

# Statika: managing cloud resources, bioinformatics tools and data

Alexey Alekhin, Evdokim Kovach, Pablo Pareja-Tobes, Marina Manrique, Eduardo Pareja, Raquel Tobes, Eduardo Pareja-Tobes

Oh no sequences! Research Group. Era7 bioinformatics

## Introduction

Statika is a set of Scala libraries which

- allows building *well-structured* module systems
- where dependencies are *correct by construction*
- and you know it *before you run* anything

## Basic notions

### Bundle

A **bundle** is a thin wrapper for a tool, library, resource or any other component of your system:

- It may have dependencies on other bundles.
- It may do something in runtime, e.g. install a tool, that it represents.

### Distribution

A **distribution** is a bundle, which can deploy other bundles (it's members):

- It represents some environment, where you're going to use your bundles.
- Being a member of a distribution means to work fine with this environment.
- Distribution takes care of installing member dependencies first, and then the member itself.

## Availability



Free as in Freedom

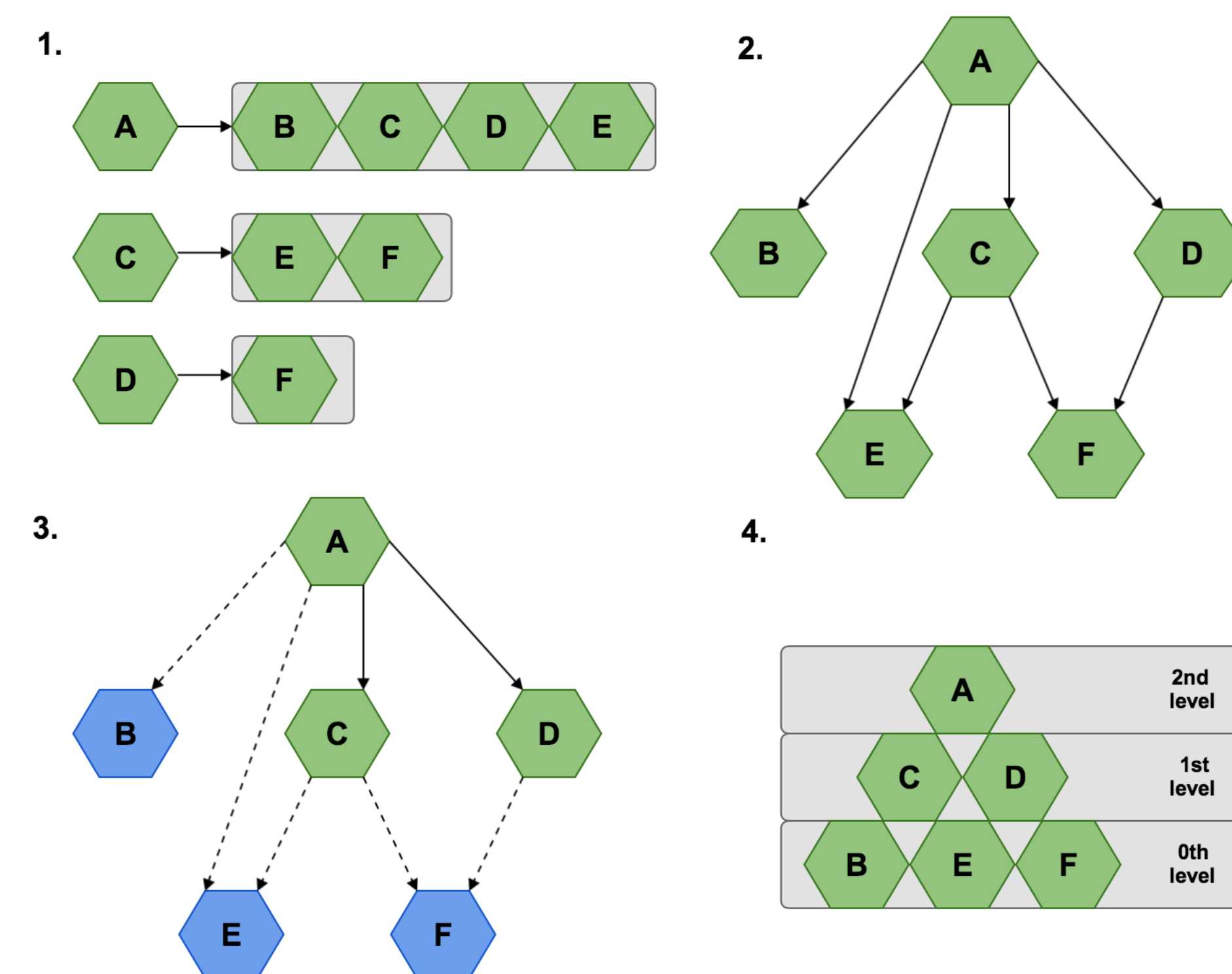
Statika is free and open-source under the AGPLv3 license

See <http://ohnosequences.com/statika>



## Abstract module system

- Bundles are represented as Scala types.
- Their dependencies on each other are validated by compiler — i.e. **statically**.
- Statika linearizes the types graph to get them in the right order.



## Artifacts management

**sbt-statika** — an sbt (simple build tool) plugin, which takes care of

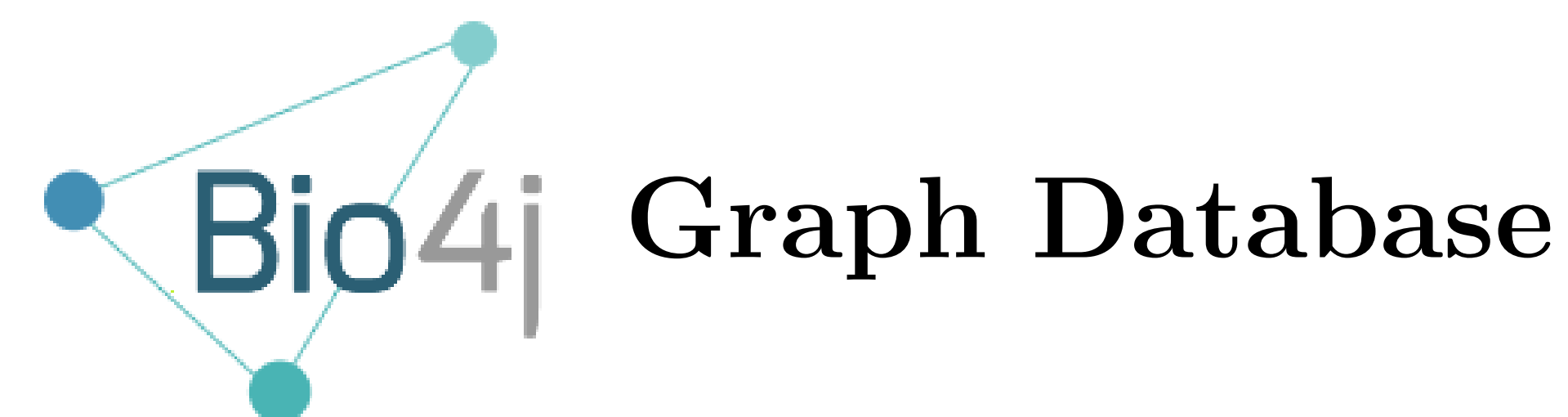
- packing bundles into versioned artifacts (jars)
- reusing sbt infrastructure to track dependencies on the artifact level
- standardizing common settings, versioning and release process

## Deployment

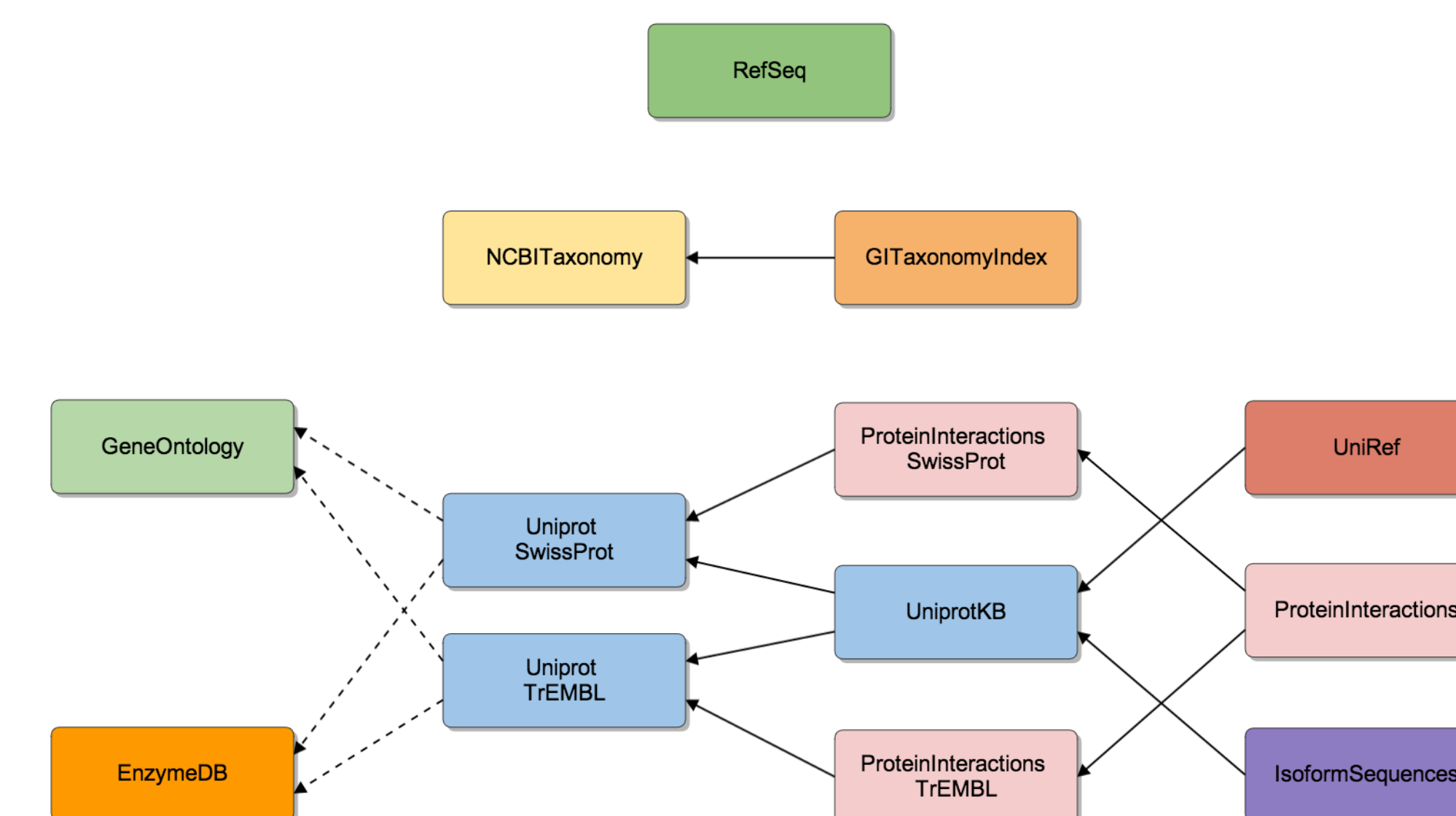
Amazon Web Services + **aws-statika** library

- Bundles can be *applied*, i.e. deployed to an EC2 instance
- Statika *distributions* abstract over the cloud infrastructure specifics

## Applications in bioinformatics

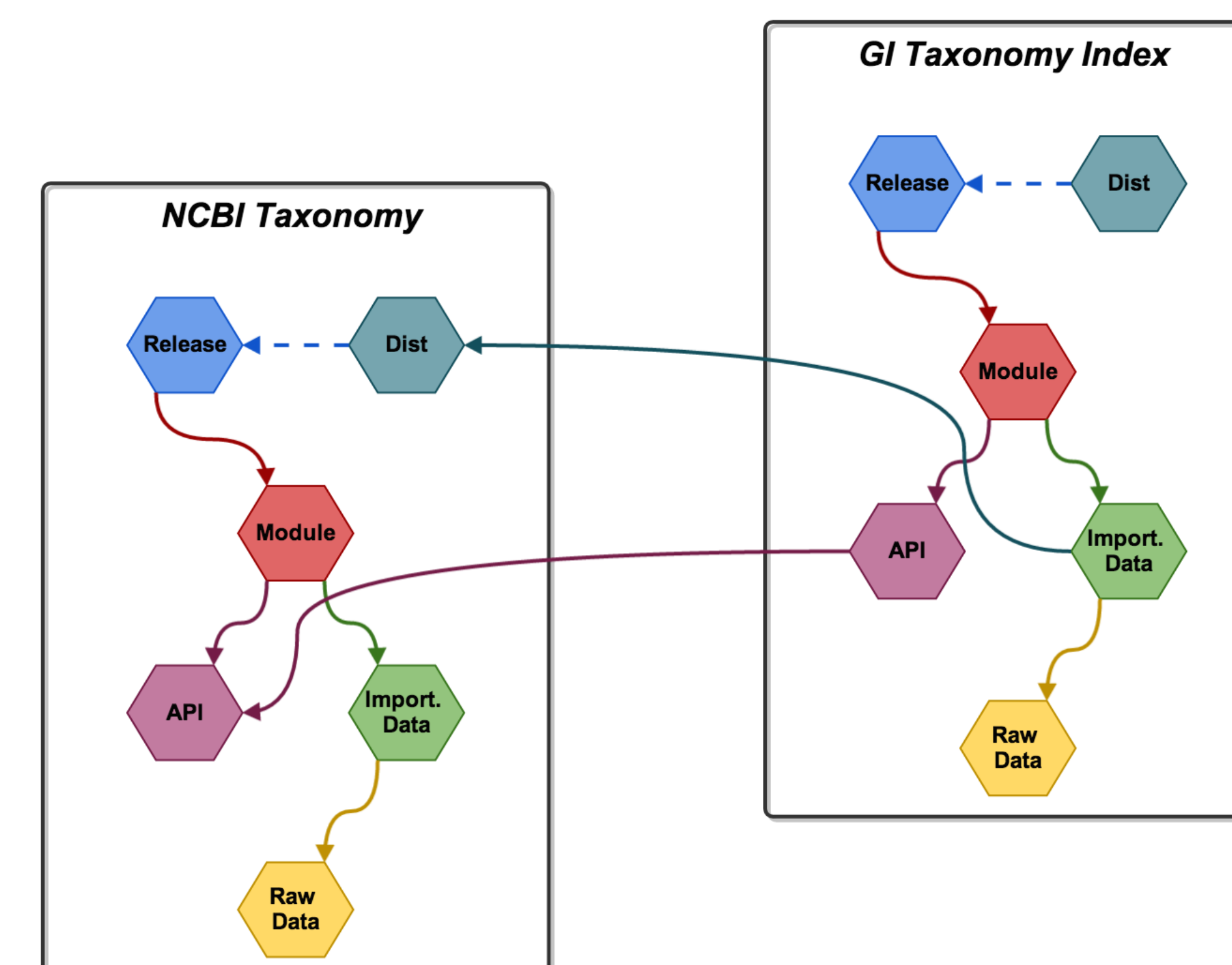


Bio4j is a bioinformatics graph database which integrates data from a lot of different sources:



Every module has some inner structure:

- raw data from a data source
- data importing process to the graph database
- nodes and relationships type definitions
- some abstract interface representing what you can do with this data



So Statika introduces to Bio4j a flexible module system with

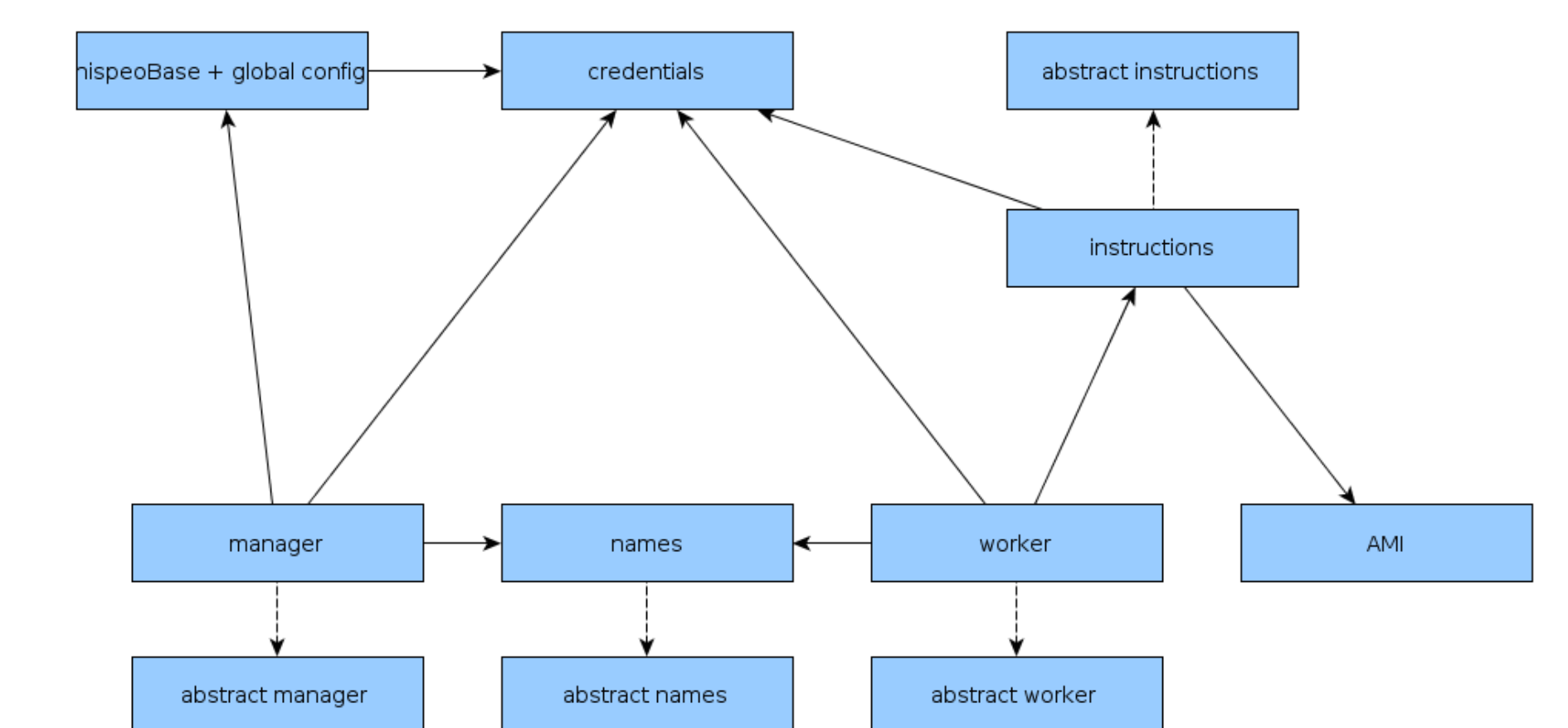
- simple data-import process
- automized dependencies management
- easy and robust deployment to AWS

## Applications in bioinformatics



### Cloud-computing System

Nispero is a toolset for declaring scalable cloud-based systems for bioinformatics computations. It has pretty complex inner components structure, which is managed with the help of Statika:



There is also a Statika distribution for bioinformatics tools, such as Velvet, Cufflinks, Tophat and Bowtie(2). See [github.com/statika/bioinfo-dist/](https://github.com/statika/bioinfo-dist/).

## Acknowledgments

Statika is developed by the R&D team of the Era7 Bioinformatics company



This project is funded in part by the ITN FP7 project INTERCROSSING (Grant 289974)

