Plan Overview

A Data Management Plan created using DMPonline

Title: Novel Bioinspired Optimization Techniques for Resilience and Sustainability

Creator: Carlos Cotta

Principal Investigator: Carlos Cotta, Antonio J. Fernández Leiva

Data Manager: José E. Gallardo, David Bueno Vallejo, Jesús Martínez Cruz

Affiliation: Other

Template: DCC Template

ORCID iD: 0000-0001-8478-7549

ORCID iD: 0000-0002-5330-5217

Project abstract:

This project revolves around the idea of resilience, both as a sought property in the design and operation of bioinspired optimization methods (BOMs), and as a property of systems of interest to be attained or boosted via the application of said BOMs. Herewith, we will use the term resilience to denote a systems ability to rebound, namely to withstand shocks in a dynamic way and bounce back to good operational conditions in an efficient way. Indeed, resilience is a major topic after dramatic disruptions such as the COVID-19 pandemics where many organizations/systems were too brittle to quickly adapt and they simply broke. BOMs can be instrumental in this context, and can be used to minimize risk exposure, to endow the system with resilience by design, and to help in the restoration process after the system has been hit by a significant disruption.

Resilience is not only a desired property of the system to which BOMs are applied, but also a sought property for these methods themselves. Consider for example they can be run on complex, large and emergent computational environments that are irregular and volatile or open to malicious attacks to the system integrity. This requires making algorithms cognizant of such features of the computational landscape, having them autonomously adapt to the irregular (dynamic and heterogeneous) environment. From a system design perspective, this naturally leads to the notion of deep BOMs exhibiting multiple interconnected layers/components that contribute the desired characteristics by encapsulating the tools required to tackle the different aspects of the complexity of the problem and the intricacy of the computational substrate. It must be also noted that resilience is intimately related to sustainability, and can be seen as a requisite for the latter. It is therefore crucial that the operation of AI tools in general and BOMs in particular is done in a sustainable way.

Therefore, this project will research on BOMs in connection to resilience. On one hand, this involves both a study of the factors and methods whereby these techniques can be endowed with resilience to different kind of disturbances, as well as the conditions under which these methods can be run in a sustainable, eco-friendly way. On the other hand, we will consider the application of BOMs to domains in which resilience is sought. This research will be conducted following Open-Science guidelines in order to maximize the impact and outreach (both scientific and societal) of the results, as well as to contribute to the sustainability of the

scientific process.

ID: 106384

Start date: 01-09-2022

End date: 31-08-2025

Last modified: 22-11-2022

Grant number / URL: PID2021-125184NB-I00 / http://bio4res.lcc.uma.es

Copyright information:

Novel Bioinspired Optimization Techniques for Resilience and Sustainability

Data Collection

What data will you collect or create?

Run traces of bioinspired algorithms providing forensic data about the execution of these algorithms, in particular performance indicators (such as values of the objective function(s) being optimized, or the solutions being found), population metrics (such as measures of diversity or any other property of the population), and algorithmic behavior descriptors (such as CPU time or energy consumption). Each individual trace will contain tabular data and may have a size ranging from a few kBs up to a few MBs. The data will be stored in an amenable format such as CSV for data sharing, reuse and preservation.

How will the data be collected or created?

Traces will be obtained by repeatedly running the bioinspired algorithms under scrutiny. These algorithms will be designed to output such an execution trace as they run. Given the stochastic nature of these techniques, multiple runs will be performed with each particular parameterization. Each batch of traces will be stored in a separate folder, named after the combination of parameters used in the experimentation, and further contained within other folders in case the parameter combinations are systematic (each level in the hierarchy of folders representing a particular parameter, and the different folder at that level being named after the value of the parameter). At the bottom of the hierarchy, the folders will contain a batch of files, each of them using names with the pattern "run#.csv', where # is the ordinal number of each corresponding run. A file "configuration.csv" will be also present at this level, containing the algorithm parameters for that batch of results. This structure can be further extended with levels describing the algorithm considered (if there are several) or the problem tackled (again, if there are several). In general, we define the concept of "experiment" as such a hierarchy, often representing a particular research work. Many experiments will be stored in parallel or within a hierarchical structure if it was deemed more readable.

Documentation and Metadata

What documentation and metadata will accompany the data?

CSV files will have self-descriptive column names. Each experiment (as defined in the "Data Collection" section) will be structured using the Data Package standard. A data-package descriptor will be included above the top level of the hierarchy of files in the experiment. This descriptor will be named "datapackage.json" and will be a valid JSON file (as defined in RFC 4627) that provides metadata about the data package, and describes its contents. The format of this JSON file will follow the standard defined in the Frictionles Data website (https://specs.frictionlessdata.io/data-package/). Data packages may be linked to analysis components (other data packages generated by processing the former, as well as the analysis scripts used) and communication components (reports and published articles using the data).

Ethics and Legal Compliance

How will you manage any ethical issues?

Whenever a human user can interact with any system developed in the project and such interaction is susceptible of being preserved, an informed consent will be provided indicating the extent of the data stored and its usage. Upon acceptance, the data preserved will be anonymized. No personal or sensitive information will be kept.

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

The intellectual property of all data generated and applications developed belongs to the University of Málaga (UMA) and the researchers involved. To promote scholarship and reproducibility, data will be publicly shared by default under a CC-BY-ND license (https://creativecommons.org/licenses/by-nd/4.0/). Aiming to maximize clarity and reusability, not all data will be made public though. The data will be curated following criteria of relevance and significance. Additionally, data may be protected when required to safeguard the interest of the UMA whenever an intellectual property instrument (including -but not limited to- patents, software registers, and utility models) is pursued. The exercise of such instruments will be done in accordance to the policies and indications of the Transference Unit of the UMA.

Storage and Backup

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

Raw data will be routinely kept in private cloud storage under the UMA institutional accounts. Publicly shared data will be stored in the Open Science Framework repository (https://osf.io). This includes Data Packages as well as analysis scripts, and accompanying reports. Published reports will be also stored in the institutional repository of the UMA (https://riuma.uma.es). Source code will be released via a version control system. The use of GitHub (https://github.com) and is envisioned for this purpose.

How will you manage access and security?

Internal access to the data by members of the research team will be done via Microsoft SharePoint, using the UMA institutional account. Given the nature of the data, which does not include sensitive information, there are no security concerns. The standard security mechanisms provided by the cloud storage platform will be in place.

Selection and Preservation

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

Regarding the execution of bioinspired algorithms, both input data (problem instances considered in the experiments) and output data (run traces describing the behavior of the algorithm, its internal state throughout the run, and/or the results attained) will be preserved for the purposes of transparency in research, and reproducibility of the results.

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

The OSF repository considered for storage has funding sufficient for ensuring read access to the data for +50 years, which is way more than the reasonably expected timeframe during which the project data can be reused.

Data Sharing

How will you share the data?

As a default rule, the data will be made available under a CC-BY-ND license (https://creativecommons.org/licenses/by-nd/4.0/) in the OSF site at the time at which scholarly reports are prepared, or prior to that in case the data were deemed particularly relevant and worth sharing on a timely basis. Generally, such an upload will be publicized in the research group and project websites, as well as in the associated social media, in order to maximize outreach. Using the OSF repository will provide permanent identifiers (DOIs) to the data, both for citation and sharing purposes.

Are any restrictions on data sharing required?

In some cases data may require being anonymized and aggregated before sharing, leading to a longer timeline before being made publicly available. Also, there may be circumstances related to intellectual property rights (application to some IPR instrument, or embargo required by publishers, to cite some examples) which may advise keeping some data/applications protected for some period of time. Whenever data or any other artifact produced by the project is made public and reused by the community, it is expected that the user will provide credit to the researchers involved by citing the data (via DOI) and/or the accompanying published report (if it exists).

Responsibilities and Resources

Who will be responsible for data management?

Data manager duties will be decentralized among the members of the research team in order to avoid bottlenecks in the process. Data capture, metadata production, data quality, storage, backup and archiving will be the responsibility of the researcher carrying out the particular investigation (or the designated researcher in case there are several involved). The principal investigators of the project (Prof. C. Cotta and Prof. A.J. Fernández) will oversee the correct implementation of the Data Management Plan.

What resources will you require to deliver your plan?

All physical resources required to implement the Data Management Plan are available via institutional plattforms (RIUMA, OneDrive cloud storage) and free tools (OSF, GitHub). Members of the research team will require some basic training on the use of these plattforms and tools.