BioData.pt

Ready for BioData Management?

How to Make a DMP

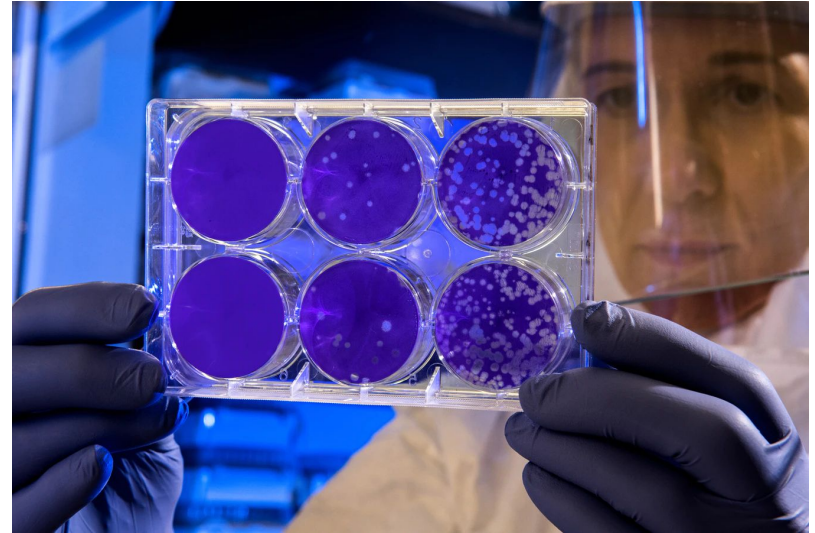**João Cardoso, Daniel Faria**

How to Make a DMP?

Learning Outcome 1:

Create a DMP according to the DMP Common Standard Model

BioData.pt

Ready for BioData Management?

Let's Consider a Mock Project

- **Project X**

    - Unveiling the mechanisms of Disease X

    - BioData.pt is applying for funding from the Fundação para a Ciência e Tecnologia (FCT).

Ready for BioData Management?

Let's Consider a Mock Project

- **Motivation**
  - The cause of **Disease X** has been recently discovered to be a **virus**, phage X.

  - It **infects** normal **gut bacteria** and leads them to become virulent and cause **chronic intestinal infection**.

  - This disease has been **spreading rapidly** in Europe, with **costs** in health-care reaching the **tens of millions of Euros**.
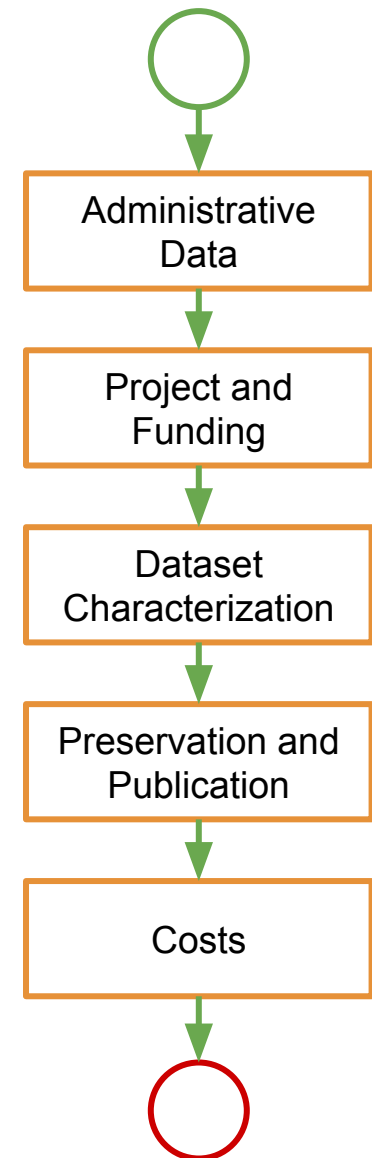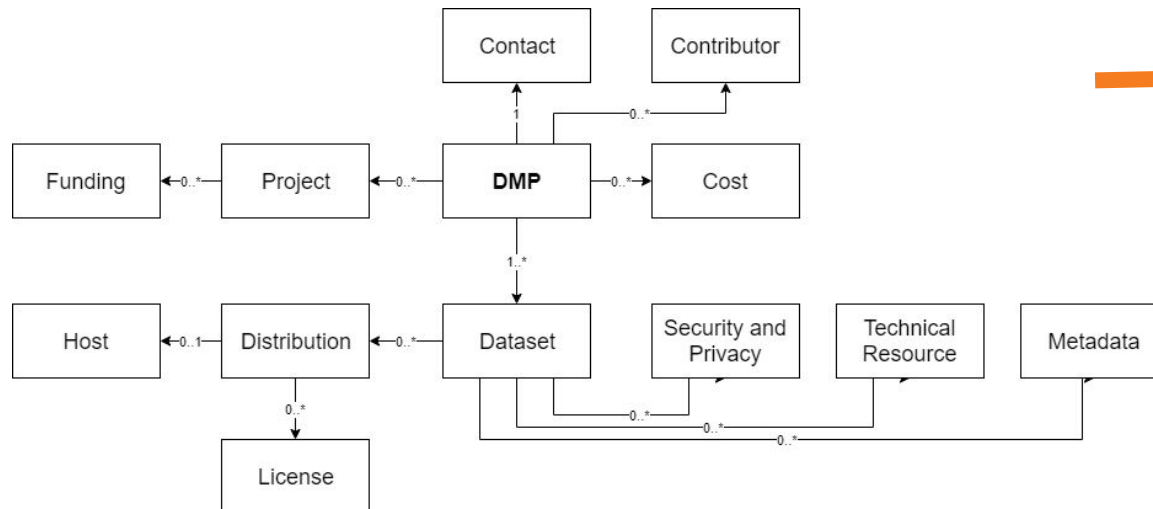
Let's Consider a Mock Project

- **Objectives**

  - **Uncover the mechanisms** of disease X by **sequencing phage X** and studying the **effects of** its **infection** in human gut microbiota at the **population and molecular level**.

  - **Assess which bacterial taxa** are **infected** by phage X and what are the **effects on** the **relative abundance** of the various taxa.

  - Study the **effects** on the **infected taxa** at the **gene expression level**.

  - **Improving treatment** for Disease X, and potentially being able to **cure it**.
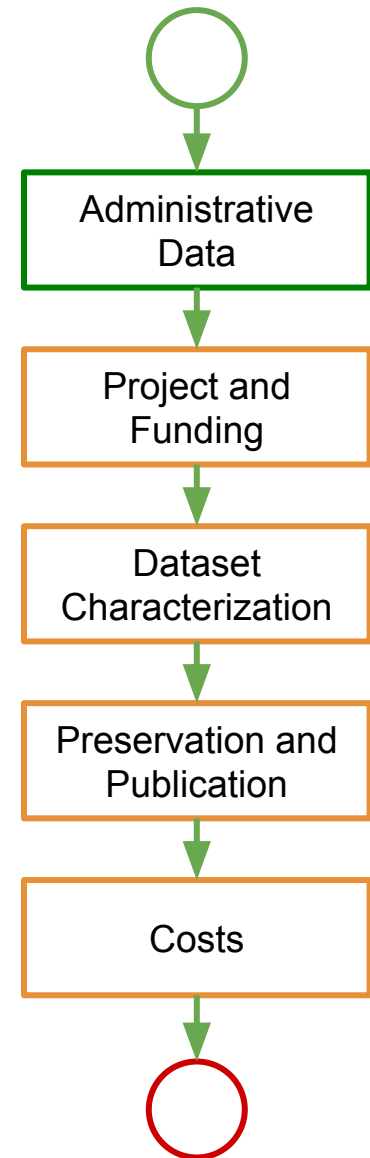
Creating a DMP

- The **DMP Creation Methodology** comprises 5 steps.

- Each step focuses on a **specific aspect** of the DMP.

- It is **based** on the **RDA's DMP Common Standards** metadata application standard.

Administrative Data

Project and Funding

Dataset Characterization

Preservation and Publication

Costs

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 1)

- **Step 1 - Administrative Data**

  - **Characterization** of the **DMP document**, and the **responsibilities** of all the **people** mentioned.

  - The information is classified in three sections:

    - **General information** characterizing the **DMP document**.

    - **Contact** (person of institution) for the DMP.

    - A listing of all **collaborators** and their **roles** in the DMP.

Administrative Data

Project and Funding

Dataset Characterization

Preservation and Publication

Costs

Ready for BioData Management?

BioData.pt

Creating a DMP (Step 1)

In the project

Project X (Application to Fundação para a Ciência e Tecnologia)

Title of the project: Unveiling the mechanisms of Disease X

Participants:

- Prof. Coor Dinator (coor.dinator@biodata.pt) [PI & DMP Coordinator]
- Dr. Dat Manger (dat.manger@biodata.pt) [Data Manager]
- Dr. Col Hector (col.hector@biodata.pt) [Clinical Data & Sample Collector]
- Dr. R. Sercher (r.sercher@biodata.pt) [Researcher]
- Mrs. A. D'Min (a.dmin@biodata.pt) [Project Manager]

Host Institution: BioData.pt

Start date: January 1st, 2021

Duration: 36 months

Ready for BioData Management?

Creating a DMP (Step 1)

In the project

Project X (Application to Fundação para a Ciência e Tecnologia)

Title of the project: Unveiling the mechanisms of Disease X

Participants:

- Prof. Coor Dinator (coor.dinator@biodata.pt) [PI & DMP Coordinator]
- Dr. Dat Manger (dat.manger@biodata.pt) [Data Manager]
- Dr. Col Hector (col.hector@biodata.pt) [Clinical Data & Sample Collector]
- Dr. R. Sercher (r.sercher@biodata.pt) [Researcher]
- Mrs. A. D'Min (a.dmin@biodata.pt) [Project Manager]

Host Institution: BioData.pt

Start date: January 1st, 2021

Duration: 36 months

- Generic information on the DMP document:
    - Title, institution, start date, duration, etc.
    - Description from the abstract.

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 1)

In the project

Abstract:

The cause of Disease X has been recently discovered to be a virus, phage X, which infects normal gut bacteria and leads them to become virulent and cause chronic intestinal infection. Although non-fatal, this disease has been spreading rapidly in Europe, with costs in health-care reaching the tens of millions of Euros.

This project aims to uncover the mechanisms of disease X by sequencing phage X and studying the effects of its infection in human gut microbiota at the population and molecular level. We will assess which bacterial taxa are infected by phage X and what effect the infection has on the relative abundance of the various taxa, as well as what e[ffect] infected taxa at the gene expression level.

The project will be a key step towards improving potentially being able to cure it.

- Generic information on the DMP document:
  - Description from the abstract.

Ready for BioData Management?

Creating a DMP (Step 1)

In the project

Project X (Application to Fundação para a Ciência e Tecnologia)

Title of the project: Unveiling the mechanisms of Disease X

Participants:

- Prof. Coor Dinator (coor.dinator@biodata.pt) [PI & DMP Coordinator]
- Dr. Dat Manger (dat.manger@biodata.pt) [Data Manager]
- Dr. Col Hector (col.hector@biodata.pt) [Clinical Data & Sample Collector]
- Dr. R. Sercher (r.sercher@biodata.pt) [Researcher]
- Mrs. A. D'Min (a.dmin@biodata.pt) [Project Manager]

Host Institution: BioData.pt

Start date: January 1st, 2021

Duration: 36 months

- **Contact** (person of institution) for the DMP.

Ready for BioData Management?

Creating a DMP (Step 1)

In the project

Project X (Application to Fundação para a Ciência e Tecnologia)

Title of the project: Unveiling the mechanisms of Disease X

Participants:

- Prof. Coor Dinator (coor.dinator@biodata.pt) [PI & DMP Coordinator]
- Dr. Dat Manger (dat.manger@biodata.pt) [Data Manager]
- Dr. Col Hector (col.hector@biodata.pt) [Clinical Data & Sample Collector]
- Dr. R. Sercher (r.sercher@biodata.pt) [Researcher]
- Mrs. A. D'Min (a.dmin@biodata.pt) [Project Manager]

Host Institution: BioData.pt

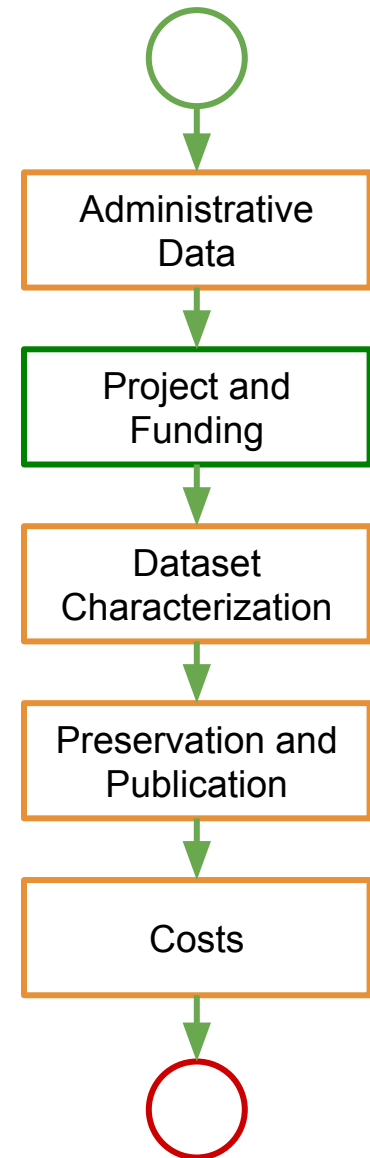**Start date:** January 1st, 2021

Duration: 36 months

- A listing of all **collaborators** and their **roles** in the DMP.

Creating a DMP (Step 2)

- **Step 2 - Project and Funding**

  - ○ **Characterization** of the **project(s)** and their sources of **funding**.

  - ○ The information is classified in two sections:

    - ■ Information regarding the **project(s)** to which the **DMP is associated**.

    - ■ Information pertaining the **funding** of a **particular project**.

Administrative Data

Project and Funding

Dataset Characterization

Preservation and Publication

Costs

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 2)

In the project

Project X (Application to Fundação para a Ciência e Tecnologia)

Title of the project: Unveiling the mechanisms of Disease X

Participants:

- Prof. Coor Dinator (coor.dinator@biodata.pt) [PI & DMP Coordinator]
- Dr. Dat Manger (dat.manger@biodata.pt) [Data Manager]
- Dr. Col Hector (col.hector@biodata.pt) [Clinical Data & Sample Collector]
- Dr. R. Sercher (r.sercher@biodata.pt) [Researcher]
- Mrs. A. D'Min (a.dmin@biodata.pt) [Project Manager]

Host Institution: BioData.pt

Start date: January 1st, 2021

Duration: 36 months

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 2)

In the project

Project X *(Application to Fundação para a Ciência e Tecnologia)*

Title of the project: Unveiling the mechanisms of Disease X

Participants:

- Prof. Coor Dinator (coor.dinator@biodata.pt) [PI & DMP Coordinator]
- Dr. Dat Manger (dat.manger@biodata.pt) [Data Manager]
- Dr. Col Hector (col.hector@biodata.pt) [Clinical Data & Sample Collector]
- Dr. R. Sercher (r.sercher@biodata.pt) [Researcher]
- Mrs. A. D'Min (a.dmin@biodata.pt) [Project Manager]

Host Institution: BioData.pt

Start date: January 1st, 2021

Duration: 36 months

- Information regarding the **project(s)** to which the **DMP is associated**.

Ready for BioData Management?

Creating a DMP (Step 2)

In the project

Project X (Application to Fundação para a Ciência e Tecnologia)

Title of the project: Unveiling the mechanisms of Disease X

Participants:

- Prof. Coor Dinator (coor.dinator@biodata.pt) [PI & DMP Coordinator]
- Dr. Dat Manger (dat.manger@biodata.pt) [Data Manager]
- Dr. Col Hector (col.hector@biodata.pt) [Clinical Data & Sample Collector]
- Dr. R. Sercher (r.sercher@biodata.pt) [Researcher]
- Mrs. A. D'Min (a.dmin@biodata.pt) [Project Manager]

Host Institution: BioData.pt

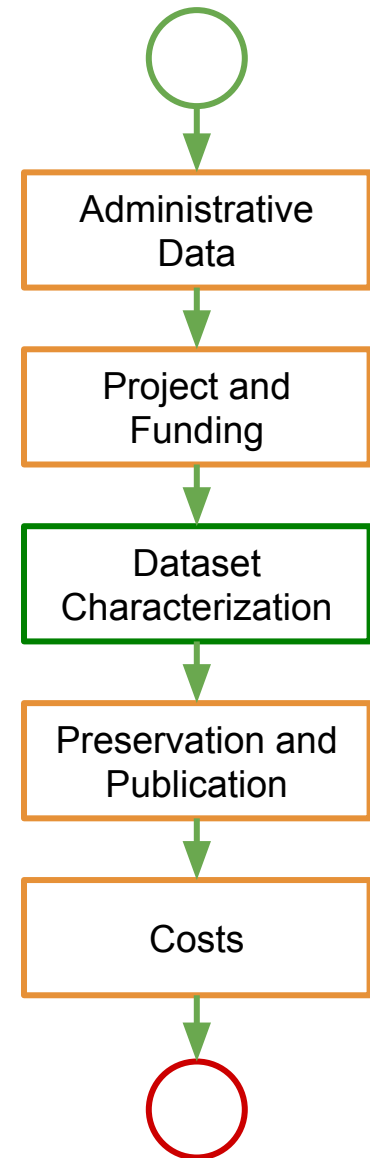**Start date:** January 1st, 2021

Duration: 36 months

- Information pertaining the **funding** of a **particular project**.

Creating a DMP (Step 3)

- **Step 3 - Dataset Characterization**

  - **Characterization** of the **dataset(s)** that are encompassed by the DMP. Apart from **generic information** on the dataset, **additional** descriptions of **security and privacy** policies, **technical resources** and **metadata** standards can also be given.

  - The information is classified in four sections:

    - **General information** about all **datasets**.

    - Any **security and privacy** policies associated with the datasets.

    - **Technical resources** associated with the datasets.

    - **Metadata** associated with the datasets.

```
        ○
        │
        ▼
  ┌─────────────┐
  │Administrative│
  │    Data     │
  └─────────────┘
        │
        ▼
  ┌─────────────┐
  │ Project and │
  │   Funding   │
  └─────────────┘
        │
        ▼
  ┌─────────────┐
  │   Dataset   │
  │Characterization│
  └─────────────┘
        │
        ▼
  ┌─────────────┐
  │Preservation and│
  │ Publication │
  └─────────────┘
        │
        ▼
  ┌─────────────┐
  │    Costs    │
  └─────────────┘
        │
        ▼
        ○
```

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 3)

In the project

> **Research plan and method summary:**
>
> The project will be divided into four activities:
> 1. Sample collection
> 2. Phage X sequencing
> 3. 16S sequencing
> 4. Metatranscriptomics

Creating a DMP (Step 3)

In the project

Research plan and method summary:

The project will be divided into four activities:

1. Sample collection
2. Phage X sequencing
3. 16S sequencing
4. Metatranscriptomics

● **General information** about all **datasets**.

Creating a DMP (Step 3)

In the project

Sample Collection:

In the sample collection activity, we will define a study group of volunteer disease X patients, numbering no less than 20, and a control group comprising their close relatives, 1-2 per patient. We will collect stool samples from each of the volunteers.

Creating a DMP (Step 3)

In the project

Sample Collection:

In the sample collection activity, we will define a study group of volunteer disease X patients, numbering no less than 20, and a control group comprising their close relatives, 1-2 per patient. We will collect stool samples from each of the volunteers.

- Any **security and privacy** policies associated with the datasets.

BioData.pt

Creating a DMP (Step 3)

In the project

> **Phage X sequencing:**
>
> In the Phage X sequencing activity, we will carry out DNA sequencing of the stool samples and assemble the genome of Phage X. In order to facilitate the assembly while enabling the reliable identification of sequence variants, we will combine the higher quality but short read sequencing technology of the Illumina NextSeq 500 sequencer with the long read but lower quality technology of the Nanopore MinION sequencer.

Creating a DMP (Step 3)
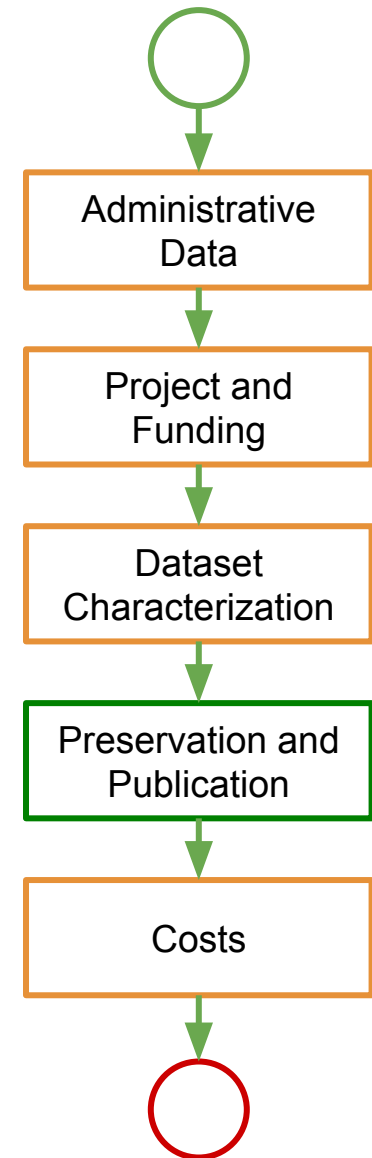
In the project

Phage X sequencing:

In the Phage X sequencing activity, we will carry out DNA sequencing of the stool samples and assemble the genome of Phage X. In order to facilitate the assembly while enabling the reliable identification of sequence variants, we will combine the higher quality but short read sequencing technology of the Illumina NextSeq 500 sequencer with the long read but lower quality technology of the Nanopore MinION sequencer.

- **Technical resources** associated with the datasets.

- **Metadata** associated with the datasets.

Ready for BioData Management?

Creating a DMP (Step 4)

- **Step 4 - Preservation and Publication**

  - **Characterization** of the **preservation** and **publication** policies **for each** of the identified **datasets.**

  - The information is classified in three sections:

    - Information regarding the **policies** on how each dataset is **distributed**.

    - Information on the data **host for each** of the identified **distributions**.

    - Characterization of the **licenses** associated with each **distribution policies**.

Administrative Data

Project and Funding

Dataset Characterization

Preservation and Publication

Costs

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 4)

In the project

BioData.pt|ELIXIR PT

Ready for BioData Management?

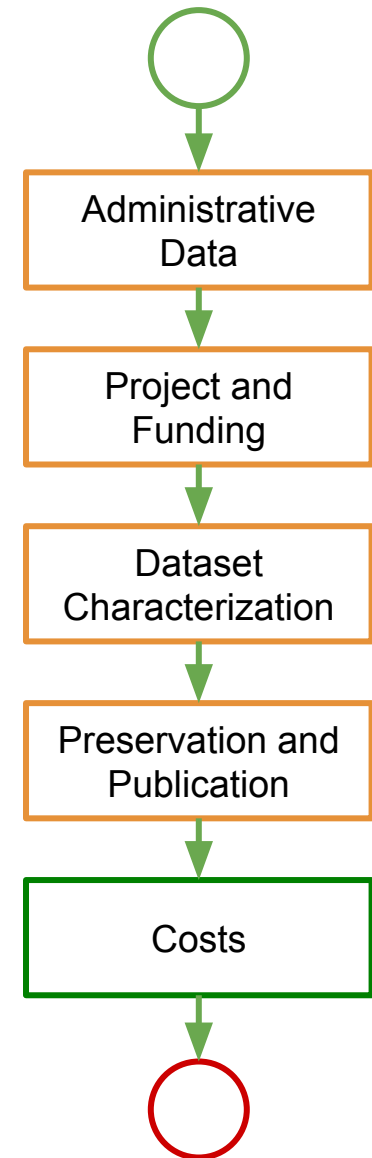# Creating a DMP (Step 4)

In the project

> **Why is this empty? Not usually covered in project descriptions. What would you write?**

- Information regarding the **policies** on how each dataset is **distributed**.

- Information on the data **host for each** of the identified **distributions**.

- Characterization of the **licenses** associated with each **distribution policies**.

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 5)

- **Step 5 - Costs**

  - **Characterization** of the **costs** associated with this DMP.

    - The numeric value associated with each cost can be a rough estimate.

Creating a DMP (Step 5)

In the project

Expected Data & Metadata Outputs:

1. Sample Collection:
   - ○ Patient clinical data (< 1 MB)
   - ○ Sample identification table (< 1 MB)
2. Phage X sequencing
   - ○ Raw FASTQ sequencing data - NextSeq (60 MB)
   - ○ Sample preparation & sequencing metadata - NextSeq (< 1 MB)
   - ○ Raw FASTQ sequencing data - MinION (1 GB)
   - ○ Sample preparation & sequencing metadata - MinION (< 1 MB)
   - ○ Assembled Phage X genome (< 1 MB)
   - ○ Assembly metadata (< 1 MB)
3. ...

BioData.pt

Ready for BioData Management?

Creating a DMP (Step 5)

In the project

Expected Data & Metadata Outputs:

1. Sample Collection:
   - Patient clinical data (< 1 MB)
   - Sample identification table (< 1 MB)
2. Phage X sequencing:
   - Raw FASTQ sequencing data - NextSeq (60 MB)
   - Sample preparation & sequencing metadata - NextSeq (< 1 MB)
   - Raw FASTQ sequencing data - MinION (1 GB)
   - Sample preparation & sequencing metadata - MinION (< 1 MB)
   - Assembled Phage X genome (< 1 MB)
   - Assembly metadata (< 1 MB)

What other costs would you consider?

- **Characterization** of the **costs** associated with this DMP.

BioData.pt

Ready for BioData Management?