# Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** EzMechanism: An Automated Tool to Propose Catalytic Mechanisms of Enzyme Reactions

**Creator:**Antonio Ribeiro

**Principal Investigator:** António J M Ribeiro

**Data Manager:** António J M Ribeiro

**Project Administrator:** Janet M Thornton

**Contributor:** Neera Borkakoti, António J M Ribeiro, Ioannis G. Riziotis, Jonathan D. Tyzack

**Affiliation:** Other

**Template:** DCC Template

**ORCID iD:** 0000-0002-2533-1231

**ID:** 124544

**Last modified:** 22-05-2023

# EzMechanism: An Automated Tool to Propose Catalytic Mechanisms of Enzyme Reactions

## Data Collection

### What data will you collect or create?

During the study we will create mainly two types of data: a) the rules of enzyme catalysis; b) prediction results for the use of EzMechanism against a test set of enzyme mechanisms.

"Rules of Enzyme Catalysis" comprise a SMARTS reaction pattern that codifies each rule, as well as information about where that rule was extracted (catalytic step, mechanism, and enzyme).

"Prediction Results" will be given several json files (one for each prediction) containing the reactants, products and intermediates of the predicted mechanisms as well as the rule used to generate each catalytic step. Additionally the steady states will be pictured as an svg file.

The volume of the created data will be less than 1GB.

### How will the data be collected or created?

"Rules of enzyme catalysis" will be automatically generated from the catalytic steps of enzymes available in the M-CSA database. Rules will have a version number, since different kind of rules might be generated for the same catalytic steps. The data about the rules, will be stored primarily in the M-CSA database and also made available as a flat file.

"Prediction Results" will be generated from the EzMechanism code that will be written as part of the project.

## Documentation and Metadata

### What documentation and metadata will accompany the data?

The generated data will be part of the M-CSA database and website, which has been described before and is well documented. A paper shall also be published describing in more detail the EzMechanism program.

## Ethics and Legal Compliance

### How will you manage any ethical issues?

We do not anticipate any ethical issues associated with this research project.

### How will you manage copyright and Intellectual Property Rights (IPR) issues?

These data generated for this project, as all data contained in M-CSA will be distributed according to a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

## Storage and Backup

### How will the data be stored and backed up during the research?

We expect the data to be part of the M-CSA database and website, which at the moment is managed by the EMBL-EBI.  We expect to have enough storage for the generated data and backups of data are also in place, as part of the EMBL-EBI infrastructure.

### How will you manage access and security?

Data will be made available through the website directly (user interface) and flat files. Data is in the public domain so no special restrictions are necessary.

## Selection and Preservation

### Which data are of long-term value and should be retained, shared, and/or preserved?

Both types of data mentioned above are important and should be preserved. "Rules of catalysis" might be used independently by other researchers in their programs. The "Prediction Results" will be important to show the validity and capabilities of the current version of EzMechanism.

### What is the long-term preservation plan for the dataset?

The data is part of the M-CSA database which has funding for at least another 6 years. We will be also sharing the data on another archiving website in case the M-CSA database is not funded further.

## Data Sharing

### How will you share the data?

At the time of publication, the data will be put in the public domain attached to a Creative Commons Attribution 4.0 International (CC BY 4.0) license. Data will be shared through the M-CSA website and an external archive to be chosen (with a persistent identifier).

### Are any restrictions on data sharing required?

No restrictions.

## Responsibilities and Resources

### Who will be responsible for data management?

António Ribeiro will be responsible for the data management of the project and all associated responsibilities.

### What resources will you require to deliver your plan?

All resources required to deliver this plan are already in place. These include mainly the M-CSA database and website.