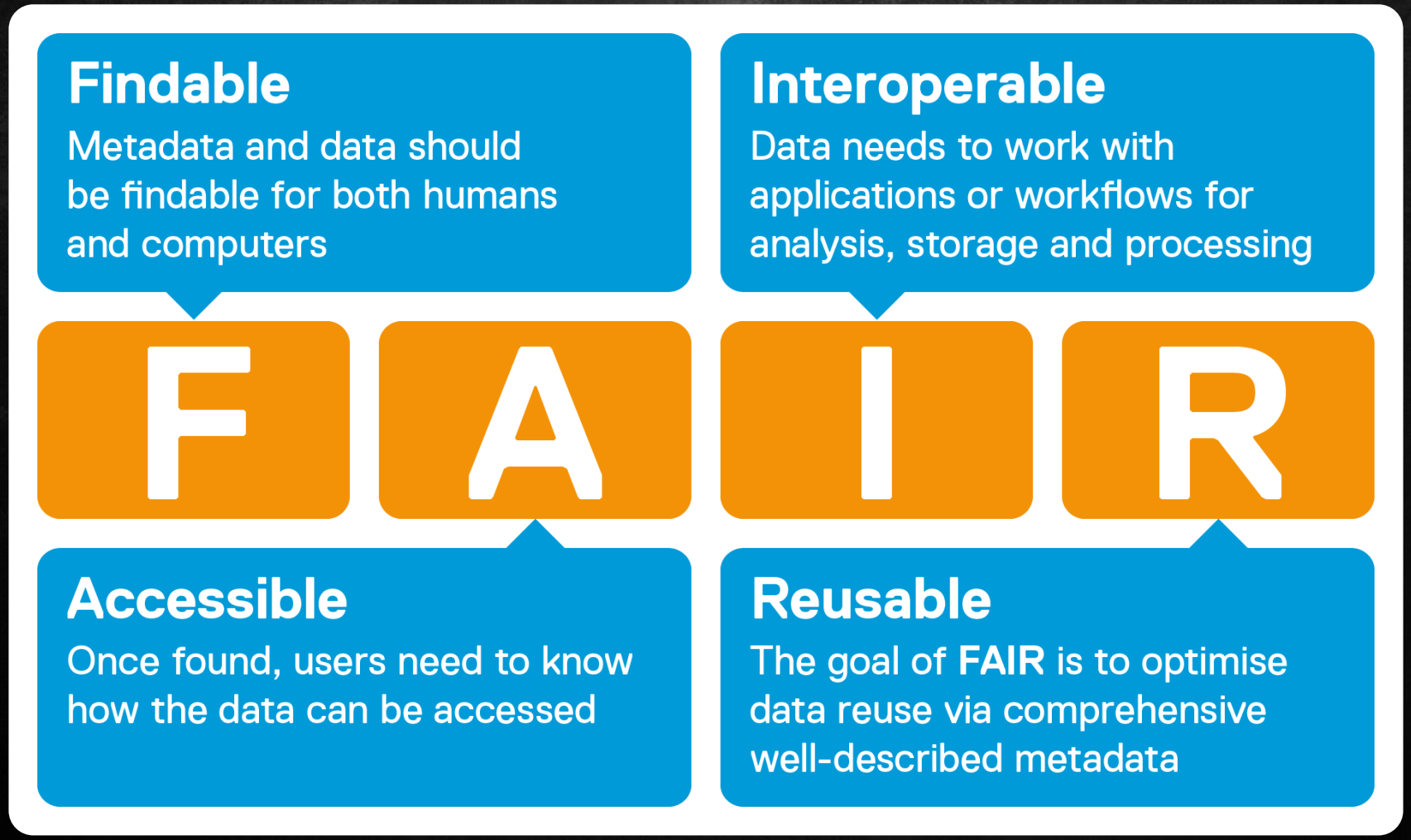
ACCURATE

FAIR

DMP

DYNAMIC

FAIR PRINCIPLES



LaCaixa DMP

FÓRUM GESTÃO DE DADOS DE INVESTIGAÇÃO

Output summary

- The data or output that will be generated or collected.
- The formats and standards that will be used.
- The value the output may have for other researchers.
- What data or output cannot be shared and why.

Responsibilities

- Monitoring the plan.
- Managing the data or outputs.
- Preservation and conservation.

Findable data

- Which metadata will be used.
- Which identifier will be used for the data.
- What version control will be undertaken.

Accessibility

- Who can access the data and when
- Availability of the data
- Methods or software to access the data
- Documentation on the necessary software included
- Where the data, metadata, documentation and code are deposited
- How access is granted.

Interoperability

- Describe how interoperability of the data through the use of controlled vocabularies will be facilitated.

Reusability

- Degree of reuse permitted when the data are made available.
- When the data will be accessible for reuse.
- Will the data generated and/or collected in the project be reusable?
- Periods during which the data may be reused.

Data recovery processing, secure storage and transfer of sensitive data

- Describe the measures adopted to safeguard data security. This includes storage and security copies, long-term storage and how security and protection of the data will be maintained.

Coverage of the ethics review procedure context

- Possible ethical issues in the gathering, retaining, processing or storing of the data.
- Ethical principles and relevant national, EU and international legislation applied.

LaCaixa DMP

Output summary

The data or output that will be generated or collected.	The FITTED project intends to develop a flexible platform for generating synthetic virus to control infectious diseases.
	Multiple data will be collected and generated.
	Data: genomic and proteomic data on the genome, genes and proteins of bacteriophages. Also, data on its properties will be collected or generated
The formats and standards that will be used.	Used international relevant standardized formats: FASTQ/ ab1/FASTA/GenBank/TSV/CSV/JSON/PZFX/JPEG/TIFF/PDF.
	Richly-described metadata, preferably in a machine-readable and machine-actionable format whenever possible.
The value the output may have for other researchers.	Compliance with the FAIR principles.
	Open Science
What data or output cannot be shared and why.	3-year embargo to assess IP

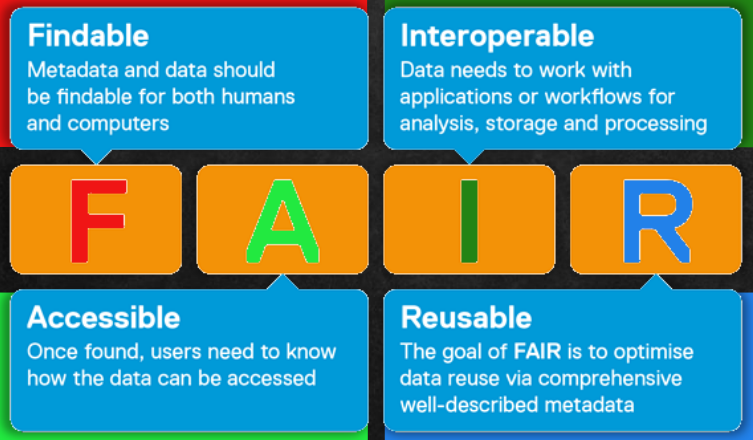
LaCaixa DMP

Responsibilities	Data Steward (DMP)	Finding, gathering, and collecting data	Managing and maintaining/preserving/ conserving the data
Monitoring the plan.	• Sílvio Santos	• Sílvio B Santos • Ivone Martins • Óscar Dias • Hugo Oliveira • Alexandra Fraga • Jorge Pedrosa • Ana Oliveira • Diana Priscila Pires • Luís Melo • Alexandre Lima • Rita Domingues • Joana Azeredo	• Ivone Martins • Óscar Dias • Hugo Oliveira • Alexandra Fraga • Miguel Rocha
Managing the data or outputs.			
Preservation and conservation.			

LaCaixa DMP

- Rich, generous and extensive metadata
- Structured and searchable way
- Keywords and ontology terms
- International standard schemas
- Metadata in a machine-readable format
- Globally unique and persistent identifier – DOI
- Domain-specific repository (genbank)
- Versioning with documented history

- Standardized formats
- Metadata in machine-readable and machine-actionable format
- Open formats
- Ontologies whenever possible
- Qualified references to other data and metadata.
- Chosen repository (dataRepositorium) established using the Dataverse Open Source



- Open science
- Institutional repository (dataRepositoriUM)
- Freely available with obligation to quote
- Embargo (3 years) to assess IP
- Standardized formats (no proprietary)
- License info available
- Persistent metadata

- Optimize data reuse
- Richly-described data and metadata
- Standardized formats
- Freely available without limitations after embargo
- Use of a clear licensing status

LaCaixa DMP

Data recovery processing, secure storage and transfer of sensitive data

Describe the measures adopted to safeguard data security. This includes storage and security copies, long-term storage and how security and protection of the data will be maintained.

- Data stored in our institutional repository (Data Repositorium)
- dataRepositorium allows a safe and secure storage, data recovery and long-term preservation.
- Access to data will be given through data protection methods like password protection, allowing a secure collaboration
- Different access levels will be defined (view; add; edit) mitigating the risk of data loss.
- Mitigate information leak risk by avoiding storage of data on computers in the lab or external hard drives
- We are not using any personal information.

LaCaixa DMP

Coverage of the ethics review procedure context

Possible ethical issues in the gathering, retaining, processing or storing of the data.

We are not using any personal information and ethical issues are not applied here.

Ethical principles and relevant national, EU and international legislation applied.

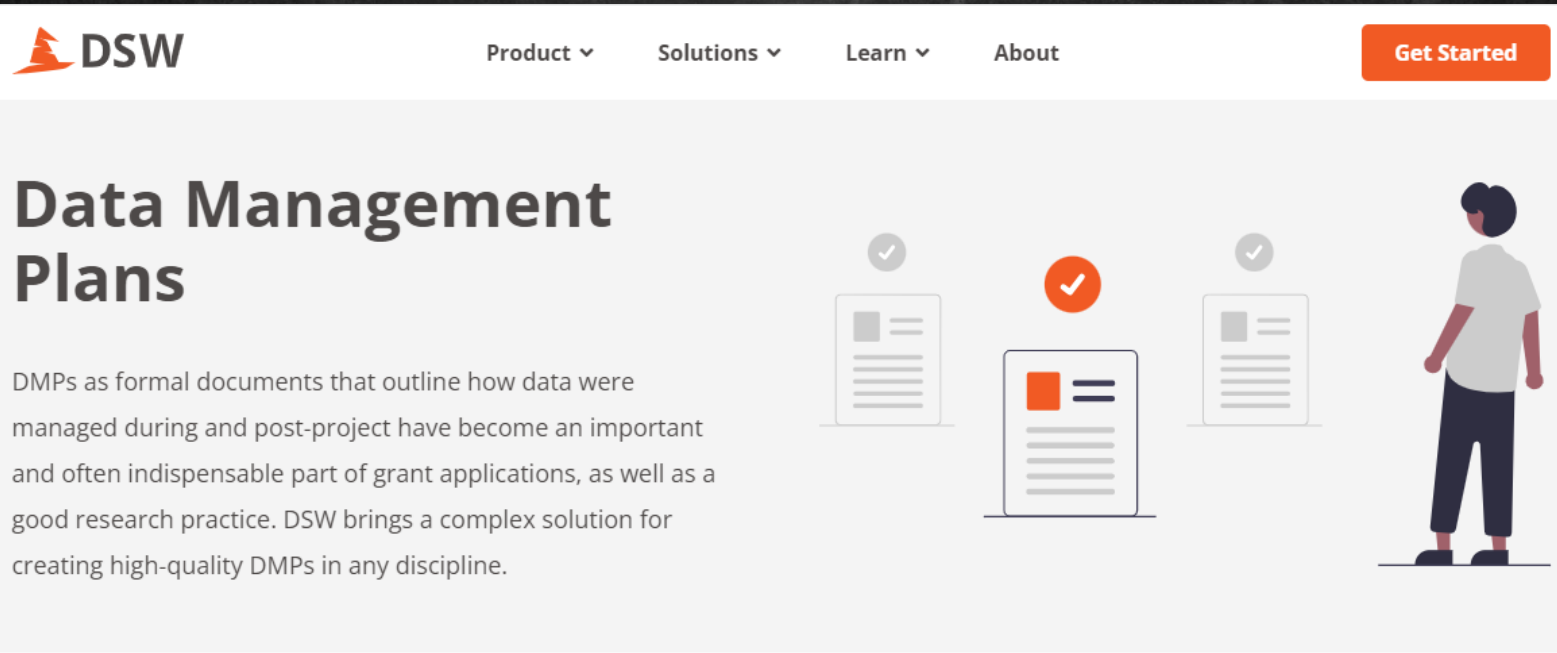
Animal experimentation will be performed at the Life and Health Sciences Research Institute UM (ICVS)

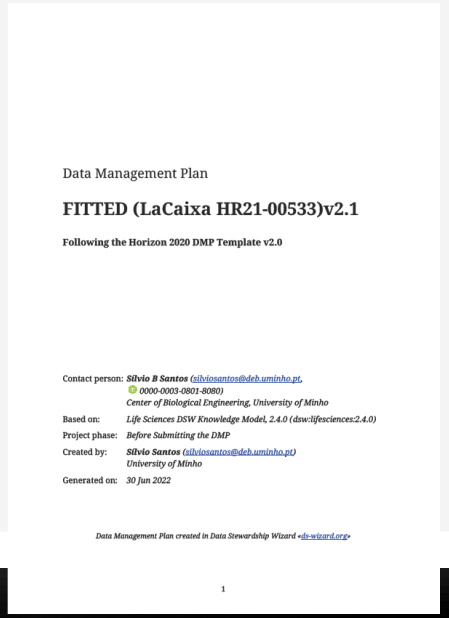
ICVS has a license in accordance with European guidelines for the care and use of animals for research purposes

Animals will be handled in accordance with Directive 2010/63/EU of the European Parliament and of the Council on the Protection of Animals Used for Scientific Purposes

DMP tool

https://ds-wizard.org/data-management-plans





Data Management Plan

FITTED (LaCaixa HR21-00533)v2.1

Following the Horizon 2020 DMP Template v2.0

Contact person: Silvio B Santos (silviosantos@deb.uminho.pt, 0000-0003-0801-8080) Center of Biological Engineering, University of Minho

Based on: Life Sciences DSW Knowledge Model, 2.4.0 (dsw:lifesciences:2.4.0)

Project phase: Before Submitting the DMP

Created by: Silvio Santos (silviosantos@deb.uminho.pt) University of Minho

Generated on: 30 Jun 2022

Data Management Plan created in Data Stewardship Wizard «ds-wizard.org»

1

Smart Questionnaires

The Data Stewardship Wizard provides a simple way to create the DMP by filling the Questionnaire in a smart way. What does "smart way" mean? Based on your previous answers in the Questionnaire, only relevant questions for your case will be shown and further followed.

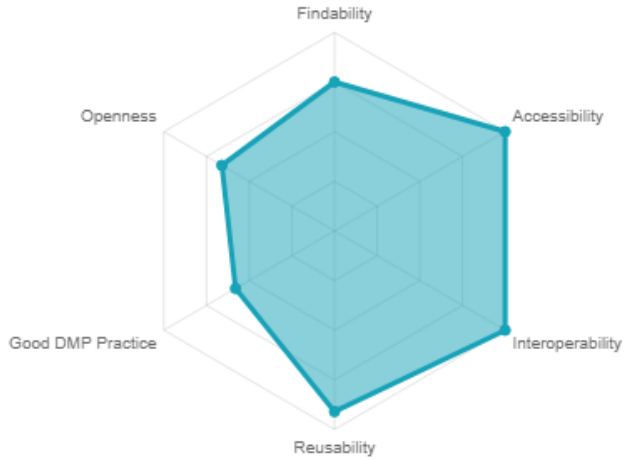
DMP tool

Summary Report

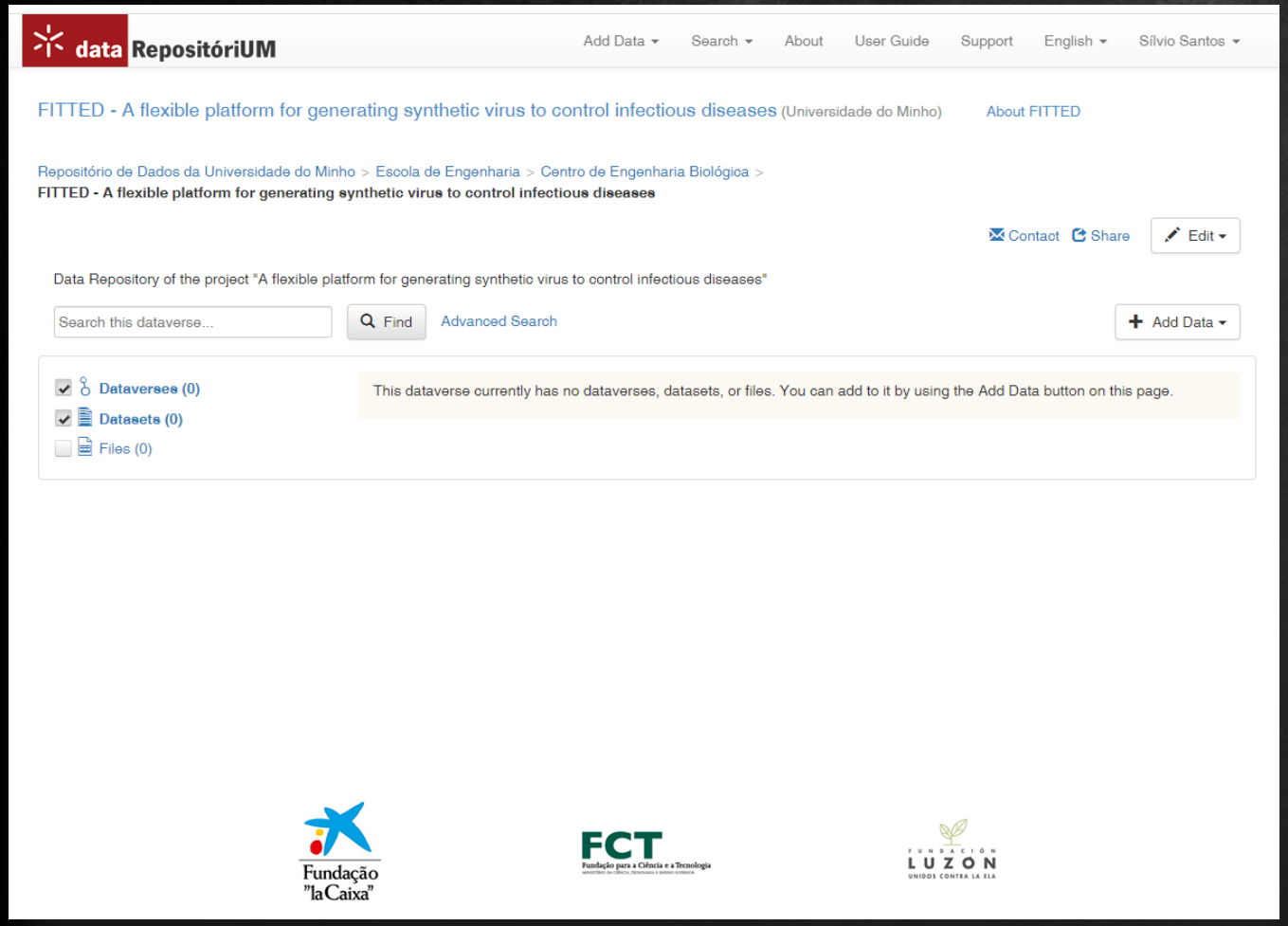
Answered (current phase): 193/193

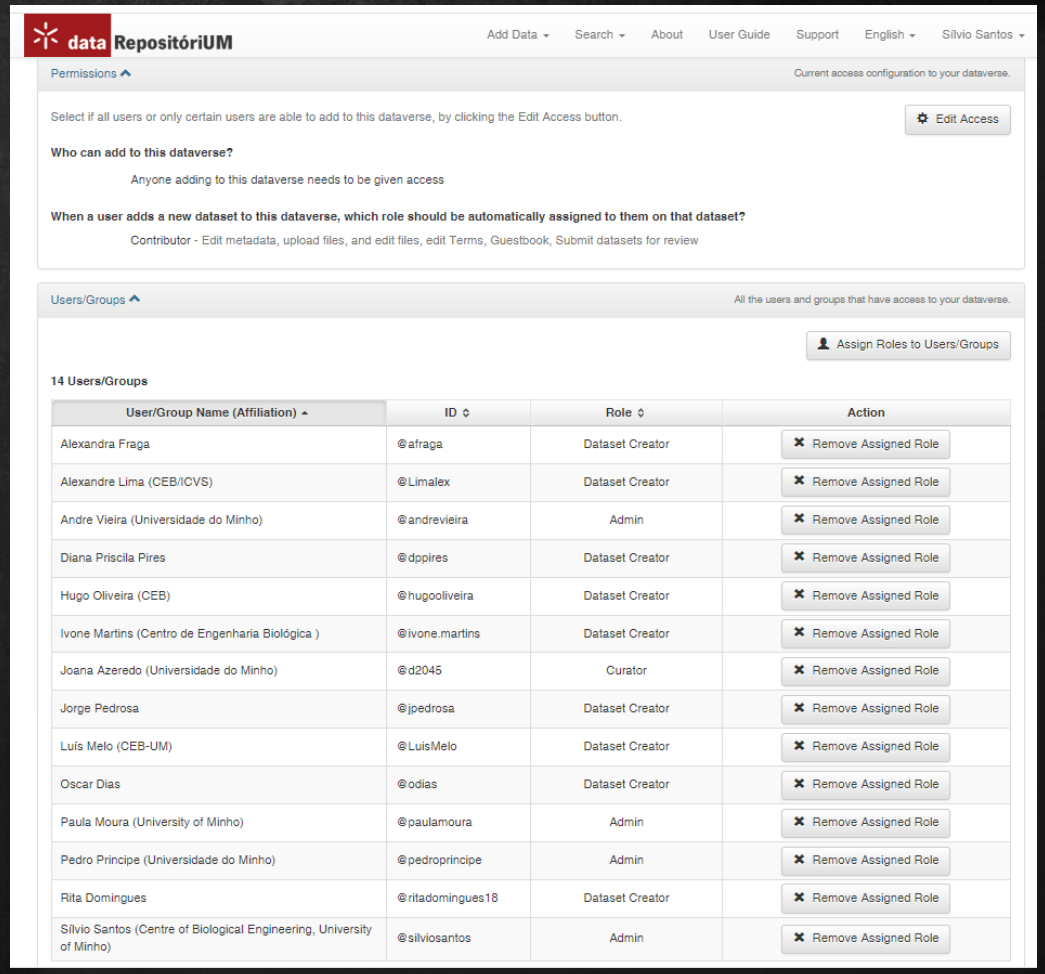
Answered: 342/349

Metric	Measure
Findability	0.75
Accessibility	1.00
Interoperability	1.00
Reusability	0.91
Good DMP Practice	0.58
Openness	0.66



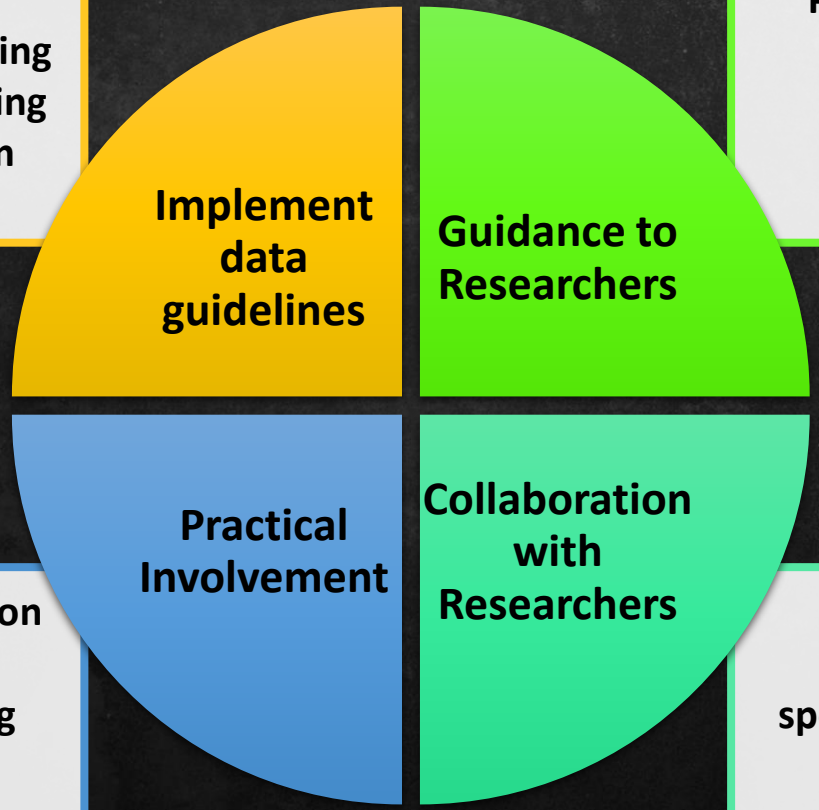
dataRepositoriUM





Data Steward | Responsabilities

Responsible for implementing specific data guidelines, translating them into the project and applying domain-specific procedures from data management plans



Provide guidance to researchers, ensuring that project data is handled in accordance with institutional policies

Responsibilities and tasks focus on supporting scientists in project data needs, as well as addressing infrastructure and service requirements.

Works closely with researchers, often being associated with a specific scientific project, and may even engage in practical work within the project

Data Steward | SKILLS

Project Management

The ability to efficiently manage data projects from planning to implementation, ensuring delivery within established timelines and budgets

Technical Knowledge

A data manager needs a solid technical understanding of tools and technologies relevant to data management, such as databases, programming languages, and data analysis platforms.

Communication Skills

Strong communication skills to collaborate effectively with researchers, data scientists, and other stakeholders, fostering a culture of open communication.

Domain Knowledge

In-depth understanding of the specific scientific domain, including terminology, methodologies, and research processes.

Problem-Solving

Being able to address complex challenges related to data quality, data integration, and technical issues that may arise during research.

Data Steward | SKILLS

Training and Development

Ease in training and developing the team in effective data management practices, fostering a data-centric culture within the organization.

Data Curation

Expertise in curating and maintaining high-quality datasets, ensuring data integrity and reliability for scientific analysis.

Data Governance

Ability to establish and enforce data governance policies, ensuring compliance with ethical standards, privacy regulations, and industry best practices.

Data Analysis and Interpretation

Capacity to critically analyze data, identifying patterns, trends, and insights relevant to the research.

Data Analysis Support

Understanding the needs of researchers and providing support in data analysis, including helping with data queries, interpretation, and troubleshooting.

Data Steward | SKILLS

Data Documentation

Thorough documentation of datasets, methodologies, and data processing steps to enhance transparency and reproducibility of scientific findings.

Metadata Management

Proficiency in creating and managing metadata, facilitating the discoverability and understanding of datasets by researchers.

Sensitive Data Management

The ability to handle sensitive and confidential data, ensuring compliance with regulations and ethical practices, is crucial.

Regulatory Knowledge

Staying updated on data protection regulations, research ethics, and legal requirements to ensure compliance.

Data Privacy and Security

Expertise in ensuring the privacy and security of sensitive scientific data, implementing measures to safeguard against unauthorized access or data breaches.

Data Steward | CHALLENGES

Undefined Career

Data Steward is a relatively new profession (~2017). The job profile is not always formally defined and the career-track may be unclear.

Data Lifecycle Management

Managing the entire data lifecycle, from data acquisition to archiving, and ensuring proper documentation at each stage.

Up-to-date Management

Handling the evolution of research methodologies, adapting to changes in the data environment and team needs.

Keeping Abreast of Technology

Staying updated on advancements in data management technologies and tools relevant to scientific research.

Data Steward | CHALLENGES

Constant Support

Researchers expect support and advice during the whole data cycle at all time.

Data Integration/ Complexity

Dealing with the complexity of integrating data from formats, structures, and sources, ensuring consistency and interoperability.

Interdisciplinary Collaboration

Overcoming challenges related to collaboration among researchers from different scientific disciplines, each with unique data requirements and practices.

Data Quality Assurance

Addressing challenges related to maintaining and assuring the quality of scientific data, including validation, cleaning, and error handling.

Data Steward | CHALLENGES

Adherence to Standards

Ensuring adherence to data standards and promoting best practices for data management and sharing within the scientific community.

Ethical and Legal Compliance

Navigating the ethical and legal landscape of scientific data, ensuring compliance with institutional review board approvals and other regulatory requirements.

Data Security

Ensuring proper security and protection of data, especially when dealing with sensitive and personal information.

Resource Allocation

Efficiently allocating resources, including time and budget, to support data management activities while balancing the needs of multiple research projects.

Data Steward | SWOT analysis

STRENGTHS (S)

Domain Expertise: Deep understanding of the scientific domain enhances the ability to effectively manage and curate domain-specific datasets.

Data Governance Skills: Maintains ethical standards compliance, privacy regulations, and industry best practices, ensuring data integrity and reliability.

Effective Communication: Facilitates collaboration with researchers, data scientists, and stakeholders, fostering a transparent and open data culture.

Data Analysis Support: Adds value by assisting researchers with queries, interpretation, and troubleshooting, enhancing the overall research process.

WEAKNESSES (W)

Resource Dependency: Limited budget and time, may be a weakness, especially if constraints limit the execution of comprehensive data management practices.

Technological Challenges: The fast-paced evolution of data management technologies may pose challenges in keeping abreast of the latest tools and ensuring that the infrastructure meets current standards.

THREATS (T)

Data Security Risks: Security threats poses risks to the security of sensitive scientific data, requiring constant vigilance and robust security measures.

Regulatory Changes: Rapid changes in data protection/privacy regulations pose challenges in adaptation, affecting data management practices.

Data Quality Assurance Challenges: May be threatened by challenges related to validation, cleaning, and handling errors in complex datasets.

Competing Priorities: Balancing diverse needs of multiple projects threatens resource allocation, impacting the quality of data stewardship.

OPPORTUNITIES (O)

Interdisciplinary Collaboration: Provides an opportunity for DS to bridge gaps between different scientific disciplines, facilitating broader research initiatives.

Data Standards Advocacy: Opportunities for DS to advocate for the establishment of data standards, enhancing consistency and interoperability.

Advancements in Technology: Embracing advancements in data management technologies can improve efficiency and effectiveness in handling scientific data.

CONCLUSION

Data Stewards

Have The Worst Seat At The Table

CONCLUSION

Integral Support in Scientific Communities

- Data Stewards play a vital role in scientific communities, actively supporting and promoting recommended data management practices.

Time-Efficient Assistance for Researchers

- Their contribution significantly saves researchers time by providing essential support in data management efforts, allowing scientists to focus more on their core research.

Informal Career Path and Background

- Data Stewards typically lack a formal career path or specific qualifications but bring valuable insights with their background in scientific or research fields.

Data Stewards are indispensable

www.ceb.uminho.pt/bbig

University of Minho
School of Engineering

CENTRE OF BIOLOGICAL ENGINEERING

AZEREDO LAB

DATASTEWARD

IN SCIENCE

SKILLS AND CHALLENGES

Sílvio Santos | silviosantos@deb.uminho.pt

FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO

- Nov -
15