flow
flow.bio

An open database and bioinformatics analysis platform

*Sam Ireland[4], *Charlotte Capitanchik[1,2,3], *J. Alex Harston[4], Marc Jones[1,2], Ira A. Iosub[1,2], Klara Kuret[5], Rupert Faraway[1,2], Oscar G. Wilkins[1,2], Charlotte West[2], Chris Cheshire[2], Anob M. Chakrabarti[1,2], Jernej Ule[1,2,3,5]

[1] UK DRI at King's College London, [2] The Francis Crick Institute, [3] Okinawa Institute of Science and Technology, Japan, [4] Goodwright Ltd, [5] National Institute of Chemistry, Ljubljana, * co-first authors

Scan me

goodwright — software for scientists | KING'S College LONDON | UK Dementia Research Institute | THE FRANCIS CRICK INSTITUTE

Flow in 6 steps:

Upload experimental data → Store data in standardised format → Run end-to-end pipelines → Explore results with interactive visualisations → Add to our public database in one click → Integrate publicly available data
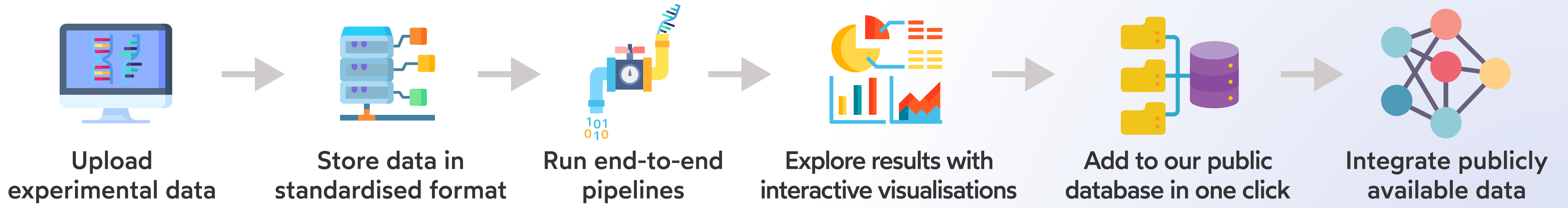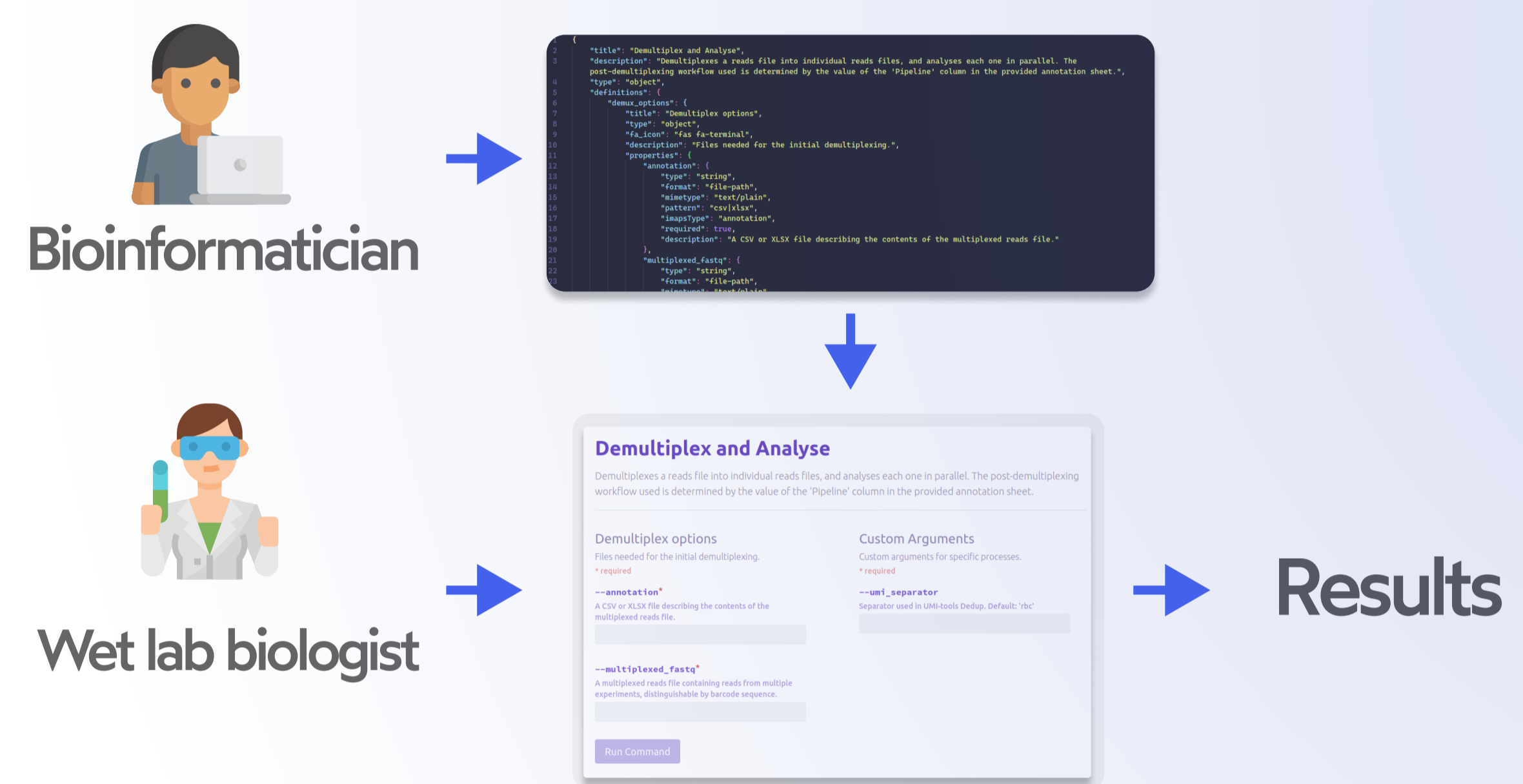
Modern bioinformatics still has issues

- **Nextflow** and **nf-core** have revolutionsed reproducibility and reduced complexity in bioinformatics, but access and data organisation can still be improved.

- **Academic databases often decay rapidly** because one lab or consortium is responsible for maintenance.

- **Data is often separated from analyses** that created it and QC, which leads to poor file organisation and inefficient meta-analyses.

Enter Flow:

a database **and** analysis platform

Flow bundles metadata together with the data objects a Nextflow pipeline produces, keeping data **searchable** and **in context**

Analyses and data are stored in a safe, structured way

Project
Sample
Pipeline Execution

Flow bundles metadata together with the data objects a Nextflow pipeline produces, keeping data **searchable** and **in context.**

Data are stored securely with **custom permissions,** and are shareable on a per-file, per execution or per-project basis.

Interact with data via web app, API or command-line

via running pipelines directly | via Flow's GraphQL API | via the Flow web platform

Generate and access the same data and analysis results, every time

Flow is built for the modern web, with the backend in Python, a Javascript React frontend

Flow helps bioinformaticians and biologists collaborate more effectively
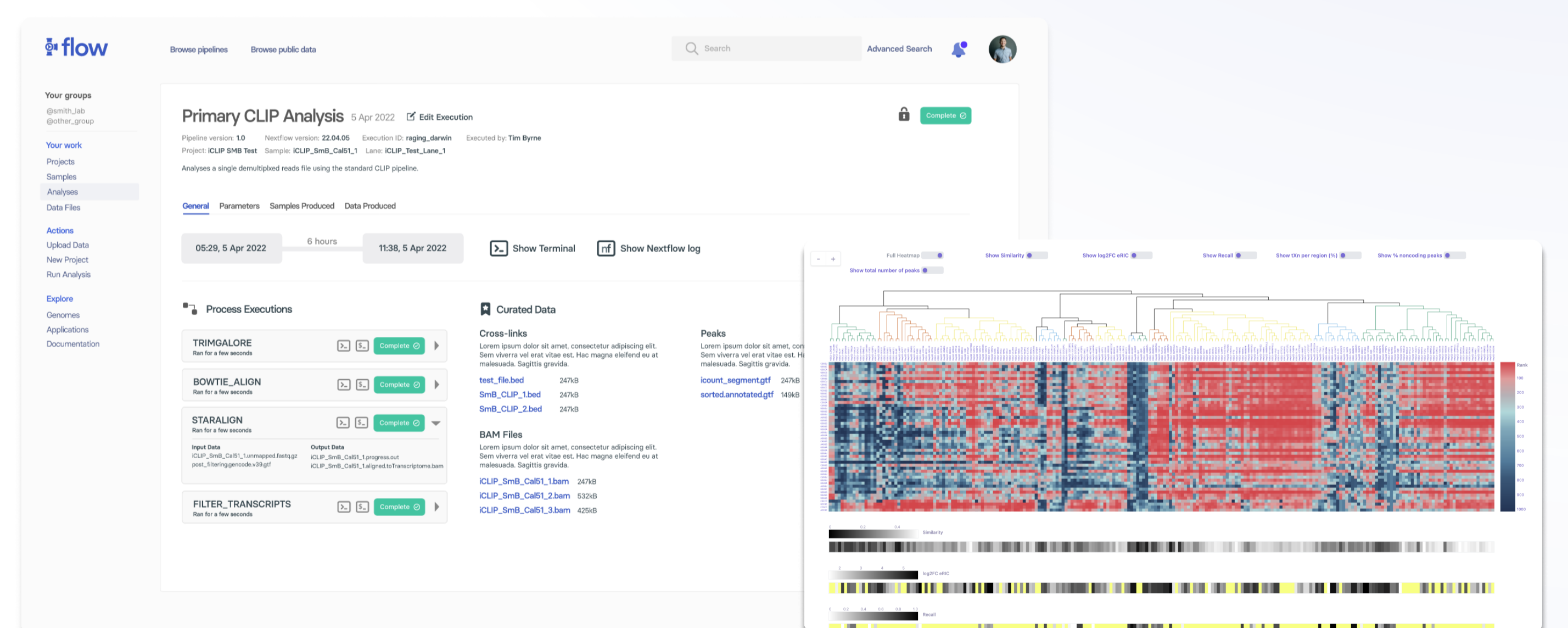

Bioinformatician

Wet lab biologist

Results

Flow lets bioinformaticians write a single JSON config file, and Flow dynamically renders a web interface for the pipeline with validated inputs, **reducing overheads to collaboration**
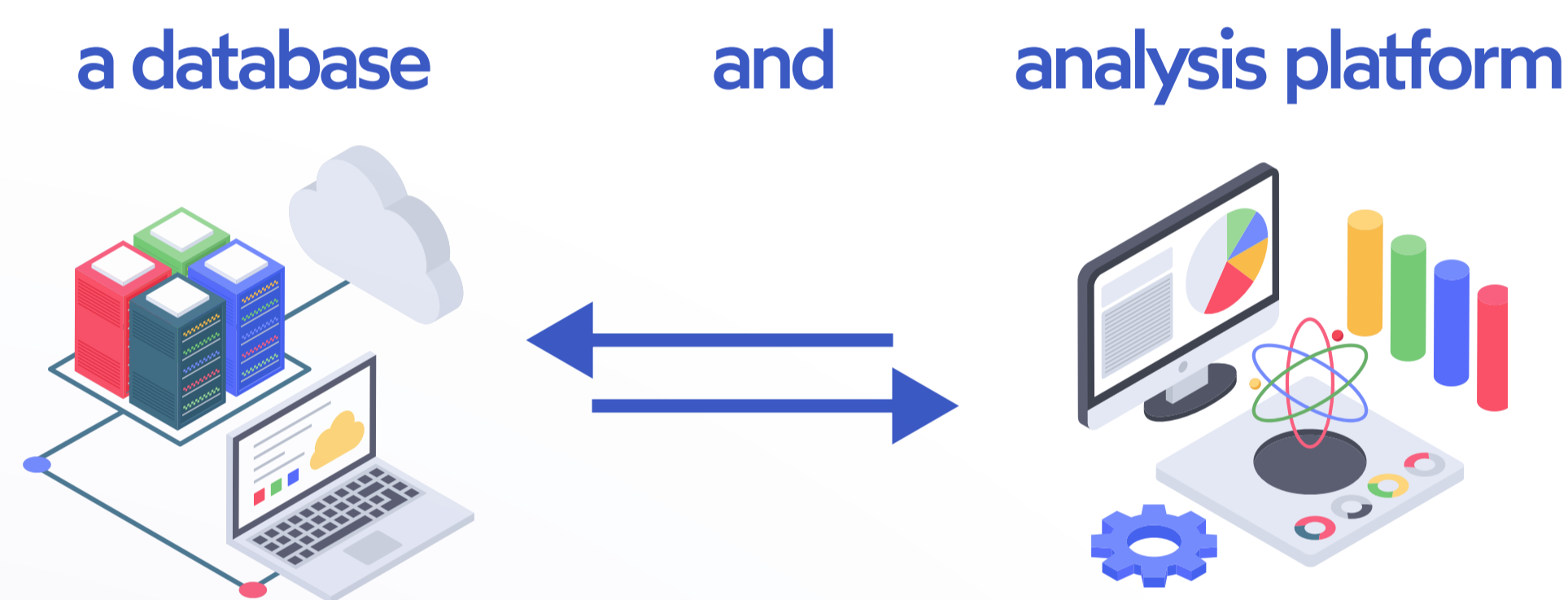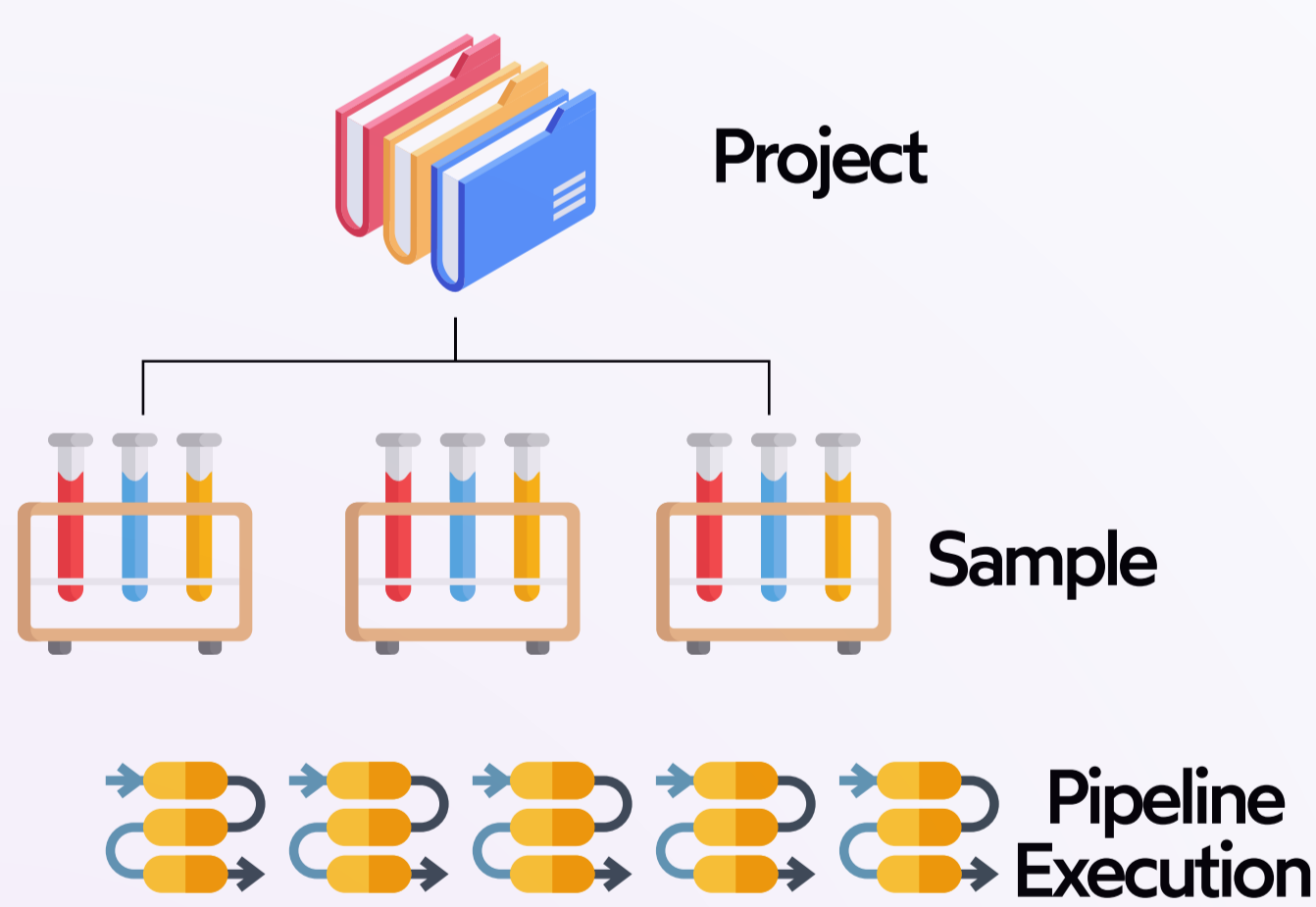
View inputs and outputs to every Nextflow pipeline module as they execute in real time

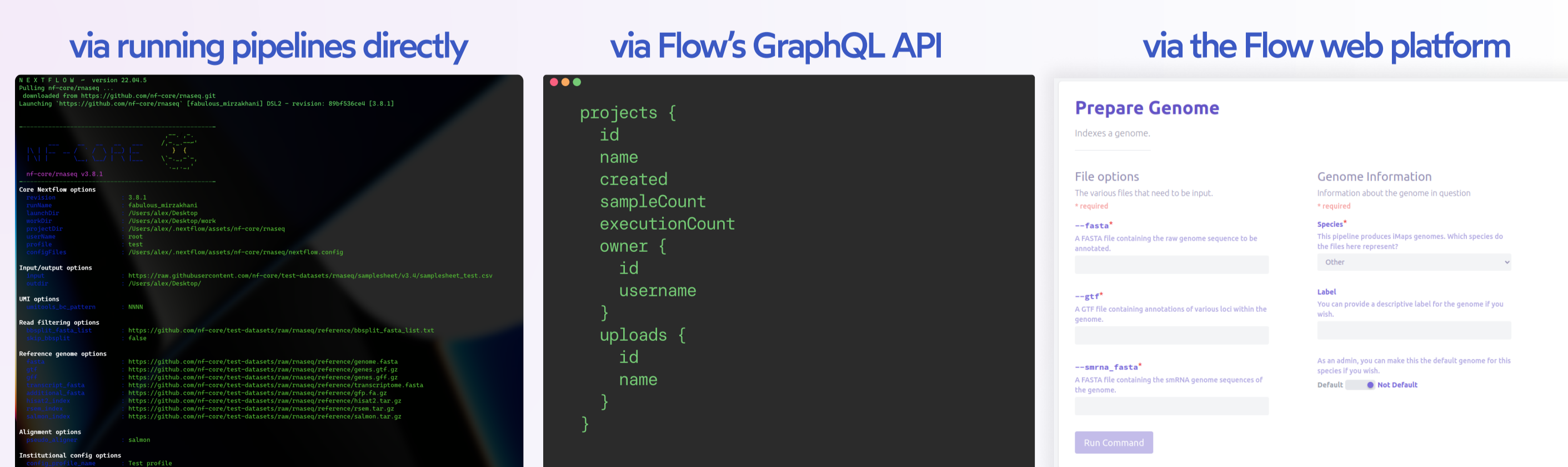## and explore results with custom interactive visualisations

How is Flow different to..

nextflow tower

Tower allows running of Nextflow pipelines in the cloud through a user interface, but doesn't provide the interactive analysis or powerful data annotation and organisation experience of Flow.

Terra

Terra lets users who can code pull data from public databases into workspaces where they can be analysed with code notebooks. Pre-processing is done with a different language called Workflow Description Language (WDL). There is no way to submit your new data to public databases, which rely on a specific team to remain current. Data objects are not directly linked to their metadata or analysis history either.

Galaxy

Galaxy lets users run pipelines through a user interface and store and manage resulting data outputs. Specialist analysis pipelines can be somewhat rigid, with no clear way to provide your input. There is no easy way to integrate your data with all publicly-available data.

Flow is built on: **nextflow** **nf-core** **django** React **GraphQL**

References
The nf-core framework for community-curated bioinformatics pipelines. Ewels et al., 2020. Nat Biotechnology.
Nextflow enables reproducible computational workflows. Di Tomasso et al., 2017. Nat Biotechnology.
GraphQL for the delivery of bioinformatics web APIs and application to ZincBind. Ireland & Martin, 2021. Bioinformatics Advances.