Genedata
solutions in silico

Genedata Profiler & iRODS

An Open & Collaborative Enterprise Software Platform for Patient and Compound Profiling

Marc Flesch, Tamas Rujan

Genedata – Corporate Snapshot

Roots

Established in 1997 | Privately owned | Headquartered in Switzerland

Global Reach

~ 200 employees | Offices in Europe (Basel, Munich), North America (Boston, San Francisco) & Asia (Tokyo)

Dedicated to Drug Discovery & Biotechnology

Innovative portfolio of enterprise systems increasing productivity of data rich & complex research processes

Domain Expertise

Experienced Ph.D. level experts coupled with efficient software engineering processes

Marquee Customer Base

Leading pharmaceutical, biotechnology, and other life science organizations

Customer Base – Pharma

Supporting the Patient Profiling Process

Patient cohorts

NGS

Patient stratification
Drug response prediction

Responder

Non-responder

Major Challenges of Patient Profiling Process

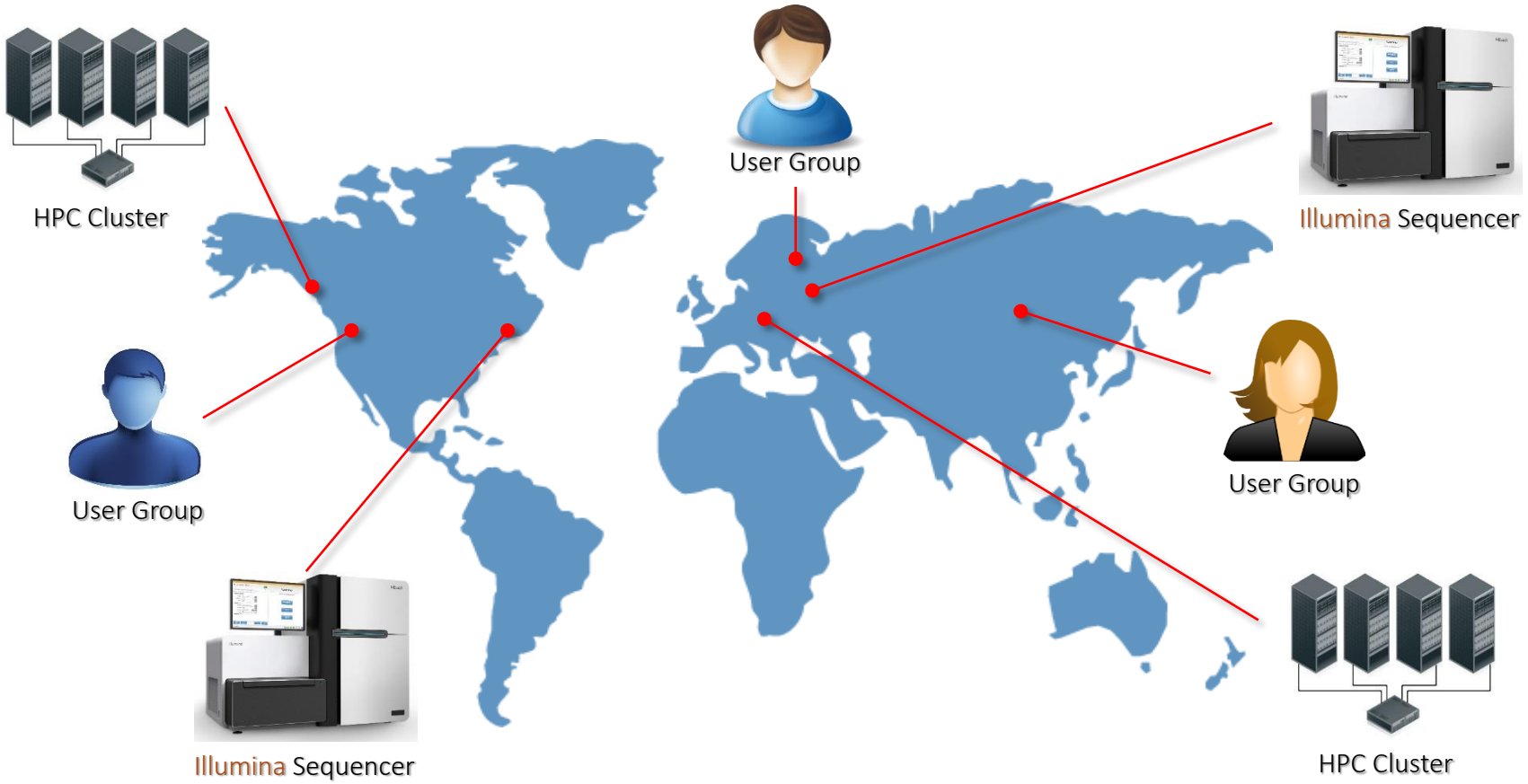
- Efficiently managing, processing, and analyzing data
  - Huge & complex datasets containing patient related omics data
  - Integrating disease & genomic information from different studies
- Facilitating collaboration within interdisciplinary teams
  - Enabling easy data, method & result sharing
  - Global distribution of data generators & data consumers
- **Working with data from human samples in research environments**
  - **Ensuring privacy of patient information**
  - **Maintaining chain of custody**

Problem Statement

"Using data from clinical samples is challenging, because we need to take patient privacy very seriously"

*Henrik Seidel, Bayer

Data privacy within a global Organization



… how-to efficiently work with distributed data?

At Present…

Common technologies applied include

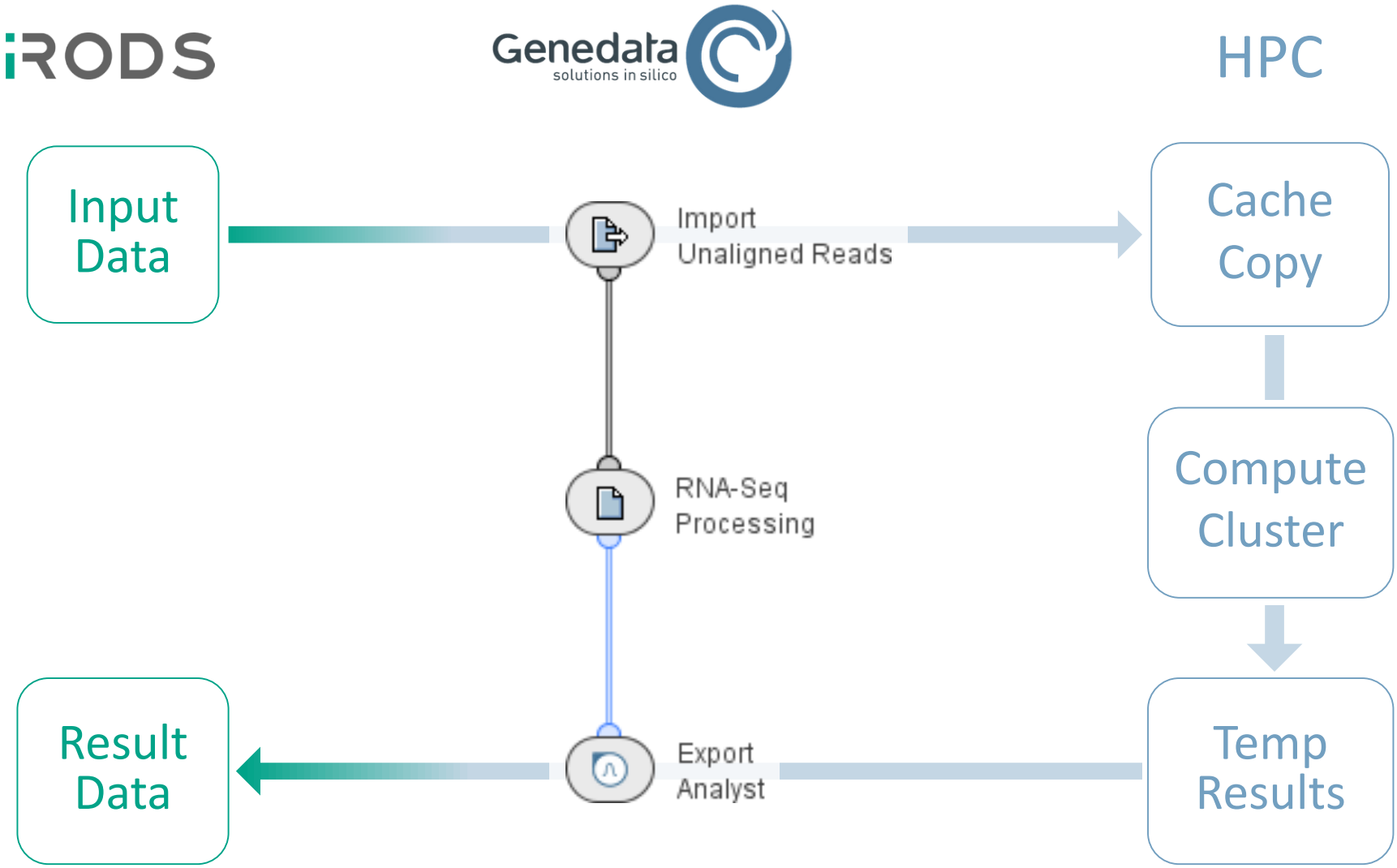
- UNIX file permissions
- POSIX Access Control Lists (ACLs)
- CIFS Shares (SAMBA)

With the following shortcomings

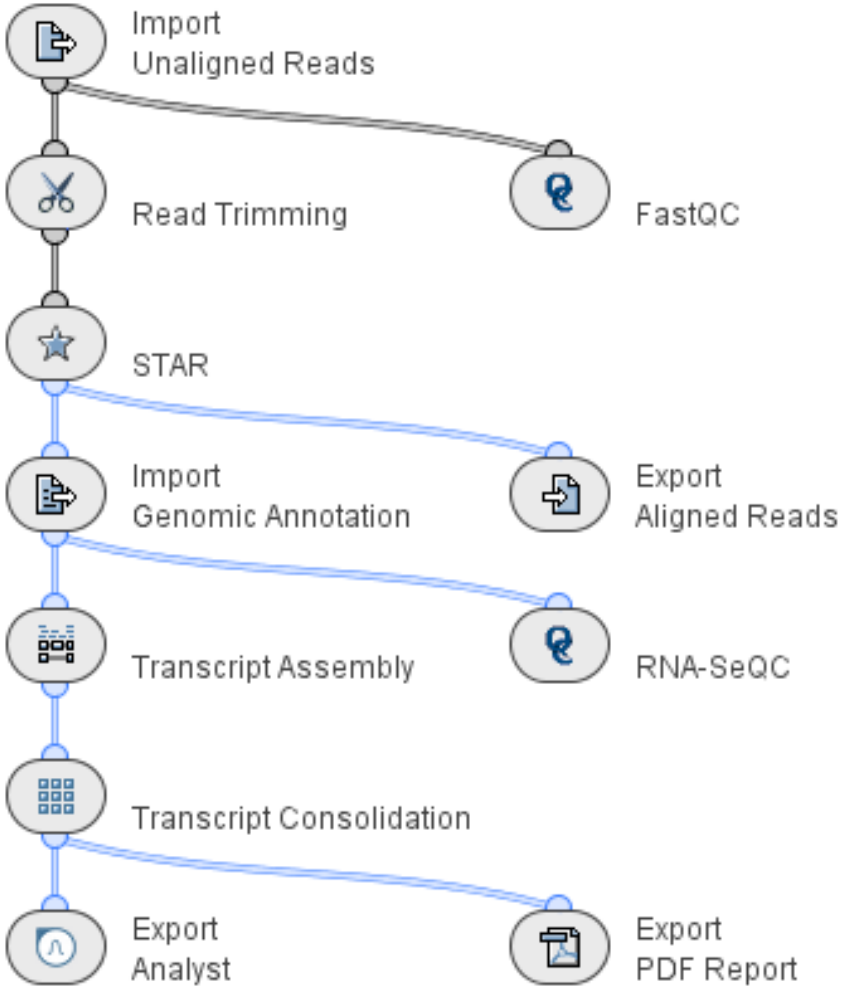
- UNIX permissions are too simple to model project centric access patterns
- paths on UNIX file systems can’t replace data management systems
- permissions have to be maintained manually which is extremely cumbersome
- ACLs are hard to manage
- distributed storage problem stays unresolved

Our Solution

Marrying Security with Performance

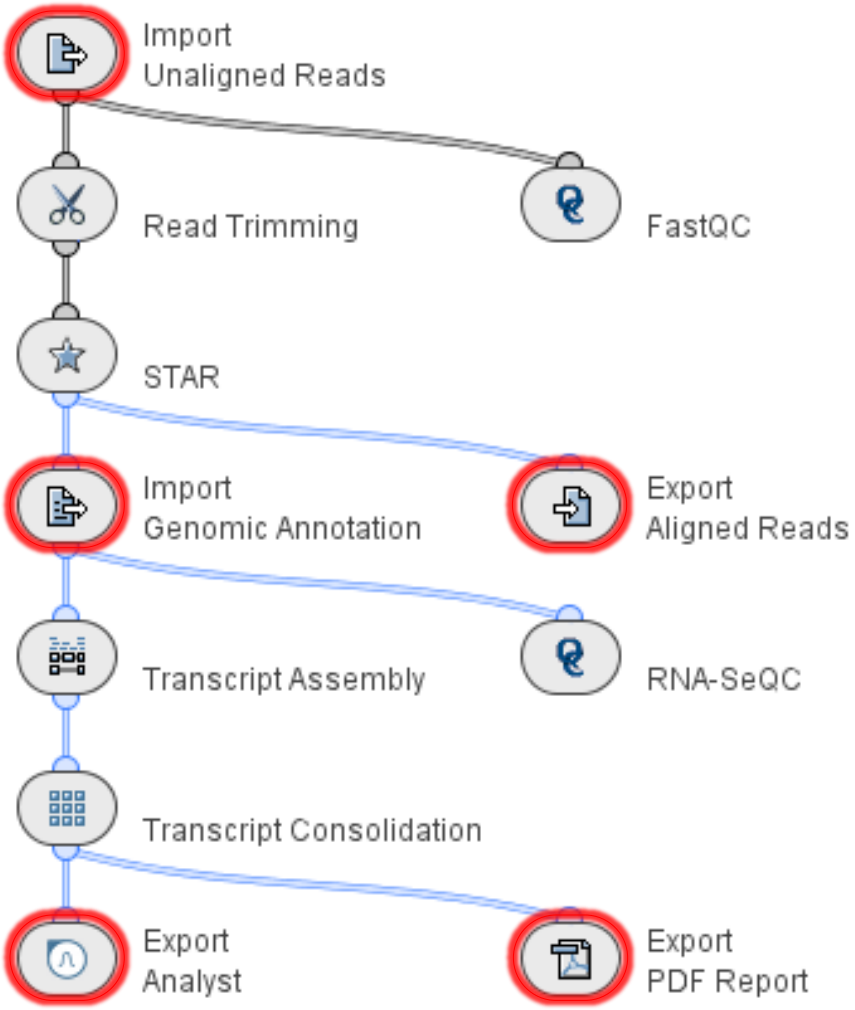


RNA-Seq Data-Processing Pipeline

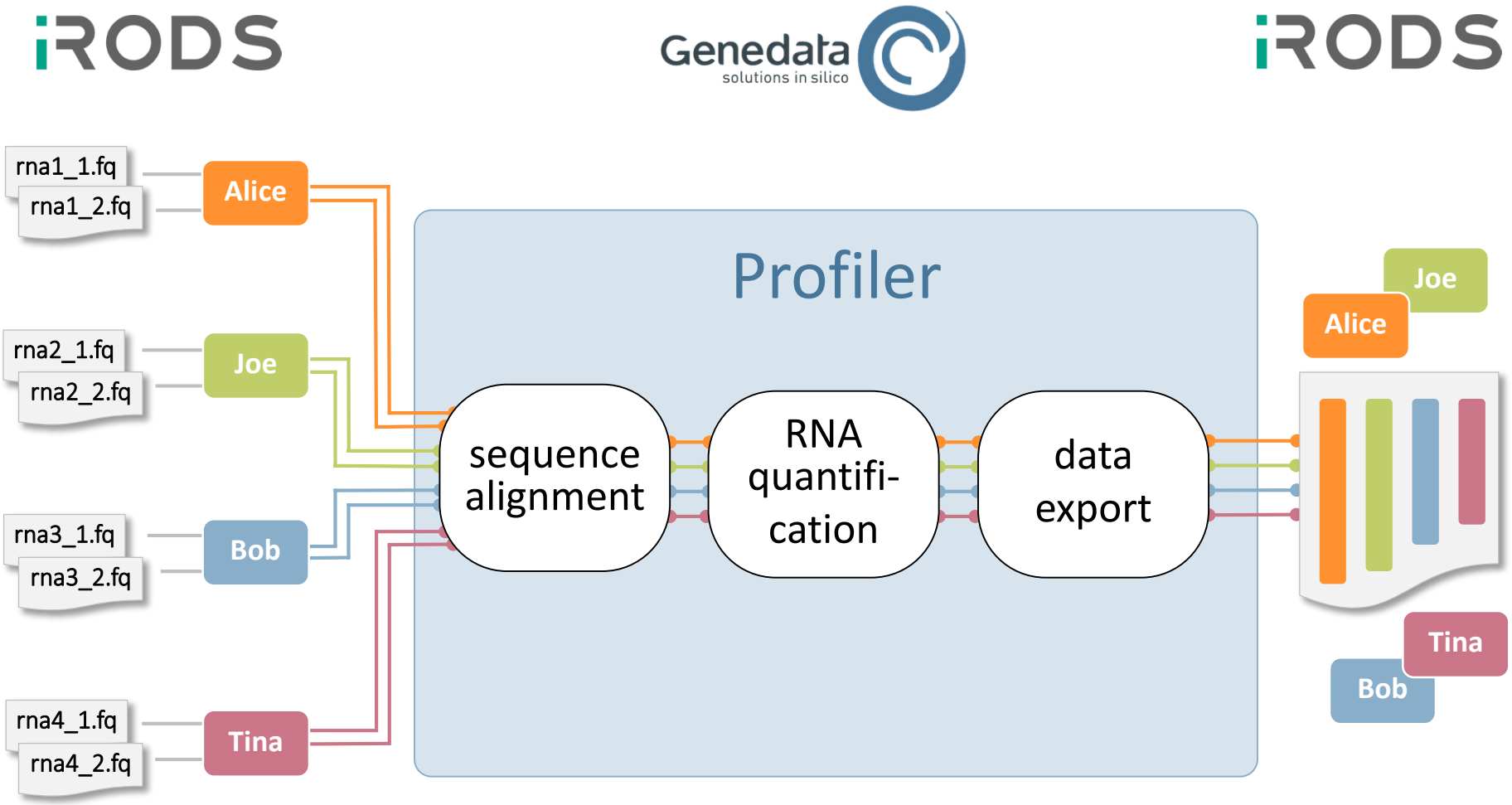


and Interaction Points with iRODS

Genedata
solutions in silico



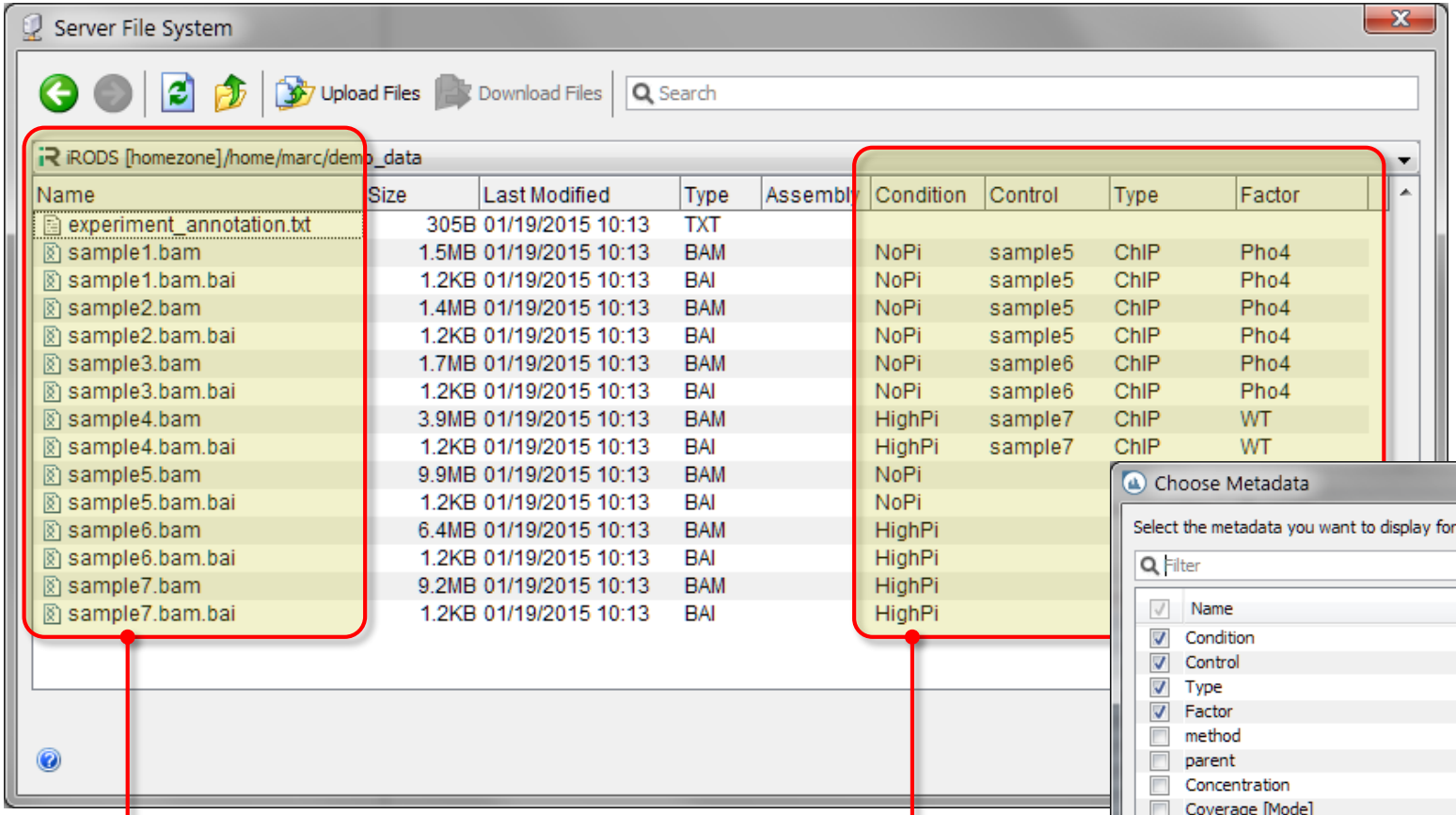
Chain-of-Custody



Enabling Intuitive Raw Data Management

1. Visualization of clinical sample annotation together with corresponding raw data
2. Flexible search functionalities across the whole database
3. Powerful annotation curation capabilities including bulk editing and annotation information protection

Marrying Raw Data with Sample Annotation

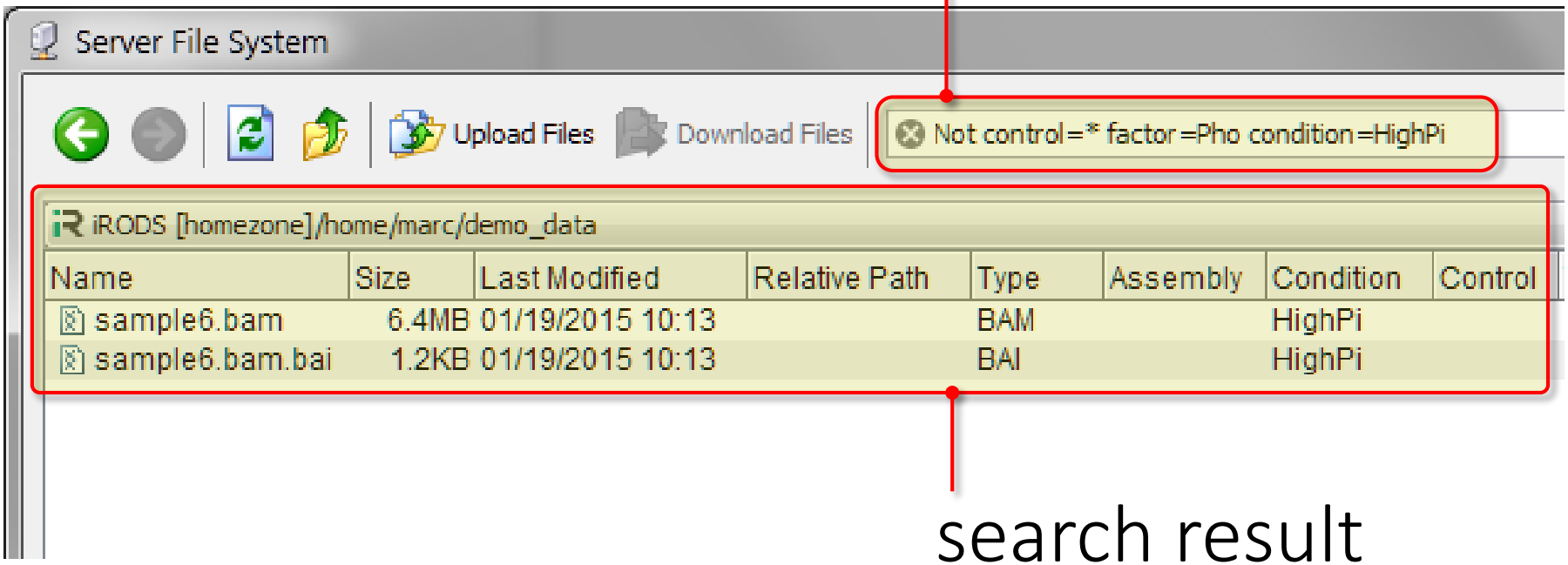


Raw Data

Sample Annotation

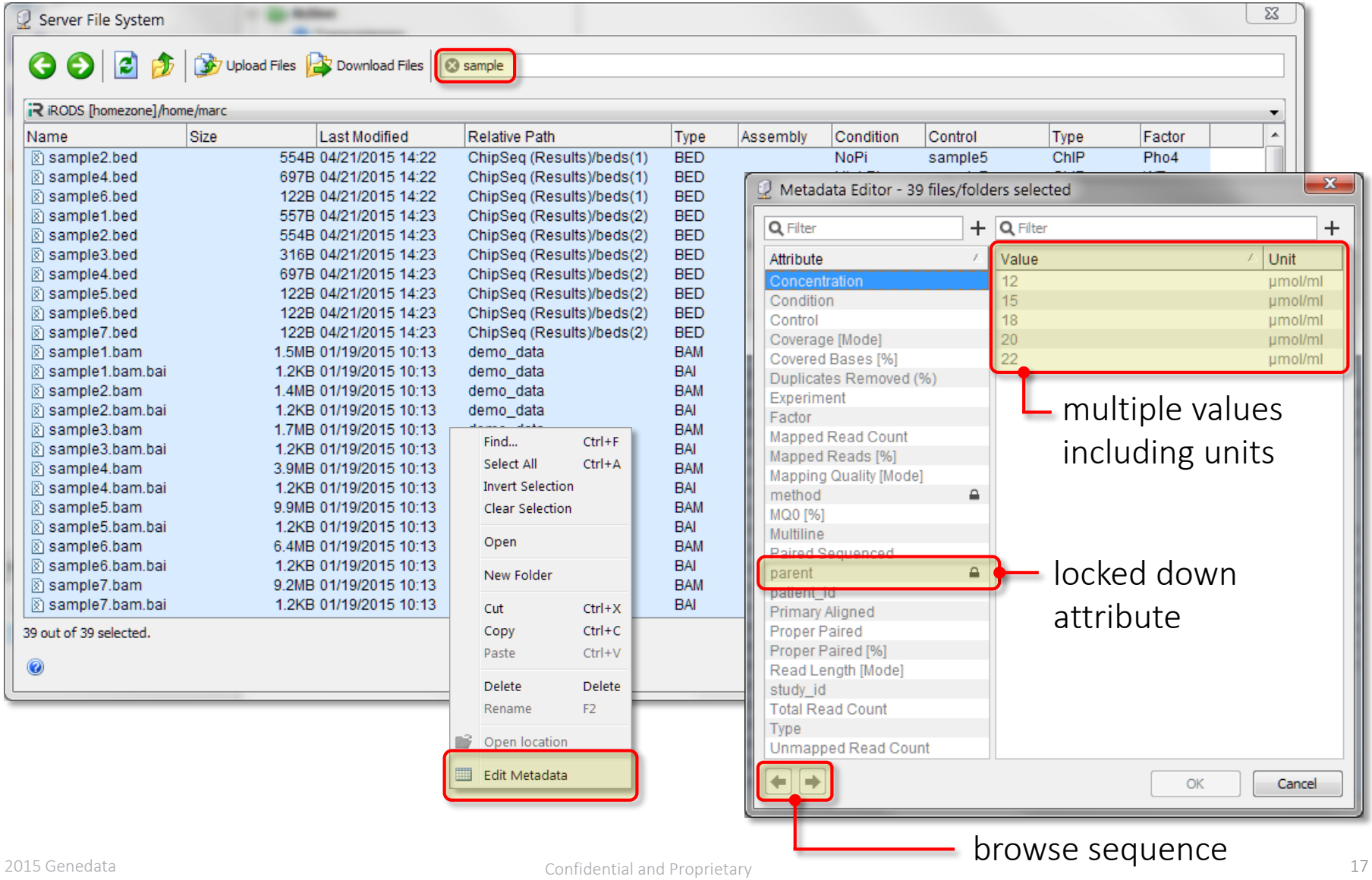
Providing ‘Google-Like’ Search

complex search



search result

Sample Annotation Curation



Summary

- The smooth integration of **Genedata Profiler** with **iRODS** enables scientists to **preserve** their **research eco-system** when working with confidential data
- **Genedata Profiler's** data processing and management capabilities together with **iRODS'** metadata and security concepts are a unique combination to establish the **chain-of-custody** for analyzing personalized medicine data