



HAL
open science

PARSEC Data and Digital Output Management Plan and Workbook

Shelley Stall, Alison Specht, Pedro-Luiz-Pizzigatti Corrêa, Romain David,
Rorie Edmunds,, Jeaneth Machicao, Laurence Mabile, Margaret O'brien,
Lesley Wyborn, Danton-Ferreira Vellenich, et al.

► **To cite this version:**

Shelley Stall, Alison Specht, Pedro-Luiz-Pizzigatti Corrêa, Romain David, Rorie Edmunds,, et al.. PARSEC Data and Digital Output Management Plan and Workbook. BELMONT Forum; AGU; ERINHA. 2023, <https://zenodo.org/record/8231464>. hal-04180992

HAL Id: hal-04180992

<https://hal.science/hal-04180992>

Submitted on 14 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Data and Digital Output Management Plan and Workbook for the Belmont Forum

Collaborative Research Action (CRA) Science-driven e-Infrastructure Innovation (SEI) for the Enhancement of Transnational, Interdisciplinary and Transdisciplinary Data Use in Environmental Change

Project: Building New Tools for Data Sharing and Reuse through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas (PARSEC)

Version 3.0 August 2023

DOCUMENT CHANGE LOG

Version, year, DOI	Editors	Features
V1.0, 2022 https://doi.org/10.5281/zenodo.3891427 Concept DOI for all versions: https://doi.org/10.5281/zenodo.3891426	all	Initial doc
V2.0, 2021 https://doi.org/10.5281/zenodo.4910115	Stall, Specht	<ol style="list-style-type: none"> 1. Added links to tracking sheets for brief reference 2. Included Jupyter notebooks as a type of software. 3. Updated text for clarity in OP - Assigning usage licenses to data and digital outputs 4. Corrected URLs
V3.0, 2023 https://doi.org/10.5281/zenodo.8231464	Specht, O'Brien, Stall	<ol style="list-style-type: none"> 1. Added the Document Change Log 2. Refined description and links to Tracking Sheets for dataset and software 3. Changed the name of the Other Digital Output Tracking Sheet to: Non Peer-reviewed Digital Outputs Tracking Sheet. 4. Added three Operational Procedures Sections (OP): OP - Support for inclusive collaboration and authorship for data and digital outputs; OP - Giving attribution for data and digital objects previously created or created by others; OP - Project Backup and Closeout. 5. Added the D(DO)MP graphic 6. Added the PARSEC Code of Conduct 7. Updated Table of Contents 8. Added new section providing author detail and information about the PARSEC project. 9. Expansion of text in OP - Ensuring quality of digital objects during the life of the project and prior to preservation. 10. Added in three guidance documents created by the PARSEC Data Strand team in: OP - Selecting datasets for preservation; OP - Support for inclusive collaboration and authorship for data and digital outputs; PARSEC Tracking Sheets and Checklists

INTRODUCTION

This Data and Digital Output Management Plan and Workbook (DDOMP) is used by the PARSEC team to establish the policies the team will follow, document the operational procedures necessary to comply with those policies, and the planned activities necessary to manage PARSEC data, software, and other digital outputs. This is a working document.

This document describes activities necessary during the PARSEC research lifecycle as well as those necessary to preserve all digital outputs for use into the future. It is the intent of the PARSEC team to make our digital outputs as open as possible, discoverable, accessible, well-documented to promote the broadest reuse. These elements are both recommended and required by the **Belmont Forum Open Data Policy and Principles**¹ and the **FAIR Guiding Principles**².

The Belmont Forum Open Data Policy and Principles state that:

Datasets should be:

- Discoverable through catalogues and search engines
- Accessible as open data by default, and made available with minimum time delay
- Understandable in a way that allows researchers—including those outside the discipline of origin—to use them
- Manageable and protected from loss for future use in sustainable, trustworthy repositories

The FAIR Guiding Principles state that data should be Findable, Accessible, Interoperable, and Reusable.

The Table of Contents (page 3) lists the operational procedures addressed. In each procedure we provide reference to the specific elements requested by the Belmont Forum to support the review process.

For an understanding of repository process steps for data preservation referenced in this DDOMP, we recommend the one preferred by our selected preservation repository, the Environmental Data Initiative, which is certified as a Trustworthy Data Repository by the CoreTrustSeal Standards and Certification Board

Glossary of terms: A well-managed glossary that partners with the Research Data Alliance, a PARSEC partner, is the CASRAI Research Data Management Glossary. Currently hosted by CODATA.

PARSEC DDOMP Workbook Checklist

A short PARSEC checklist is used by our team members to help prompt the use of common tools, and necessary reporting and tracking requirements for scholarly publication, datasets, software, and other digital outputs.

Stall, Shelley, Specht, Alison, Corrêa, Pedro Luiz Pizzigatti, David, Romain, Edmunds, Rorie, Mabile, Laurence, Machicao, Jeaneth, O'Brien, Margaret, Wyborn, Lesley, Vellenich, Danton Ferreira, Miyairi, Nobuko, & Murayama, Yasuhiro. (2023). PARSEC DDOMP Workbook Checklist. Zenodo. <https://doi.org/10.5281/zenodo.4909851>

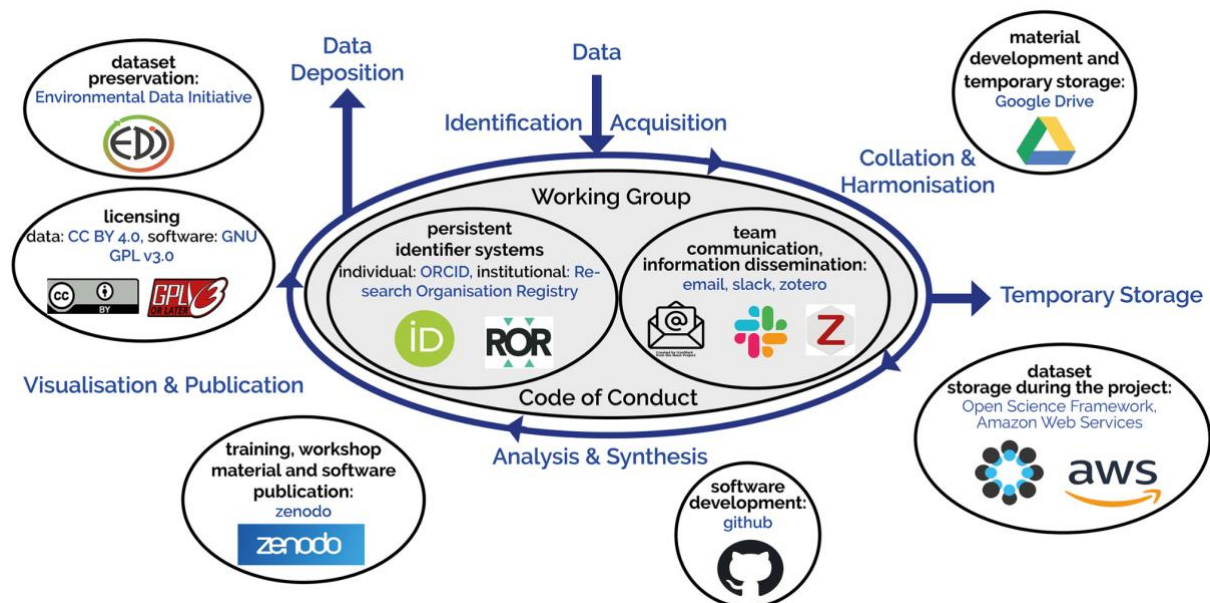
¹ <http://www.belmontforum.org/about/open-data-policy-and-principles/>

² <https://www.nature.com/articles/sdata201618>.



TABLE OF CONTENTS

Document Change Log	0
Introduction.....	1
Table of Contents	2
PARSEC Team Member Roles.....	3
PARSEC - Operational Procedures (OP)	4
OP1 - Selecting datasets for preservation.	4
OP2 - Protecting digital objects during the life of the project.....	5
OP3 - Ensuring quality of digital objects during the life of the project and prior to preservation.	7
OP4 - Ensuring data security, privacy, and intellectual property restrictions associated with datasets and digital outputs are honored and preserved in derivative products.....	8
OP5 - Support for inclusive collaboration and authorship for data and digital outputs.	8
OP6 - Giving attribution for data and digital outputs previously created or created by others. ..	9
OP7 - Assigning usage licences to data and digital outputs.	9
OP8 - Preparing supporting documentation (i.e., metadata) for data and digital outputs that supports public access, discovery, and long-term reuse.....	10
OP9 - Selection of a digital object Preservation Repository	11
OP10 - Sensitive or Restricted Data	13
OP11 - Determining costs associated with long-term data management.	13
OP12 - Project backup and closeout.	14
PARSEC Tracking Sheets	14
Tracking Sheet: Scholarly Publications.....	15
Tracking Sheet: Non Peer-Reviewed Digital Outputs.....	15
Tracking Sheet: Dataset Log - working log of datasets reviewed for use by the project.	15
Tracking Sheet: Software Log - working log for software reviewed by the project	15
PARSEC Code of Conduct	16
Authors and PARSEC Resources	18



PARSEC TEAM MEMBER ROLES

Belmont Forum Rubric Reference: 3.0: Team Members responsible for developing, implementing, overseeing, and updating the Data and Digital Outputs Management Plan

The Principal Investigators (PI) for each of the PARSEC strands are responsible for the development, implementation, oversight, and maintenance of the Data and Digital Output Management Plan and Workbook. The type of outputs are different for each strand. This collaboration allows for all types of digital objects to be addressed.

- Responsible for Synthesis Science Digital Content Used and Developed
David Mouillot, Synthesis Science Strand PI
david.mouillot@umontpellier.fr
[0000-0003-0402-2605](tel:0000-0003-0402-2605)
- Responsible for Data Science Strand Digital Content Used and Developed
Shelley Stall, Data Science Strand PI
sstall@agu.org
[0000-0003-2926-8353](tel:0000-0003-2926-8353)
- Responsible for both strands complying with the DDOMP operating procedures and checklists
Nicolas Mouquet, PARSEC PI
nicolas.mouquet@cnrs.fr
[0000-0003-1840-6984](tel:0000-0003-1840-6984)
Alison Specht, PARSEC PI
a.specht@uq.edu.au
[0000-0002-2623-0854](tel:0000-0002-2623-0854)
- Responsible for Complying with Belmont Forum Open Data Policy and Principles
All team members.



All PARSEC team members have signed a Code of Conduct document concerning data and digital output management that states “Derived data and digital outputs generated during PARSEC activity will be documented and released at the time of publication where possible into the public domain in compliance with Belmont Forum requirements. Policies for broad access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements, metadata description, intended repositories, and so on will be clearly described in the Project’s Data and Digital Outputs Management Plan.”

Signed PARSEC Code of Conduct forms from all team members are kept on file and available for review.

Responsibility following the end of the project for data and digital objects created by the team is entrusted to the selected preservation repositories. Selected repositories have indicated a commitment to maintaining the preserved objects and their persistent identifiers in support of the scientific record.

All PARSEC team members are responsible to have and manage an ORCID profile that links to their digital products. This promotes credit and automated attribution.

PARSEC - OPERATIONAL PROCEDURES (OP)

This section includes the operational procedures or processes PARSEC leadership uses to determine the procedure for:

1. Selecting datasets requiring preservation
2. Protecting digital objects during the life of the project
3. Ensuring quality of digital objects during the life of the project and prior to preservation.
4. Ensuring data security, privacy, and intellectual property restrictions associated with datasets and digital outputs are honored and preserved in derivative products.
5. Support for inclusive collaboration and authorship for data and digital outputs.
6. Giving attribution for data and digital outputs previously created or created by others.
7. Assigning usage licences to data and digital outputs.
8. Preparing supporting documentation (i.e., metadata) for data and digital outputs that supports public access, discovery, and long-term reuse.
9. Selection of a digital object Preservation Repository
10. Managing sensitive or restricted data
11. Determining costs associated with long-term data management.
12. Project backup and close out.

OP1 - Selecting datasets for preservation.

Belmont Forum Rubric Reference: 1.1 Datasets and other digital objects of long-term value are identified, including data type and encoding, 1.2 How the data and other digital objects will be collected, captured, or created.

A fundamental principle of the PARSEC project is to promote data reuse and will NOT create any new earth or environmental observation data. All raw data needed for our work already exists and includes household survey datasets and satellite imagery. Datasets significant to our research are logged in our [Dataset Tracking](#) tool as well as any processed or derived dataset.

PARSEC is using existing raw data that includes satellite imagery available at various resolutions, time-series, and costs from instruments associated with our country partners: US, Brazil, France, and Japan. To perform ground truth analysis with our models we are using local household surveys for towns and villages associated with our selected sites. We have identified country-specific sources to cover the time-frames and scale required. Depending on the selection of protected area sites, we are likely to need to access additional, already existing, survey data.

PARSEC output data will include global gridded socio-economic information and socio-economic dynamics specific to our study sites.

Our primary data and digital products planned for preservation include:

- Cleaned observation records and measurements
- Aggregated datasets subjected to quality control protocols, used for machine-learning and artificial intelligence models. This includes the provenance of what data was used to develop the aggregated dataset.
- Documentation of the quality control protocols used
- Algorithms developed to prepare data (cleaning scripts) and perform analyses
- Software supporting machine learning and artificial intelligence models
- Results of our analysis
- Training and workshop material

Dataset Preservation: If a dataset is used for analysis, it is to be preserved and made available prior to the end of the project or if more time is needed then before the final report is submitted to the Belmont Forum. We track this in the table where all datasets are logged. For any papers written during the PARSEC project, datasets used to support that research must be preserved and available at the time of publication.

Software Preservation: All software developed is to be preserved at each major release. This includes computational notebooks (e.g., Jupyter notebooks) and workflows.

Training and workshop materials: All materials are to be preserved on or near the time of the workshop.

The PARSEC team developed a checklist to help document, select, and preserve any team's research files. It provides:

1. How to ensure your project files have an effective backup plan in place.
2. Determine what digital objects to preserve and the supporting documentation.
3. Determine how open and FAIR to make your digital objects.

Stall, Shelley, Specht, Alison, O'Brien, Margaret, Machicao, Jeaneth, Corrêa, Pedro Luiz Pizzigatti, David, Romain, Edmunds, Rorie, Miyairi, Nobuko, Murayama, Yasuhiro, Santos, Solange, Wyborn, Lesley, Vellenich, Danton Ferreira, Vrouwenvelder, Kristina, & Mabile, Laurence. (2023). Digital Objects Preservation Checklist For Teams. Zenodo. <https://doi.org/10.5281/zenodo.7402540>

OP2 - Protecting digital objects during the life of the project.

Belmont Forum Rubric Reference: 4.1 Describe security measures for the data and other digital outputs to prevent unauthorized access with specific examples.

The PARSEC project has selected a number of tools to use during the project. We have elements of our project that we need to keep confidential during development. Other work benefits from being developed openly and in collaboration with a very broad community. Entities such as the Research Data Alliance, CODATA, and the World Data System allow participation in review opportunities of PARSEC research

output during co-located meetings. The PARSEC Data Strand provides support and expertise for the data management needs of the entire team.

For **material development and temporary storage**: Google Drive

We have organized the content allowing different permissions for leadership, team members, and other sub-groups.

Security measures: minimum protocols are in place. Links to the materials are easy to share across the team, partners, and other stakeholders.

For **team communications and information dissemination**: Email, Slack, Zotero

Security measures: minimum protocols are in place. The email list for the project is managed by the leadership team and is available for all team members to access. The Slack channel is available only to project members, partners, and other invited stakeholders. [PARSEC's group in Zotero](#) is open to those interested in the references gathered by contacting the PIs and being added to the group.

For **dataset storage during the project**: [Amazon Web Services](#) (AWS), [Open Science Framework](#) (OSF)

Our PARSEC [OSF workspace](#) is integrated with AWS as well as our Google Drive and GitHub. Each of these platforms supports access/controls for our team members to protect the project space.

Security measures AWS: AWS provides robust security measures for project teams. The full set of best practices can be found here: <https://aws.amazon.com/blogs/security/tag/best-practices/>

Security measures OSF: Each member of PARSEC must have an OSF login in order to manage access/controls. The project space can be kept private or public. Project files are backed up each night. More information here: <https://help.osf.io/hc/en-us/articles/360019930993-FAQs-Security>

For **dataset and supporting object preservation** (except for software and training/workshop materials): Environmental Data Initiative

The [Environmental Data Initiative](#) (EDI, independently funded by the U.S. National Science Foundation) is a partner on the PARSEC project, and a preferred domain repository. EDI promotes and enables curation and re-use of environmental data for large ecosystem-level research projects including the U.S. Long Term Ecological Research Program (LTER) and many field stations, individual laboratories and synthesis research projects. The EDI repository fully supports metadata and data immutability and strong versioning for auditing and reproducibility. Each dataset version receives a Digital Object Identifier (DOI), registered and resolved through DataCite. EDI has CoreTrustSeal certification and collaborates with other data professionals to operationalize FAIR principles. EDI is a signatory of the Enabling FAIR Data Commitment Statement.

Security measures: The EDI repository ensures stable access via servers hosted on the campus of the University of New Mexico (UNM) and connected via UNM's Science DMZ research network, DOE Energy Sciences Network (ESNet) and the Western Regional Network. EDI's data storage capacity is currently at 30TB (expandable), with an equivalent mirror storage for near-time backups. EDI also uses the AWS Glacier storage as a long-term "cold data" archive.

EDI's authentication procedures use organizational LDAP systems (Lightweight Directory Access Protocol), with mapping to 3rd-party identity service providers (Google, GitHub and ORCID), using the OAuth/OpenID Connect protocols. EDI metadata employs Ecological Metadata Language (EML), which includes access rules at the data package and data entity levels to support embargo when necessary. EDI advocates for open and unfettered access, but does not forbid providers from declaring more restrictive licensing agreements for use of their data packages. If providers do not include a statement of Intellectual Rights in EDI adds a default declaration based on the Creative Commons CC0 "No Rights Reserved" waiver. Details are posted: <https://environmentaldatainitiative.org/edi-data-policy-2/>

For **software development**: GitHub

GitHub allows a team of software developers to work together with version control, collaboration, and easy access. During the project the team site will be kept private. At the end of the project or if referenced in a paper, the site will be made open. The PARSEC GitHub repository is located at <https://github.com/PARSECworld>.

Security measures: GitHub is GDPR compliant and adheres to the [Privacy Shield Framework](#), certified January 26, 2017. More information on GitHub security can be found here: <https://github.com/security/trust>

For **software preservation**: Zenodo (using integrated link with GitHub)

Zenodo is a well known general repository that has specialized in software preservation through their partnership with GitHub.

Security measures: Our digital objects preserved in Zenodo are open and licensed CC BY 4.0. More information here: <https://about.zenodo.org/policies/>

For **training and workshop materials preservation**: Zenodo, [PARSEC Community](#)

Security measures: Our digital objects preserved in Zenodo are open and licensed CC BY 4.0. More information here: <https://about.zenodo.org/policies/>

For **credit and automated attribution**: PARSEC team members will have active and up-to-date ORCID profiles that allow for automated updates from Crossref and DataCite.

OP3 - Ensuring quality of digital objects during the life of the project and prior to preservation.

Belmont Forum Rubric Reference: none.

The PARSEC project has identified a set of criteria to ensure that the quality of datasets generated by the team through aggregation or analysis meets all the quality requirements and are consistent across the country teams.

Additionally, the team is working with managed international datasets that combine country data important to our work. We uncovered several issues when comparing the country data to the combined data that has led to a paper currently in development documenting this experience and bringing awareness to other researchers. Our hope from this paper is that we influence quality improvements in international datasets.

Processes to ensure quality and consistency include:

1. Methods for handling missing data
2. Methods for handling data at different granularities.
3. Methods for handling data using different time series.
4. Use of file naming standards.
5. Completion and accuracy of metadata.
6. Use of standard vocabularies.
7. Adequate capture of changes made to digital objects to maintain transparency.
8. Tracking of those involved to ensure an accurate list of creators/authors with their ORCIDs.
9. Correct levels of access/controls for sensitive/restricted data.
10. Compliance with all elements defined in the DDOMP.

OP4 - Ensuring data security, privacy, and intellectual property restrictions associated with datasets and digital outputs are honored and preserved in derivative products.

*Belmont Forum Rubric Reference: 7.1 Describe **intellectual property rights** to the data and other digital outputs.*

All PARSEC team members have signed a Code of Conduct document, as mentioned previously, that along with expectations about/for data and digital output management, also includes the following statements about how intellectual property rights will be handled:

Investigators will adhere to the appropriate standards for reporting the results of their scientific activities including respecting the intellectual property rights of others consistent with the European Code of Conduct for Research Integrity (2011) downloadable from:
<http://www.esf.org/coordinating-research/mo-fora/research-integrity.html>.

Derived data and digital outputs generated during PARSEC activity will be documented and released at the time of publication where possible into the public domain in compliance with Belmont Forum requirements. Policies for broad access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements, metadata description, intended repositories, and so on will be clearly described in the Project's Data and Digital Outputs Management Plan.

The Belmont Forum claims no intellectual property rights (including copyright) in any reports or data produced by the funded party except for summary reports on grant outcomes required by the Belmont Foundation or other bodies to whom the Belmont Foundation is required to provide reports.

Signed PARSEC Code of Conduct forms from all team members are kept on file and available for review.

OP5 - Support for inclusive collaboration and authorship for data and digital outputs.

Belmont Forum Rubric Reference: None.

Transnational collaboration. The creation and delivery of PARSEC digital outputs are to have representation from each of the countries involved in the project as much as possible. This requires that PARSEC team members develop trusted relationships across each of the country teams and be supported in achieving this effectively.

PARSEC team members will, for each nominated task/effort:

- Determine lead participants,
- Agree to the roles of each in the task/effort,
- Ensure practices are compliant with the Code of Conduct,
- Determine 'friendly' team members for first revision/checking
- Determine a sharing schedule

The PARSEC team developed a guideline to establish for your team their inclusive and collaborative expectations, commonly referred to open science practices. These can be codified in your team's Code of Open Science Practice.

1. How to encourage open behaviors with your research team such as collaboration, transparency, and working openly.

2. Which responsibilities and tasks to assign for open science.
3. How to establish consistent communication that facilitates common understanding and flexibility.
4. What and how to share your digital objects openly.

Stall, Shelley, Specht, Alison, O'Brien, Margaret, Machicao, Jeaneth, Corrêa, Pedro Luiz Pizzigatti, David, Romain, Edmunds, Rorie, Miyairi, Nobuko, Murayama, Yasuhiro, Santos, Solange, Wyborn, Lesley, Vellenich, Danton Ferreira, Vrouwenvelder, Kristina, & Mabile, Laurence. (2023). Open Science Practices for Teams. Zenodo. <https://doi.org/10.5281/zenodo.7402075>

OP6 - Giving attribution for data and digital outputs previously created or created by others.

Belmont Forum Rubric Reference: None.

PARSEC team members are to give attribution to all data and digital objects used to support the PARSEC project that were previously created by team members, or created by others. This ensures proper credit to the creators of those digital objects. By signing the [Code of Conduct](#), PARSEC team members acknowledge this requirement.

It should be noted that adherence to this established code is reliant on the ethical practice that research data and software must be given the same attribution as peer-reviewed publications.

OP7 - Assigning usage licences to data and digital outputs.

*Belmont Forum Rubric Reference: 7.2 Describe **licensing** of data and other digital outputs.*

PARSEC digital outputs are to be as **open as possible and closed as necessary**. PARSEC team members will follow these steps to ensure reuse of datasets and other digital objects or for created digital objects.

Existing Datasets or Digital Objects:

Step 1 - determine the terms of use and licensing of datasets (of digital objects) being used by the team. Determine if there are additional country or regional laws that must be considered, or indigenous sovereignty governance required of the data.

PARSEC team members will:

- Give credit to the dataset authors
- Be responsible with the data
- Share what is learned
- Comply with the dataset license or waiver
- Understand and follow any restrictions or regulations

Step 2 - track the usage license of datasets being aggregated to determine how the final derived data product can be licenced. Derived data products (to which the project has added intellectual content) will cite the original source(s) in the metadata showing provenance.

PARSEC Created Datasets/Digital Objects (not software):

Step 1 - determine if any usage or access/control requirement must be followed. Otherwise, the dataset is as open as possible with the ability to create derived data products. Creative Commons CC BY 4.0 is the default for the project, but CC0 for US government agencies as required along with attribution.

Step 2 - confirm that the selected repository supports the desired access/controls. The digital object must have the most open license possible. Creative Commons CC BY 4.0 or CC0

Software: the license PARSEC will use is: **GNU GPLv3** which allows others to do almost anything they want with the software, *except* distributing closed source versions. It's important to note the Creative Commons licensing [does not contain specific terms about the distribution of source code](#) and should not be used for software especially if you intend usage by others to remain open.

Training and Workshop Materials: the license used is Creative Commons CC BY 4.0.

Note:

CC0 is a waiver and allows any type of usage of the object assigned. This includes access, mining, redistribution & reuse.

CC BY 4.0 applies internationally and requires all derivative works to acknowledge all contributors to all works from which they are derived. Digital objects can be accessed, mined, redistributed, and reused as long as attribution is indicated.

References used by the team making these determinations:

Creative Commons Licenses: <https://creativecommons.org/about/cclicenses/>

RDA-CODATA Legal Interoperability Interest Group: Legal Interoperability of Research Data: Principles and Implementation Guidelines: https://rd-alliance.org/sites/default/files/attachment/Legal%20Interoperability%20Principles%20and%20Implementation%20Guidelines_Final.pdf

CARE Principles for Indigenous Data Governance: <https://www.gida-global.org/care>

Legal Side of Open Source: <https://opensource.guide/legal/>

How to License Research Data: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>

Guidance for selecting open source license for software: <https://choosealicense.com>

OP8 - Preparing supporting documentation (i.e., metadata) for data and digital outputs that supports public access, discovery, and long-term reuse.

*Belmont Forum Rubric Reference: 8 Describe **supporting documentation and metadata** that will be created to make data and digital outputs publicly accessible, discoverable, and reusable.*

PARSEC will follow the community standards for each product type.

For data and supporting products preserved in the Environment Data Initiative repository, each dataset package includes metadata in the Ecological Metadata Language (EML, <https://eml.ecoinformatics.org/>).

For software preserved in Zenodo we use the Force11 recommendation of [CodeMeta](#), with support from the online tool to [generate CodeMeta Files](#).

For training and workshop material preserved in Zenodo we use metadata criteria defined in the Data Management Training Clearing House: <https://dmtclearinghouse.esipfed.org>. We will also register the materials in the clearinghouse.

For any digital product not included above we use metadata criteria defined by the selected repository, specifically Zenodo, enhanced as needed depending on product type.

All digital products, no matter the repository, will include the following core metadata:

Provided by submitter:

- Creator/Author(s) (with given name, family name, ORCID, and affiliation/ROR)
- Title

- Resource Type (e.g. paper, pre-print, dataset, software, etc.)
- Contributors (with given name, family name, ORCID and affiliation/ROR)
- Date of submission
- Related identifier and relationship (if it applies)
- Version (if it applies)
- Description/Abstract
- Geolocation information (if it applies)
- Key parameters and units used
- Size of file (if it applies)

Controlled by repository:

- Persistent Identifier (e.g. Digital Object Identifier)
- Publisher or Repository Name
- Publication Date
- Recommended citation

To enable automated attribution and credit, ORCIDs for authors will be included with all products preserved in repositories. Where possible we will also include affiliation/institutional IDs (e.g., ROR) as well as Funder IDs and Grant IDs.

OP9 - Selection of a digital object Preservation Repository

Belmont Forum Rubric Reference: 2.3 Name of the Repository; 5.1 Repository T&C and/or sustainability plan / cost, 5.2 Entity responsible for managing long-term accessibility

The PARSEC team used the following guidelines when selecting repositories for our digital output preservation.

Desirable Repository Characteristics:

1. Supportive of the [Belmont Forum Open Data Policy and Principles](#) - clipped here:
 - a. Discoverable through catalogues and search engines
 - b. Accessible as open data by default, and made available with minimum time delay
 - c. Understandable in a way that allows researchers—including those outside the discipline of origin—to use them
 - d. Manageable and protected from loss for future use in sustainable, trustworthy repositories
2. Compliance with the [Enabling FAIR Data Commitment Statement](#) for Repositories - clipped here:
 - a. Ensure that research outputs (e.g., data, software, technology, and physical samples) curated by repositories are open and FAIR, have essential documentation, and include human-readable and machine-readable metadata (e.g., on landing pages) in standard formats that are exposed and publicly discoverable.
 - b. Ingest and expose data to promote interoperability and reuse.
 - c. Ensure that unique, persistent identifiers are used for authors (e.g., [ORCID](#)), research objects (e.g., [Digital Object Identifier](#)), and physical samples (e.g., [IGSN](#)).
 - d. Create associations among the research outputs that they manage and other related entities.

- e. Ensure that data and software have licenses that are as open as possible, and as protected as necessary.
 - f. Support peer-review of related manuscripts by enabling access to the research outputs prior to publication.
 - g. Gain third-party validation of trustworthy and sustainable practices and capabilities (e.g., CoreTrustSeal).
3. Dataset or file size limits can support PARSEC requirements.
 4. Costs for preservation during the project are within the PARSEC budget
 - a. PARSEC country members have accounted for data management in their budgets. Each country-lead monitors their budget and usage.
 5. Has access and controls necessary to protect sensitive data.
 6. Indicates in the terms of use (or similar) their commitment to long-term preservation and persistence of the digital objects in their care as part of the scientific record.
 7. For domain repositories, has a third-party certification of services such as CoreTrustSeal (for domain repositories only) such as a [World Data System Regular Member](#) compliant with their [Data Sharing Principles](#).
 8. Uses English as a primary language, or has an English version that allows for searching, and other discovery activity in English. [Note: Journals prefer that preserved datasets to be in a repository that supports the English language.]
 9. Is compliant or working towards compliance with the [TRUST Principles for digital repositories](#)³.
 10. Complies with all PARSEC funder requirements

PARSEC Funder Repository Requirements – Reference Information:

São Paulo Research Foundation, FAPESP (São Paulo, Brazil): This workbook follows the FAPESP guidelines that recommends to deposit in repositories where the data will be maintained and preserved according to the standards of each scientific domain, under the ethical and security constraints defined by the project and its associated domains. FAPESP has been working since 2017 in the open research data repositories that will be considered as soon as available (<http://www.fapesp.br/openscience/en>).

<http://www.fapesp.br/openscience/en>

Agence Nationale de la Recherche, ANR (France): This workbook aligns with ANR's [Open Science Policy](#). In it they reference "[Practical Guide to the International Alignment of Research Data Management](#)" which has an extensive list of guidance for selecting trustworthy repositories (pp. 26-30).

National Science Foundation, NSF (US) - The NSF currently does not have guidance for selecting repositories. In NSF 19-069, [Dear Colleague Letter: Effective Practices for Data](#) they indicated the importance of data having persistent identifiers. This workbook supports that best practice.

Japan Science and Technology Agency, JST (Japan) - This workbook is aligned with [Implementation Guidelines: JST Policy on Open Access to Research Publications and Research Data Management](#). Here it is indicated that "It is preferable that research data be made openly available by depositing them in an existing public database or common disciplinary repositories established by academic societies to facilitate reuse of research results. If there is no

³ Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020). <https://doi.org/10.1038/s41597-020-0486-7>

appropriate public database, institutional repositories could also be used. In some cases, the academic journals to which research publications are submitted suggest designated repositories to deposit data.”

OP10 - Sensitive or Restricted Data

Belmont Forum Rubric Reference: 6.1 - Describe how sensitive data and other digital outputs are made available beyond the project team, 6.2 - Describe any limitations on the ability to share data and other digital outputs.

PARSEC is using socio-economic data sourced from the Living Standards Measurement Study (LSMS) surveys, conducted with technical assistance from the World Bank. This source (and other global sources) have proved not to be frequent enough or of the correct scale in most circumstances, so country-based census-derived sources have been used, for example, the Brazilian Institute of Geography and Statistics (IBGE) in Brazil.

These datasets include details about communities that are defined as **sensitive** but we have not needed to gain access to that level of data requiring additional access and control protections.

For data that does not have a specified data access agreement or terms of use:

Before sharing data, PARSEC team members will:

- Understand the Belmont Forum’s policies on data sharing, and any additional institution or country requirements.
- Only share data that has been anonymized in compliance with the laws of all countries associated with the PARSEC Grant.
- Only share images that have faces of individuals and other personal information (e.g., vehicle license plates) blurred.

Before using existing datasets, PARSEC team members will:

- Give credit in publications, presentations, and other materials
- Respect the assigned license
- Protect the datasets appropriately

The PARSEC team may need to access data associated with indigenous people. Should that occur, the team will consult the [CARE Principles](#) for Indigenous Data Governance. At the time of this report there is no implementation guidance, but we do expect that to be developed by CARE during the time of this project.

OP11 - Determining costs associated with long-term data management.

Belmont Forum Rubric Reference: 9 Describe costs or estimated costs associated with long-term data management or an assigned data manager.

Cost Description

Acquiring External Data Sources: For the PARSEC project we expected we might need to pay for satellite images but the research method has not yet required this source.

Quality Check Data: Develop and use common methods for mitigating quality challenges.

Prepare Metadata: Develop and use common templates for data and digital objects.

Data Storage and Backup During Project Duration: Ensure data are protected throughout the project. This includes cloud usage and storage costs.



Data Access and Security: All members of the team, from six countries, need access to the project data and digital objects.

Data Preservation: Includes conversion costs for preparing data.

Data Sharing and Reuse: Costs for copyright clearance or legal advice. Costs for curation of datasets.

The PARSEC team members are providing the above data management services as in-kind contributions except for the following items:

- Fees for satellite imagery (if needed) - the team from France is managing this budget item.
- Access to AWS - the team from Brazil has allotted \$11,000 USD per year to support the PARSEC team. This access started in June 2020.

OP12 - Project backup and closeout.

Belmont Forum Rubric Reference: None.

Determine prior to the end of the grant or project the actions needed to backup and preserve important artifacts not included in the preservation steps for digital outputs. This includes the project google drive, project website, and other working platforms. These platforms are meant to be working areas used by the team to support the research activities and create digital outputs. They are valuable to the team as a way to review prior work as well as to back up the team's efforts while the work is in progress.

Project Website: The <http://parsecproject.org/> will not be maintained after the Belmont Forum grant has ended. The project website content will be transitioned to the data.agu.org website as a project page as well as archived in the internet archive [Wayback Machine](#).

Project Google Drive: A backup is managed on our OSF site through the linked platforms feature.

PARSEC TRACKING SHEETS

Belmont Forum Rubric Reference: 1.1 - Types of data and other digital outputs of long-term value, 1.3 - How much data are anticipated, 2.1 - Metadata standards or formats for the data and other digital objects, 2.2 - When data and other digital outputs will be made available outside and within the project team, 2.3 - How data and other digital outputs will be made available beyond the project team, 6.2 - Limitations on the ability to share data and other digital outputs, 7.2 - licensing of the data and other digital outputs

This section provides links to all the digital objects created for PARSEC grouped by type in the different Tracking Sheets as well as the team Checklist prompting the time intervals for making sure these lists are updated.

This section includes tracking sheet information for:

1. Scholarly Publications
2. Non Peer-Reviewed Digital Outputs (to include Datasets, Software, Conferences, Posters, Presentations, Training Materials, Workshop Materials, etc.)
3. Datasets - working log of datasets reviewed for use by the project
4. Software - working log for software reviewed by the project (to include Models, Notebooks, Workflow, etc.)

Project tracking sheets can be found here:

Specht Alison, Stall, Shelley, O'Brien Margaret, Correa Pedro Luiz Pizzigatti, Murayma Yasuhiro, David Romain, Edmunds, Rorie, MABLE, Laurence, MACHICAO, Jeaneth, MIYAIRI, Nobuko, VELLENICH,

Danton Ferreira, Delbar, Vincent, Trammell, Jamie, Kondo, Yasuhisa (2023). PARSEC D(DO)MP. Project tracksheets: scholarly publications, non peer-reviewed digital outputs, dataset log, and software log. Zenodo. <https://doi.org/10.5281/zenodo.8232020>

The PARSEC team has developed guidance to streamline a team's research needs with resources that ensure teams are as efficient, collaborative and open as possible.

1. How to establish common team resources and digital object registry.
2. How to manage and track your team's digital objects.
3. How to ensure equitable team contributions to research outputs and provide appropriate credit.

Stall, Shelley, Specht, Alison, O'Brien, Margaret, Machicao, Jeaneth, Corrêa, Pedro Luiz Pizzigatti, David, Romain, Edmunds, Rorie, Miyairi, Nobuko, Murayama, Yasuhiro, Santos, Solange, Wyborn, Lesley, Vellenich, Danton Ferreira, Vrouwenvelder, Kristina, & Mabile, Laurence. (2023). Open Science Resources and Guidance for Teams. Zenodo. <https://doi.org/10.5281/zenodo.7402270>

Tracking Sheet: Scholarly Publications

Scholarly publications written for the purpose of reporting PARSEC project related research. These are peer reviewed publications as well as pre-prints.

PARSEC [Scholarly Publications Tracking Tool](#)

Tracking Sheet: Non Peer-Reviewed Digital Outputs

– to include Datasets, Software, Conferences, Posters, Presentations, Training Materials, Workshop Materials, etc.

Non Peer-Reviewed Digital Output developed by the PARSEC project that includes the date of identification, name, metadata standard used, available formats, name of repository, date of initial submission. These objects include posters, presentations, training materials, and workshop materials organized by event where they were presented. Reporting periods are tracked on separate tabs.

PARSEC [Non Peer-Reviewed Digital Output Tracking Tool](#)

Tracking Sheet: Dataset Log - working log of datasets reviewed for use by the project.

Datasets reviewed by the PARSEC project, including description, filename, metadata standard, licensing, available formats, name of repository, citation and persistent identifier (e.g., DOI).

Note: published PARSEC datasets are included in Non Peer-Reviewed Digital Output above.

PARSEC [Dataset Log](#)

Tracking Sheet: Software Log - working log for software reviewed by the project

– to include Models, Notebooks, Workflow, etc.

Software reviewed by the PARSEC project to include description, filename, metadata standard used (CodeMeta), licensing, language, name of repository, citation, persistent identifier (e.g., DOI).

Note: published PARSEC software are included in Non Peer-Reviewed Digital Output above.

PARSEC [Software Log](#)



PARSEC CODE OF CONDUCT

The undersigned participating member in PARSEC agrees to ensure activities are collaborative and interdisciplinary, the support of Belmont and the participating institutions and countries is acknowledged as appropriate, and any data generated are recorded and released for future use according to the Belmont Forum Open Data Principles « <http://www.belmontforum.org/about/open-data-policy-and-principles> ».

- Participants will not discriminate against others in the course of their work.
- Investigators will adhere to the appropriate standards for reporting the results of their scientific activities including respecting the intellectual property rights of others consistent with the **European Code of Conduct for Research Integrity** (2011) downloadable from: <http://www.esf.org/coordinating-research/mo-fora/research-integrity.html>.
- Authorship of outputs will be determined according to author contribution. When initiated, each output will have an agreement form signed by each author to establish authorship (see Appendix for sample form)
- Derived data and digital outputs generated during PARSEC activity will be documented and released at the time of publication where possible into the public domain in compliance with Belmont Forum requirements. Policies for broad access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements, metadata description, intended repositories, and so on will be clearly described in the Project's **Data and Digital Outputs Management Plan**
- The original data sets, the workflow, and the derived data products are all to be cited in any publication or project outcome
- Belmont will receive acknowledgement for its support of the research activities conducted by each funded party in **each publication or product** resulting from the activity. A model for this is shown below, and will be completed depending on the participating partners.

'This research is a product of the PARSEC group funded by the Belmont Forum as part of its Collaborative Research Action (CRA) on Science-Driven e-Infrastructures Innovation (SEI) + relevant research agencies + synthesis centre CESAB of the French Foundation for Research on Biodiversity.'
- Copies of outputs resulting from Belmont support will be provided to Belmont at the time of publication for record of achievements.
- Extenuating circumstances that may prevent full compliance will be discussed with Belmont Forum representatives.

Belmont claims no intellectual property rights (including copyright) in any reports or data produced by the funded party except for summary reports on grant outcomes required by the Belmont Foundation or other bodies to whom the Belmont Foundation is required to provide reports.

Signed by

.....

print name

date



AUTHORSHIP ACKNOWLEDGEMENT FORM TEMPLATE

Title of publication: _____

Type of publication: (please select from list)

	Book
	Book Chapter
	Journal Article
	Conference Paper or Poster
	Audio/Visual Recording
	Computer Software or model
	Design
	Patent
	Creative Work
	Research Report
	Workshop materials
	Webinars
	Reference Work

Proposed order of authors for this publication (add rows as required)

1.
2.
3.

Corresponding/designated senior PARSEC author: listed author number <insert number>

Country of corresponding/designated senior PARSEC author :

Details of substantive intellectual contribution* (add rows as required)

author 1	
author 2	
author 3	

Confirmation of agreement with listed authors and the order of authorship in the above publication.
(add rows as required)

Author 1	Signed	Date
	Name	
Author 2	Signed	Date
	Name	
Author 3	Signed	Date
	Name	

* Attribution of authorship depends to some extent on the discipline but, in all cases, must be based on substantive contribution and responsibility for at least one, and usually more than one, of the activities listed in the CRediT (Contributor Roles Taxonomy) guidelines: <https://credit.niso.org/>

**All authors should maintain a copy of this completed form for their records

AUTHORS AND PARSEC RESOURCES

STALL, Shelley; American Geophysical Union, ORCID: 0003-2926-8353

SPECHT, Alison; TERN, The University of Queensland, ORCID: 0002-2623-0854

CORREA, Pedro Luiz Pizzigatti; University of São Paulo, ORCID: 0002-8743-4244

DAVID, Romain; ERINHA, ORCID: 0003-4073-7456

EDMUNDS, Rorie; DataCite

MABILE, Laurence; University Paul Sabatier Toulouse III, ORCID: 0002-7724-1721

MACHICAO, Jeaneth; University of São Paulo, ORCID: 0002-1202-0194

MIYAIRI, Nobuko; National Institute of Information and Communications Technology, ORCID: 0002-3229-5662

MURAYAMA, Yasuhiro; National Institute of Information and Communications Technology, ORCID: 0003-1129-334X

O'BRIEN, Margaret; University of California, Santa Barbara, ORCID: 0002-1693-8322

WYBORN, Lesley; National Computing Infrastructure, Australian National University, ORCID: 0001-5976-4943

VELLENICH, Danton Ferreira; University of São Paulo, ORCID: 0002-3223-6996

Preferred Citation

Stall, Shelley, Specht, Alison, Corrêa, Pedro Luiz Pizzigatti, David, Romain, Edmunds, Rorie, Mabile, Laurence, Machicao, Jeaneth, Miyairi, Nobuko, Murayama, Yasuhiro, O'Brien, Margaret, Wyborn, Lesley, & Vellenich, Danton Ferreira. (2023). PARSEC Data and Digital Output Management Plan and Workbook. Zenodo. DOI: <https://doi.org/10.5281/zenodo.3891426>

Project Website: <http://parsecproject.org/>

Zenodo Community: <https://zenodo.org/communities/parsec/>