ResNet Data Management Plan

John Clark

2021-04-07

Contents

Start

suppress Start heading

versioning info

expand authorship

Chapter 1

Introduction

1.1 FAIR Guiding Principles

From the GO FAIR Initiative:

Findable
The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

Accessible
Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable
The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

Reusable
The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).

1.2 Roles

or responsibilities. Who is required to do what? How are they held accountable, and by who?

1.3 Workflows

1.3.1 Internal Data

1.3.2 External Data

Chapter 2

Standards

2.1 Identifiers

2.1.1 Researchers: ORCID

2.1.2 Data: DOI

2.1.3 Physical Samples: IGSN

2.2 Repositories

2.2.1 Portage

2.2.2 GLOBUS

2.2.3 ResNet Data Portal/GeoNode

2.3 Metadata

retitle to documentation?

Define metadata and significance

2.3.1 ISO 19115

Links to standard (http://rd-alliance.github.io/metadata-directory/standards/iso-19115.html)

Enumerate required fields

2.3.2 Tools

Metadata creation and validation tools

ESRI/ArcGIS:

Python:

https://pycsw.org/

https://github.com/geopython/pygeometa

R:

https://github.com/eblondel/geometa

Stand Alone:

Web:

2.4 Data Formats

2.4.1 Raster

2.4.1.1 Formats

2.4.1.1.1 geotiff

2.4.1.1.2 NetCDF

2.4.2 Vector

2.4.2.1 Shapefile

2.4.2.2 GeoJSON

2.4.3 Tabular

2.4.3.1 CSV

Chapter 3

Resources

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005510

https://docs.computecanada.ca/wiki/Research_Data_Management

https://earthdata.nasa.gov/esdis/eso/standards-and-references/data-product-development-guide-for-data-producers

https://daac.ornl.gov/datamanagement/

https://www.usgs.gov/products/data-and-tools/data-management/data-management-plans