

## BITS :: Call for Abstracts 2024 - Oral communication

|                    |  |
|--------------------|--|
| <i>Type</i>        | Oral communication   |
| <i>Session</i>     | Data management, sharing and FAIRness  |
| <i>Title</i>       | CREDO: Customizable, Reproducible, Dockerfile Generator for Bioinformatics Applications                      |
| <i>All Authors</i> | Eliseo Martelli(1, 2), Maria Luisa Ratto(2), Beatrice Nuvolari(2), Maddalena Arigoni(2), Luca Alessandri(2). |

### *Affiliation*

(1) Department of Computer Science, University of Turin, Turin, Italy

(2) Department of Molecular Biotechnologies and Health Sciences, University of Turin, Italy

### *Motivation*

Bioinformatics research faces significant challenges due to a diverse and segmented toolkit and ecosystem, making it difficult to achieve fully reproducible environments. The absence of standardization in research intensifies the difficulty of implementing a solution that functions seamlessly across various scenarios. While containerization offers some answers, it falls short in providing an effective dependency management framework and relies heavily on a central hub like Docker. To address these issues, we propose a new tool, CREDO, to tackle these challenges. Designed for flexibility, adaptability, and scalability, CREDO supports diverse platforms, data formats, standards, and workflows, aligning with the FAIR principles. It generates a self-contained executable that operates the required environment under the most important operating systems and architectures, promising to overcome the current limitations.

### *Methods*

The application is programmed in Go, which facilitates lightweight parallelization of long-running tasks. Due to the challenges of tracking dependencies, especially R package dependencies, we are populating a key-value store to maintain this information. Additionally, we have developed a fuzzy search service to efficiently locate dependencies. The architecture of CREDO is modular and taps into a registry, allowing for the development of self-contained modules. Modules generate and consume Spells. A spell is the unit of configuration for CREDO and contains all the necessary steps to produce a properly configured dependency in the target system.

Each Module acts as an interface for different environments -- like CRAN or Bioconductor for R, pip or conda for Python, apt or brew for system libraries, and git. These Modules can be interdependent, enabling CREDO to discern if a CRAN library requires a package from apt, for example. This capability enhances the utility and extendibility of our framework, allowing these modules and the libraries created to interact with various subsystems to be reused in other projects.

To commit a Spell into the full configuration, CREDO tests the feasibility of introducing that dependency by trying the installation in a temporary environment, and only if said installation succeeds, the spell gets appended to the configuration.

To ensure multi-platform compatibility, CREDO can include a container runtime in the generated .CREDO file, eliminating the need for external dependencies such as Docker. This built-in runtime simulates the environment, making the .CREDO file self-sufficient and automatically adjusting to the user's architecture if it is among those predefined, thus simplifying deployment across different systems.

### *Results*

CREDO has the potential to revolutionize bioinformatic analyses by providing a robust framework for enhancing experiment reproducibility. It simplifies the process of defining and managing dependencies, allowing researchers to focus more on the analysis rather than tedious setup tasks. Self-documentation ensures a transparent record of all steps involved, facilitating easy reproduction and extension of the results by others. CREDO also integrates seamlessly with existing workflows and containerization platforms, enabling smooth deployment and execution of analyses.

The final product of CREDO is a self-contained .CREDO executable package that equips researchers with all necessary tools to effectively handle data across multiple platforms. Launching this executable initiates a shell that sets up an environment precisely configured with the required libraries. It is crucial that the architecture on which the executable was built matches or is compatible with the architecture where it is launched. If there is a discrepancy in architecture between different machines, CREDO automatically suggests running the .CREDO file through tools that can emulate the recommended architecture, ensuring that the environment is accurately reconstructed on any system.

CREDO significantly simplifies life for peer reviewers and researchers attempting to replicate analyses from published papers:

For Peer Reviewers: CREDO's .CREDO executable instantly recreates the exact computational environment used in the study, allowing reviewers to focus on the scientific validity rather than setup issues.

For Replicating Analyses: The executable ensures that anyone can reproduce the analysis in precisely the same software environment, eliminating inconsistencies due to different software versions or configurations. If hardware architectures differ, CREDO suggests running the environment through an emulation layer to match the setup accurately.

### *Info*

|   |
|---|
| -   |
| <i>filename</i> -   |
| <i>Figure</i>   |
| -   |
| <i>Availability</i> <a href="https://pubmed.ncbi.nlm.nih.gov/38475691/">https://pubmed.ncbi.nlm.nih.gov/38475691/</a>   |
| <b>Dissemination Material</b>   |
| <i>Social</i>   |
| <a href="https://linkedin.com/in/eliseomartelli/">https://linkedin.com/in/eliseomartelli/</a><br><a href="https://calogerolab.it/">https://calogerolab.it/</a><br><a href="https://eliseomartelli.it/">https://eliseomartelli.it/</a>         |
| <i>Summary</i>  |
| CREDO (Customizable, Reproducible, Dockerfile Generator) aims to revolutionize bioinformatics research by providing a robust framework for experiment reproducibility, simplifying dependency management and self-documenting analysis steps. |
| <b>Corresponding Author</b>   |
| <i>Name, Surname</i> Eliseo, Martelli   |
| <i>Email</i> <a href="mailto:me@eliseomartelli.it">me@eliseomartelli.it</a>   |
| <i>Submitted on</i> 30.04.2024  |

**Società Italiana di Bioinformatica**  
C.F. / P.IVA 97319460586  
E-mail [bits@bioinformatics.it](mailto:bits@bioinformatics.it)

Sede legale Viale G. Mazzini, 114/B - 00195 Roma  
Website [bioinformatics.it](http://bioinformatics.it)

message generated by [sciencedev.com](http://sciencedev.com) for <https://bioinformatics.it> 09:07:44 30.04.2024