

Borse di studio GARR
Orio Carlini

Reproducible Bioinformatics?

CREDO: Customizable, REproducible, DOcker file
generator for bioinformatics applications

Eliseo Martelli

Università degli Studi di Torino

Tutor: Luca Alessandrì

**Bioinformatics is a discipline
that intersects various
fields of research.**

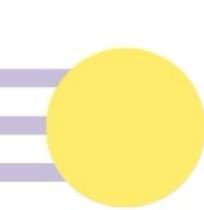


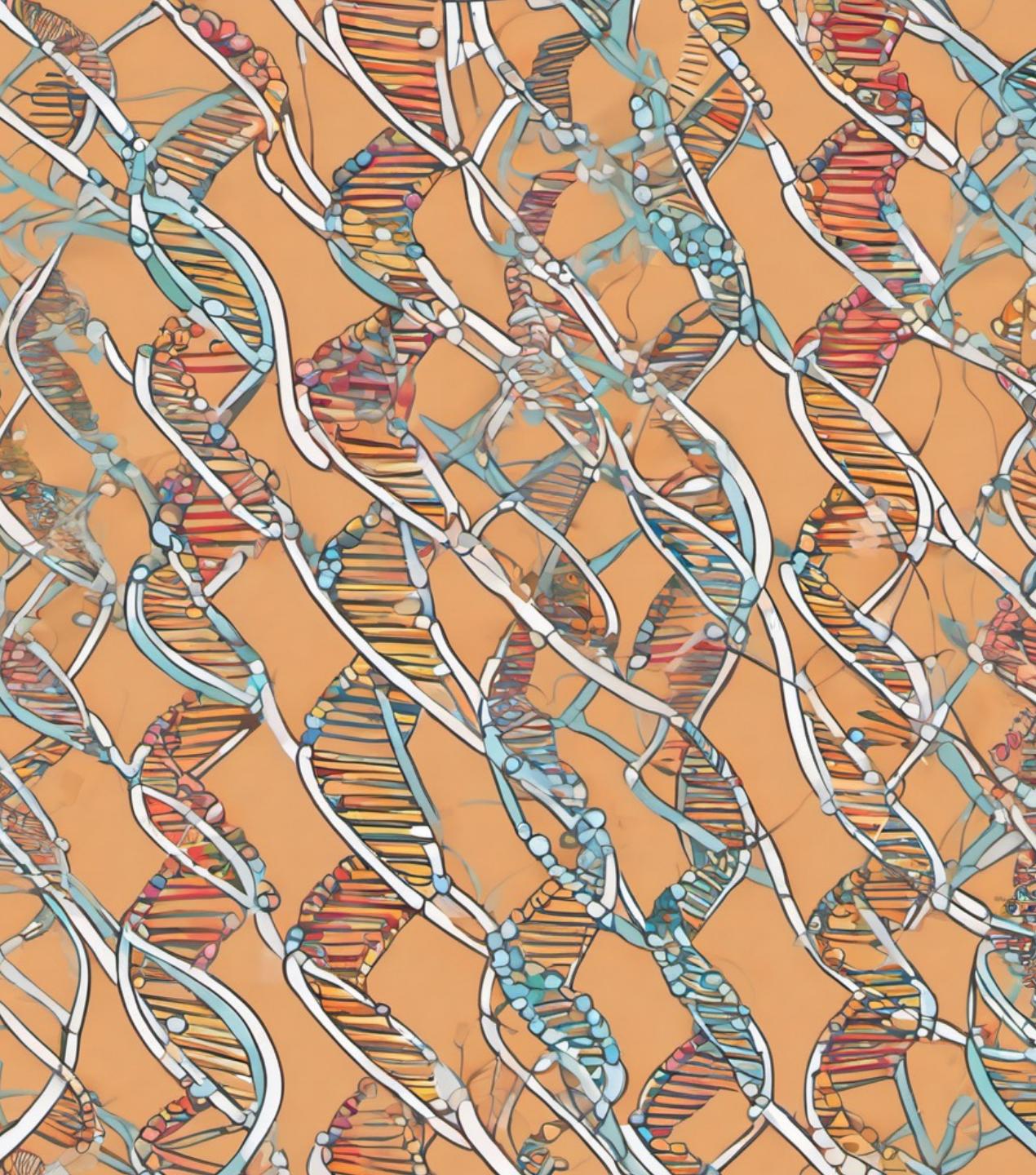
Image generated with stable-diffusion-2
Prompt:
DNA strand illustration minimal, monoline, drawing, warm palette

**Bioinformatics is a discipline
that intersects various
fields of research.**

Biology

Computer Science

Statistics



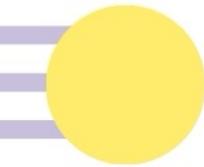


Image generated with stable-diffusion-2
Prompt:
Computer scientist, monoline, b&w, desk, computers

**Bioinformatics is a discipline
that intersects various
fields of research.**

Biology
Computer Science
Statistics



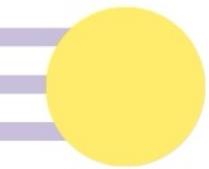


Image generated with stable-diffusion-2

Prompt:

statistic dashboard, drawing, minimalistic, warm palette

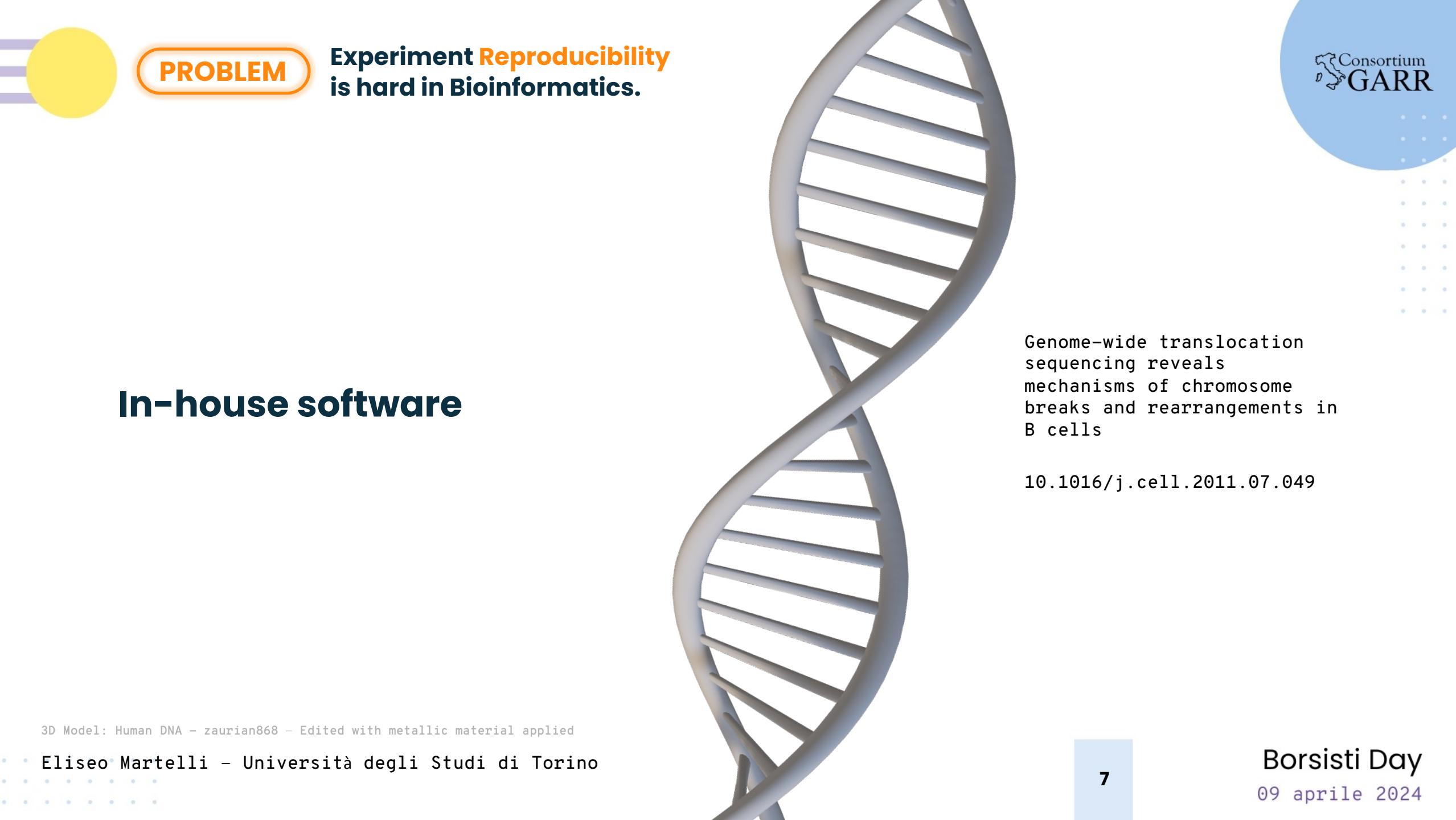
**Bioinformatics is a discipline
that intersects various
fields of research.**

Biology
Computer Science
Statistics



PROBLEM

Experiment Reproducibility is hard in Bioinformatics.



PROBLEM

Experiment Reproducibility
is hard in Bioinformatics.

In-house software

Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells

[10.1016/j.cell.2011.07.049](https://doi.org/10.1016/j.cell.2011.07.049)

3D Model: Human DNA – zaurian868 – Edited with metallic material applied

- Eliseo Martelli – Università degli Studi di Torino
- • • • • • •



PROBLEM

Experiment Reproducibility
is hard in Bioinformatics.

In-house software

DNA pieces scatter

[10.1016/j.cell.2011.07.049](https://doi.org/10.1016/j.cell.2011.07.049)



3D Model: Human DNA - zaurian868 - Edited with metallic material applied

• Eliseo Martelli - Università degli Studi di Torino



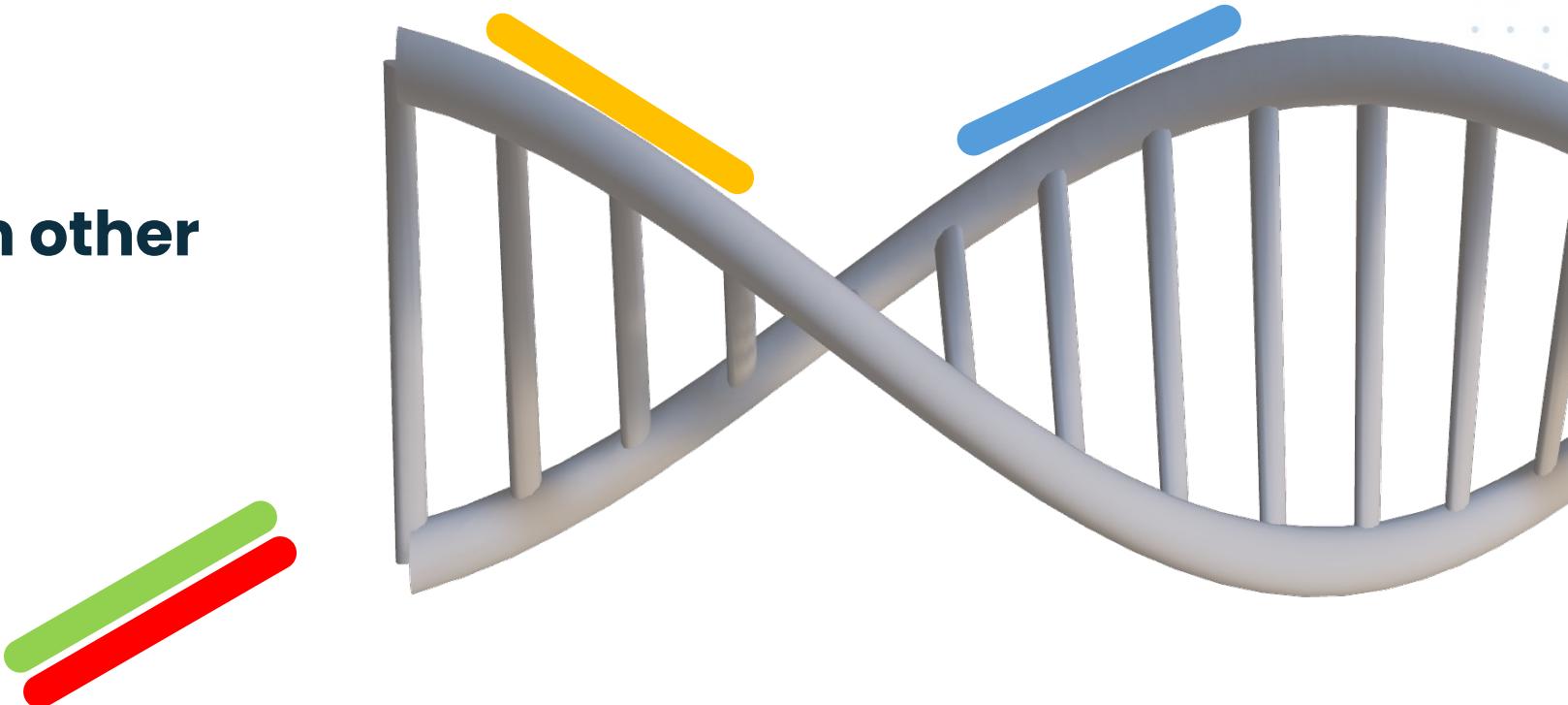
PROBLEM

Experiment Reproducibility
is hard in Bioinformatics.

In-house software

DNA pieces scatter

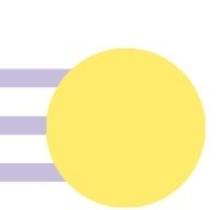
**They join together or in other
places of the DNA**



10.1101/j.cell.2011.07.049

3D Model: Human DNA - zaurian868 - Edited with metallic material applied

- Eliseo Martelli - Università degli Studi di Torino
- • • • • • •



PROBLEM

Experiment Reproducibility
is hard in Bioinformatics.



BLAST is the **core dependency of
this project.**

Basic local alignment search tool - 10.1016/S0022-2836(05)80360-2

3D Model: Human DNA - zaurian868 - Edited with metallic material applied

Eliseo Martelli - Università degli Studi di Torino

PROBLEM

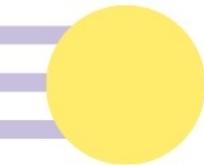
Experiment Reproducibility
is hard in Bioinformatics.

But there is a catch...

3D Model: Apple MacBook Pro 15 – texture removed – applied metallic material

Eliseo Martelli - Università degli Studi di Torino





PROBLEM

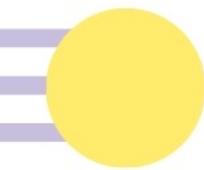
Experiment Reproducibility
is hard in Bioinformatics.



Let's compare 2 software versions

blast 2.12.0, build Mar 8 2022 16:19:08

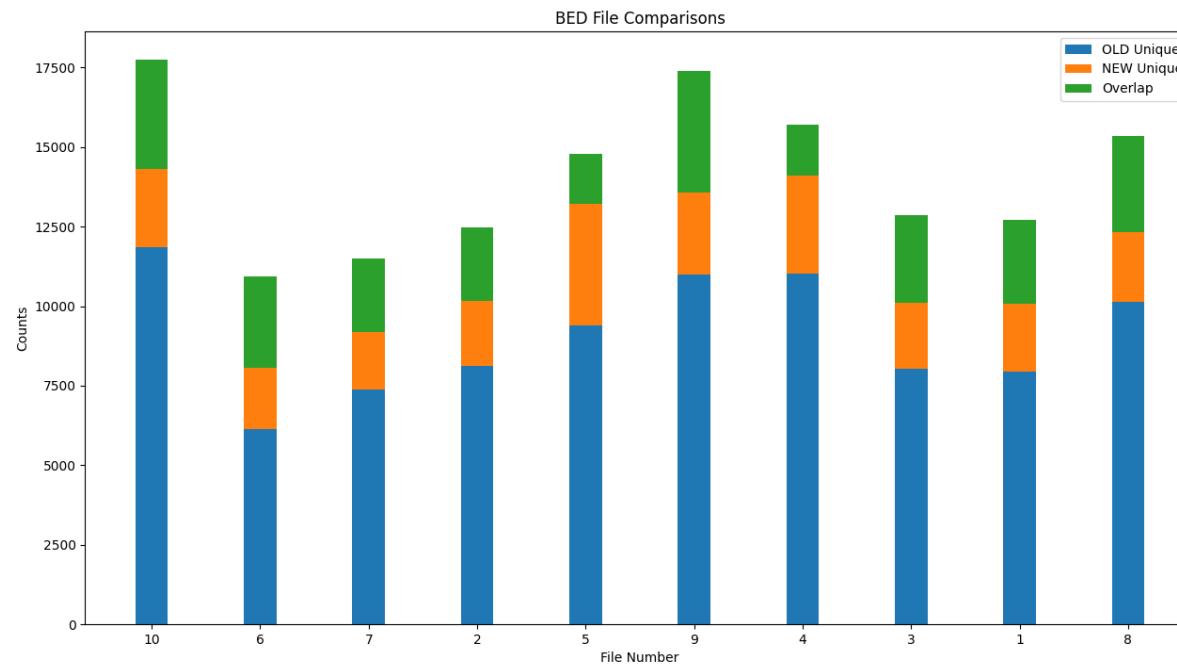
blast 2.2.31, build Jun 2 2015 10:47:46



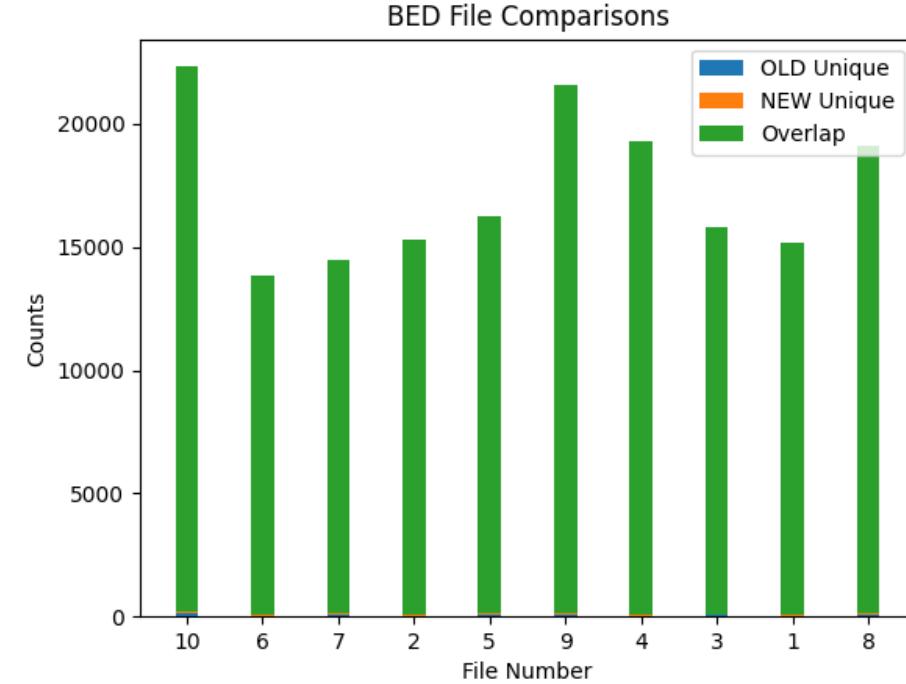
PROBLEM

Experiment Reproducibility
is hard in Bioinformatics.

Let's compare 2 software versions



blast 2.12.0, build Mar 8 2022 16:19:08



blast 2.2.31, build Jun 2 2015 10:47:46



PROBLEM

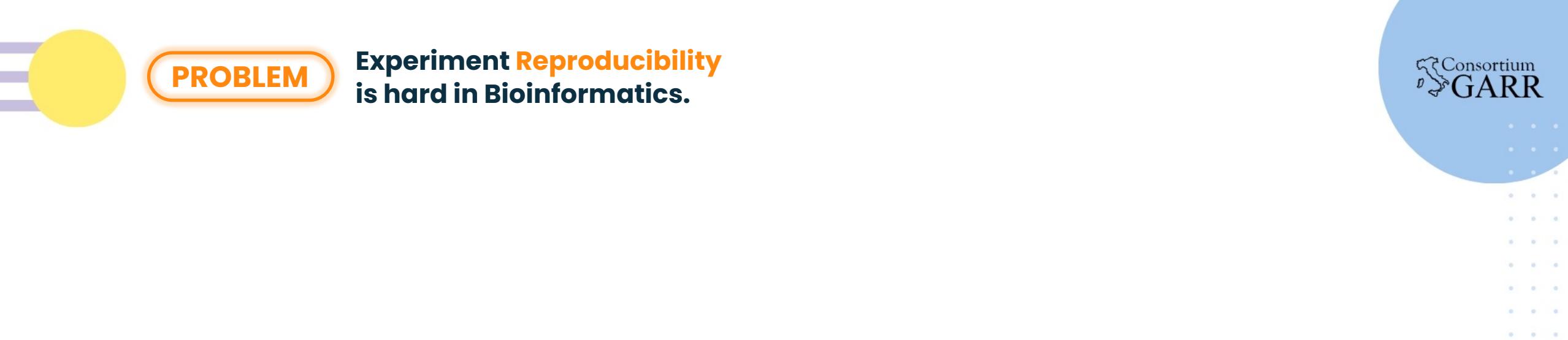
Experiment Reproducibility
is hard in Bioinformatics.

Consortium
GARR

Validity & Reliability

Insufficient Documentation

Lack of Standardization



PROBLEM

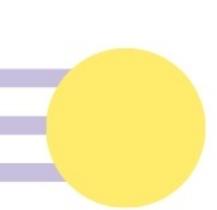
Experiment Reproducibility
is hard in Bioinformatics.

Consortium
GARR

Validity & Reliability

Insufficient Documentation

Lack of Standardization



PROBLEM

**Experiment Reproducibility
is hard in Bioinformatics.**

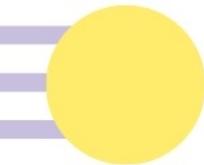


Validity & Reliability

Insufficient Documentation

Lack of Standardization





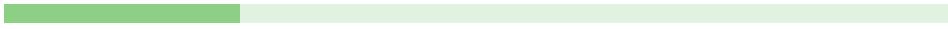
**CREDO stands
by the**



FAIR principles.

FINDABILITY

Consolidation of sources for researchers
to access, install, and utilize the software using centralized platforms.



ACCESSIBILITY



Simplification of the process of creating reproducible environments to users without previous bioinformatics knowledge.



INTEROPERABILITY

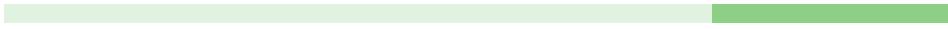
Ensuring compatibility with the most widely available operating systems to maximize the range of users that can use the application.



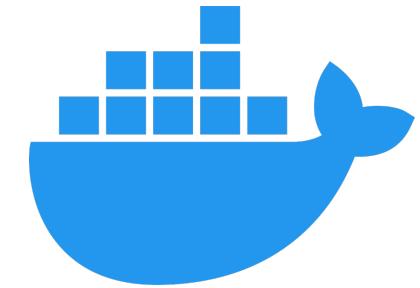
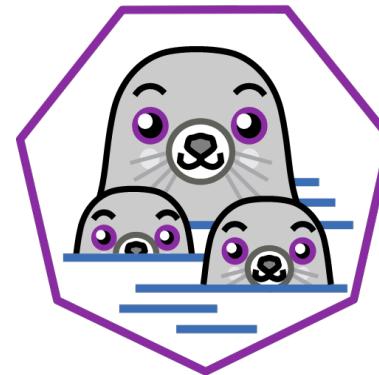
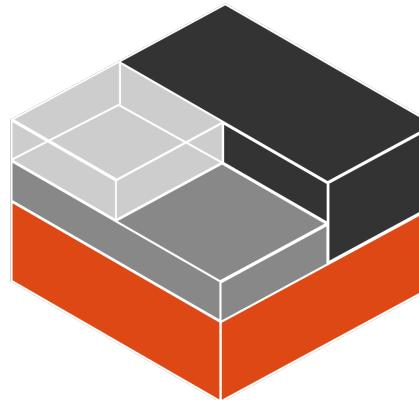
REUSABILITY

Facilitation of software customization.

Ensure archival of dependencies to prevent future inaccessibility of dependencies.



A possible solution? Containerization.



Docker Logo is a registered trademark of Docker, Inc.
Podman logo and LXC logo are public domain.

Eliseo Martelli - Università degli Studi di Torino



But... Containerization doesn't solve one big problem:
Dependency Management

Docker Logo is a registered trademark of Docker, Inc.
Podman logo and LXC logo are public domain.

Eliseo Martelli - Università degli Studi di Torino

Dependency Management

Involves efficiently handling and controlling the software components and libraries needed for an application to function properly, ensuring compatibility and mitigating conflicts.

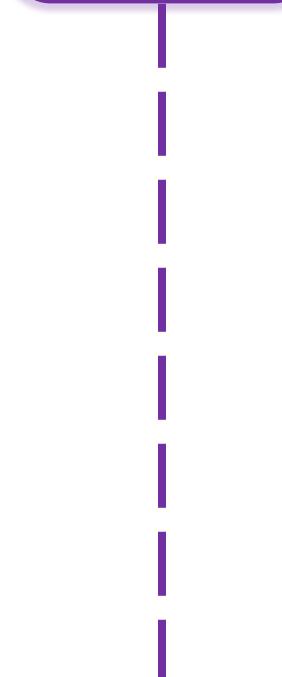
BEFORE

CREDO existed as a Proof of Concept.



CREDO: a friendly Customizable, REproducible, DOcker file generator for bioinformatics applications – 10.1186/s12859-024-05695-9

PLAN



Modularity & Optimization

Achieved through code-refactoring and implementing strong architectural patterns.

UI Redesign

Improve UI design and performance to increase user interaction quality and speed.

Multiuser Environment

Restructure the application to support multiple users and permissions.

DOI Tagging

Assign DOI references to artifacts generated to facilitate citations and references.

Quality Assurance & Testing

Configure safeguards and robust testing methodologies via CI and GitOps.

Documentation

Produce In-depth documentation
by following a doc-first approach.

We ❤️ Open Source

OSS Contributions

Using industry standard libraries
Building tools for the community

OSS Contributions

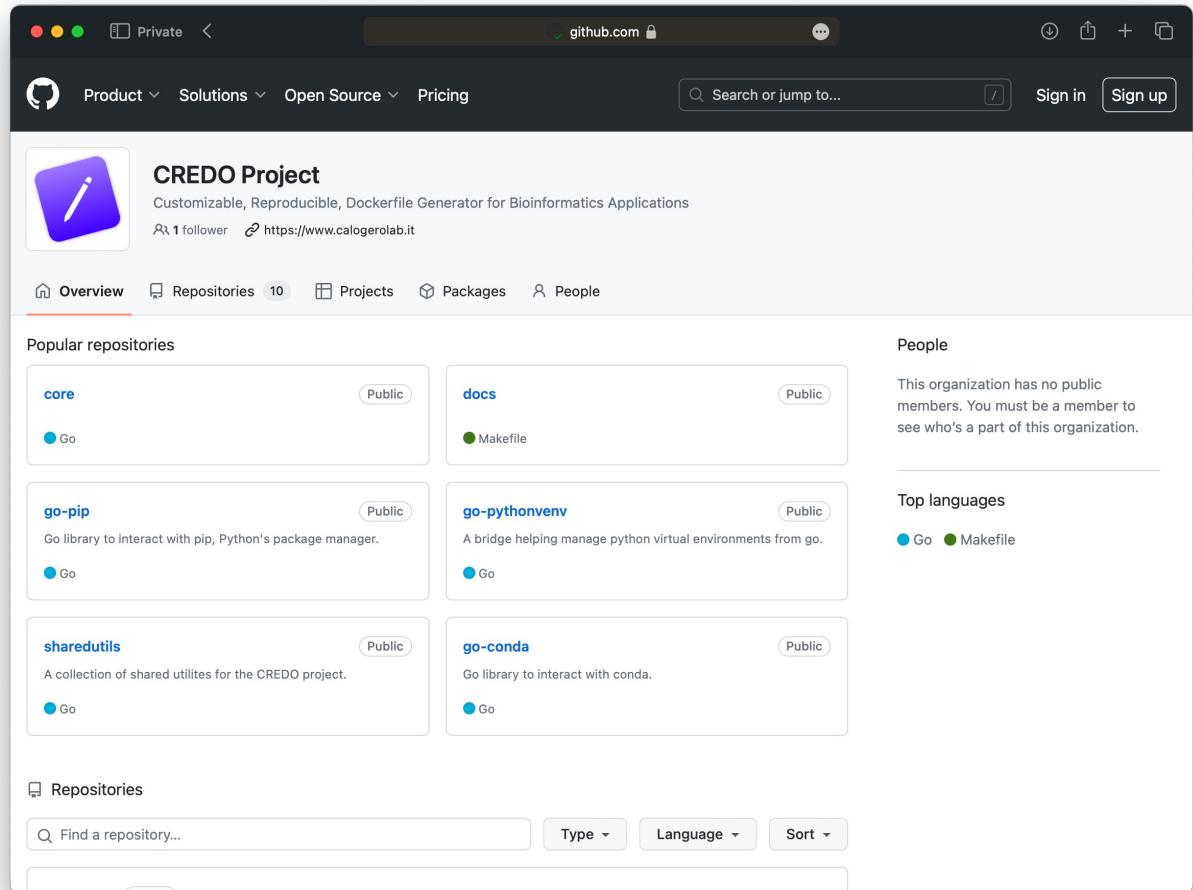
Using industry standard libraries

Building tools for the community

OSS Contributions
Using industry standard libraries
Building tools for the community



[Github.com/CREDOProject](https://github.com/CREDOProject)



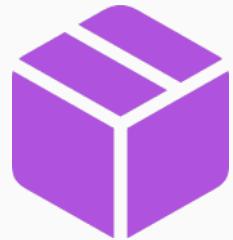
The screenshot shows the GitHub organization page for "CREDO Project". The header includes the GitHub logo, navigation links for Product, Solutions, Open Source, and Pricing, a search bar, and sign-in options. The main content area displays the organization's profile information, including its name, a brief description ("Customizable, Reproducible, Dockerfile Generator for Bioinformatics Applications"), and a link to their website. Below this, a section titled "Popular repositories" lists six public repositories: "core", "docs", "go-pip", "go-pythonvenv", "sharedutils", and "go-conda". Each repository card shows its name, language (Go or Makefile), and a brief description. To the right, there are sections for "People" (indicating no public members) and "Top languages" (Go and Makefile). At the bottom, there is a search bar and filters for "Repositories".

NEW

A new **evolution** of CREDO. **Built from the ground up.**

NEW

A new evolution of CREDO.
Built from the ground up.



Modular Architecture



Goroutines
Channels
WaitGroups



Registry
Pattern



Golang



Idempotence

Symbols provided by Hero Icons, MIT License

Eliseo Martelli - Università degli Studi di Torino

NEW

A new evolution of CREDO.
Built from the ground up.



Modular Architecture



Goroutines
Channels
WaitGroups

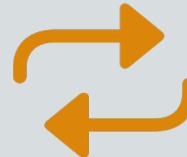


Registry
Pattern

YAML

=GO

Golang



Idempotence

Symbols provided by Hero Icons, MIT License

Eliseo Martelli - Università degli Studi di Torino

NEW

A new evolution of CREDO.
Built from the ground up.



Modular Architecture



Goroutines
Channels
WaitGroups

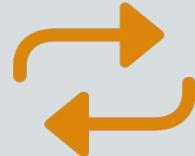


Registry
Pattern

YAML

=GO

Golang



Idempotence

Symbols provided by Hero Icons, MIT License

Eliseo Martelli - Università degli Studi di Torino

NEW

A new evolution of CREDO.
Built from the ground up.



Modular Architecture



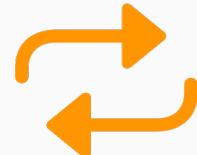
Goroutines
Channels
WaitGroups



Registry
Pattern



Golang



Idempotence

Symbols provided by Hero Icons, MIT License

Eliseo Martelli - Università degli Studi di Torino

NEW

A new evolution of CREDO.
Built from the ground up.



Modular Architecture



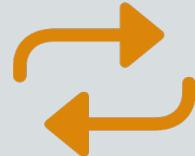
**Goroutines
Channels
WaitGroups**



**Registry
Pattern**



Golang



Idempotence

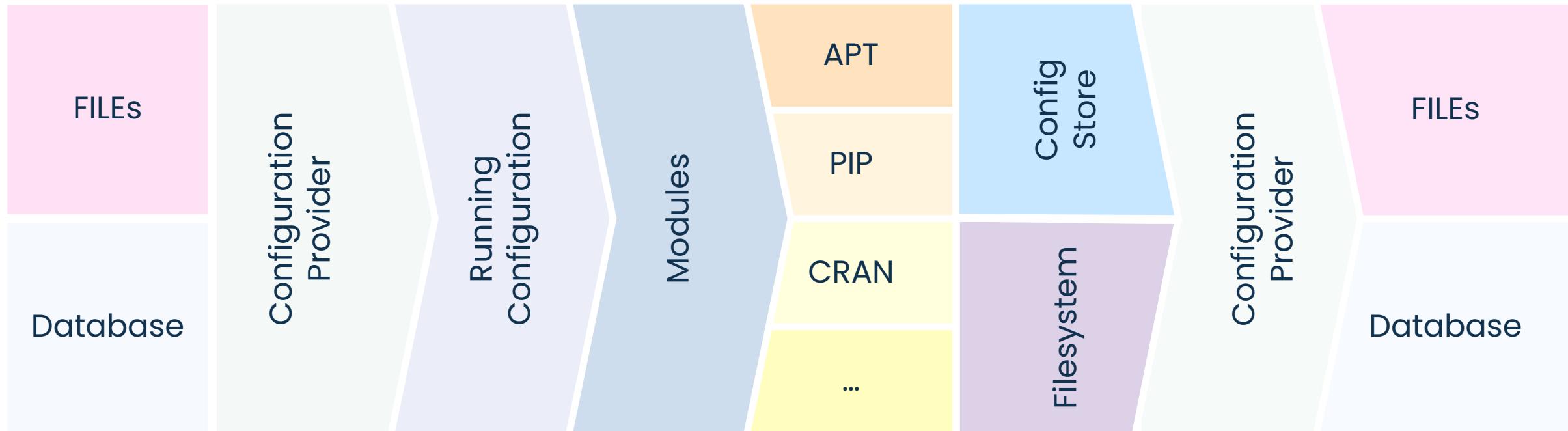
Symbols provided by Hero Icons, MIT License

Eliseo Martelli - Università degli Studi di Torino

ARCHITECTURE

CREDO's new building blocks.

ARCHITECTURE



ARCHITECTURE

Where are we now?

ARCHITECTURE

Where are we now?

8 +

**Open Source
Libraries Released**

ARCHITECTURE

Where are we now?

3 +

Integrated
Package Managers

What's next?

Evolution to a GARR-hosted service.

Exploration of techniques of session replay.

Integration of more package managers.

Feedback system for users.

Feedback system for users.

Borse di studio GARR
Orio Carlini

Thank you!