Pistoia Alliance

Building a business from a Pistoia Alliance project

GENESTACK

"Genestack was formed just three months before the original Pistoia Alliance Sequence Services RFP was issued and by chance we happened to already have identified the same gap in the market. Responding to the RFP so that we could work with the members of the Pistoia Alliance to develop our plans seemed the obvious thing to do."

Misha Kapushesky,
CEO,
Genestack

The Pistoia Alliance Sequence Services project set out to address a widespread issue in the life sciences industry: to efficiently install and maintain local copies of key public software tools and data resources without straining the already limited capacity of bioinformatics teams. In response to an RFP a number of pilot solutions were developed. Genestack, a bioinformatics company headquartered in Cambridge, UK, has developed a successful business around their solution that is, in part, thanks to their involvement with the Pistoia Alliance.

Manual Labour

In bioinformatics labs across the world there is a common need to install public software tools in order to carry out the analyses required by researchers. These tools typically come from the open source community or the academic world and generally do not have much support.

Every organization that deploys any of these tools internally ends up having to invest in maintaining, supporting, and understanding the inner workings of the tool, and possibly going so far as improving and bug-fixing the source code themselves. On top of this they also have to develop the infrastructure required to run the tools. This is costly and generally hard to outsource on an ad-hoc basis, and is clearly pre-competitive.

The problem isn't limited to managing local installations of public software tools. Public data sources, including Ensembl, PubMed, SRA, and many, many others, are a valuable asset not easily recreated outside the academic world. It is important to be able to access them securely, as is being able to access the computational resources required to process the data efficiently.

Many companies wish to use public data by installing copies into a secure sandboxed environment in order to combine them with proprietary or commercially sensitive data. But, as with the third-party software tools they use, they face the challenge of maintaining these local data mirrors, keeping up backwards compatibility, integrating them with their internal resources, and developing infrastructure capable of processing it all.
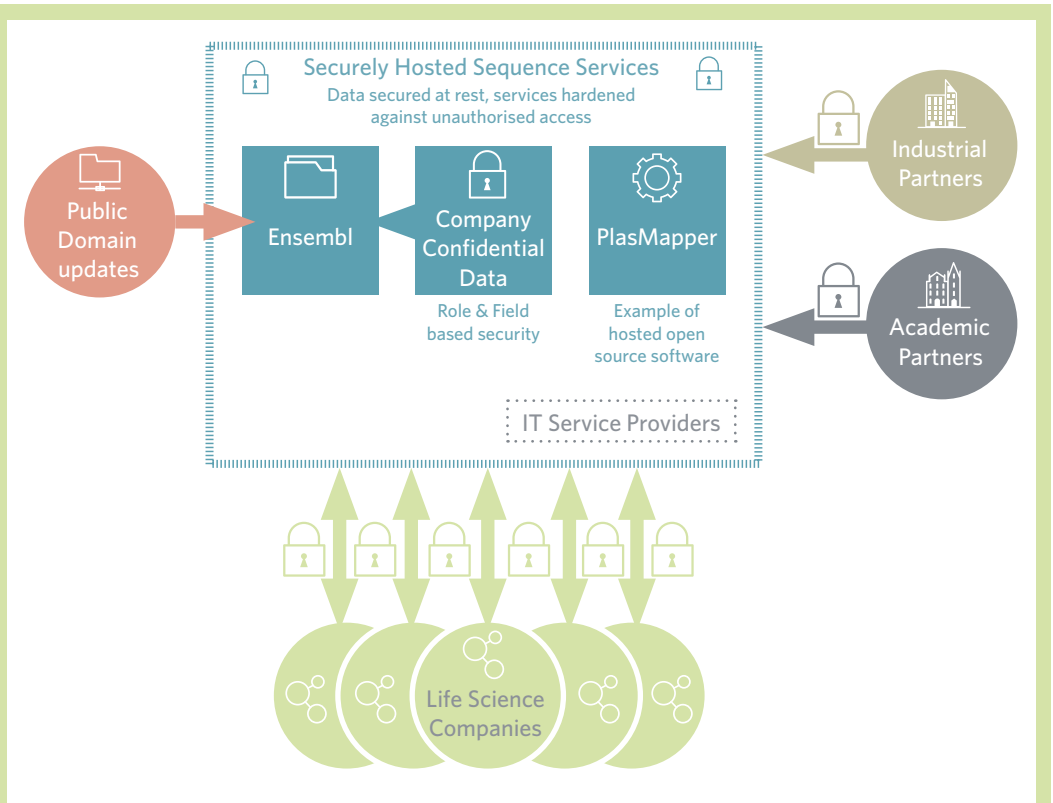
Ecosystem of Solutions

Before the Pistoia Alliance Sequence Services project came along there was no universal solution available. Bioinformaticians simply made local copies of everything they needed in whatever way they saw fit. Occasional attempts were made to build custom solutions, outsource the work, or bring in external help to set up isolated individual instances of key tools or databases, but this approach was by no means uniform.

The pain points of managing local copies of software tools and public data were the core issues that the Sequence Services project set out to address. The project created an ecosystem of commercial Sequence Services solutions via an RFP process. Each of these solutions had the ability to provide secure environments for combining private and public data and tools with a minimum of fuss, backed by commercial-grade support and guaranteed maintenance.

Pistoia Sequence Services Phase 1

Sequence Services phase 1 envisaged a secure access arrangement for a genome browser and a small number of supporting tools and datasets.

Phase 2 expanded this concept to include almost any selection of tool or data to be deployed within a flexible, extensible framework.

Securely Hosted Sequence Services
Data secured at rest, services hardened against unauthorised access

Public Domain updates

Ensembl

Company Confidential Data
Role & Field based security

PlasMapper
Example of hosted open source software

Industrial Partners

Academic Partners

IT Service Providers

Life Science Companies

Expensive Problem

There is an obvious cost in having to maintain local copies of public resources in that it requires manpower to simply install and update each new version that comes out. Additional costs include capital expenditure on obtaining the proper hardware to install and run the resources is one aspect, and the pressure that is put on internal IT teams to deliver and maintain the physical infrastructure required.

But, most painfully, there is still a shortage of qualified bioinformatics specialists that is exacerbated by the need to use them to carry out basic installation and maintenance tasks. In the common situation where there is a computational biology unit consisting of a small number of bioinformaticians providing research support to a large and globally distributed team of scientists, it is a waste of a scarce and valuable resource to have them spend their time compiling software or troubleshooting database mirrors.

The key advantage of having a Pistoia Alliance Sequence Services solution in place is that bioinformatics labs have more time to invest in their core role of providing support to their scientists. It enables them to concentrate on developing new analytical pipelines, designing specialist algorithms, and working on important research questions rather than worrying about infrastructure maintenance issues.

"The Pistoia Alliance catalyzed the industry into thinking about the problem of maintaining local copies of public tools and data and got them together to think about generating a solution. It got people thinking and brought the issue to the fore. Many people were aware of it but had not necessarily thought that the issue was pervasive and that many others shared their pain."

Misha Kapushesky, CEO, Genestack

Screenshots from the Genestack platform.
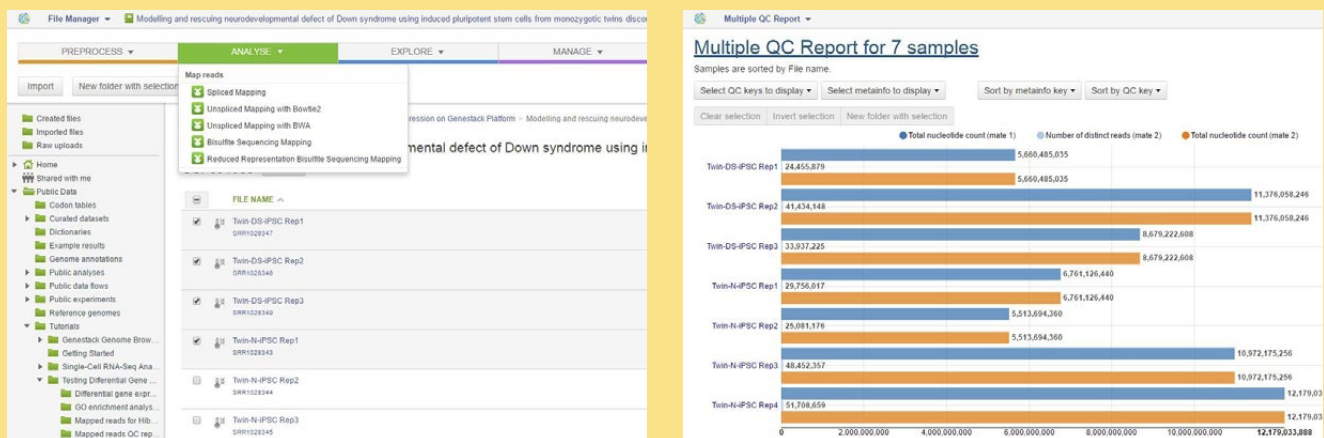
A pilot of this platform was created and demonstrated as part of the Pistoia Alliance Sequence Services project.

Growth of a Startup

Across the two phases of the Pistoia Alliance Sequence Services project a number of companies were involved in responding to the RFPs and developing prototype solutions, often in partnership with one another. These included Cognizant, Eagle Genomics, Cycle Computing, Hewlett Packard, the Functional Genomics Center Zurich, Constellation Technologies, Microsoft, Infosys, and Genestack. Almost three years on from the completion of the last phase of the project, Genestack has built a successful business around the concept.

Involvement in the Sequence Services project served as a useful requirements gathering exercise for Genestack, providing a lot of information within the RFP documentation and subsequent discussions that they would otherwise have had to identify and gather themselves. It also helped expose them to relevant people in the market and gave them a testing ground for their ideas.

The Genestack platform has evolved massively since the initial prototype was completed and presented to the Pistoia Alliance Sequence Services project but the core ideas from the start are still there in the foundations and have been built on ever since.

Genestack has grown from a two-person startup to a 18-strong company. They built this product and took it to market independently. Sequence Services however, provided the crucial validation and a perfect test-bed for their ideas. The boost given to them by working with the Pistoia Alliance in the early stages has accelerated the company's growth.

Genestack currently has pilots and agreements in place to deploy its core product with a number of major pharmaceutical companies, including several of those involved in setting up the original Sequence Services project.

More information

http://www.pistoiaalliance.org/projects/sequence-services/

info@pistoiaalliance.org

http://www.genestack.com
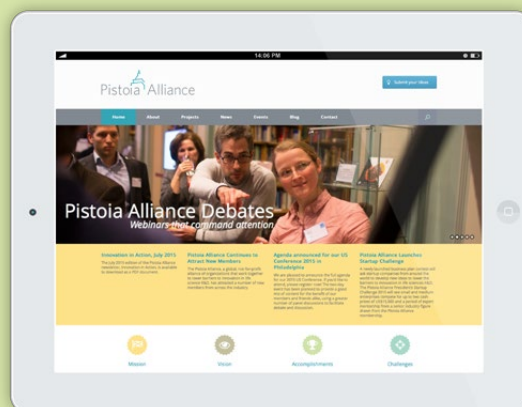
Pistoia Alliance: Lowering barriers to R&D innovation

The Pistoia Alliance is a global, not-for-profit alliance of life science companies, vendors, publishers, and academic groups that work together to lower barriers to innovation in R&D.

Our members collaborate as equals on open projects that generate significant value for the worldwide life sciences community.