

# Trusted Friend Computing: data mining federated OMICS knowledge source

B. Wolf, P. Kuonen, M. Lourenço, J. Stoppani, *University of Applied Sciences Western Switzerland*  
T. Dandekar, *University of Würzburg, Germany*  
D. Atlan, *Phenosystems, Belgium*

## Introduction

Interpretation of OMICs data requires comparison and thus access to a large number of previously analysed samples. But sharing of human OMICs data has many complex legal and ethical constraints. Because of this, many laboratories restrict their comparison to internal OMICs knowledge bases. To solve this problem, we propose a new approach called TFC (Trusted Friend Computing), which lets laboratories transparently and securely share their internal knowledge bases, as well as other resources, especially computing resources. To implement our vision, we launched a research project with two parts. In the first part, we augment an existing distributed computing programming language, POP-Java, to allow for an easier development of distributed applications using the TFC model. This work is freely available as open-source software. In a second phase, we integrate the TFC model into an existing NGS data analysis application, GensearchNGS, to allow for a secure and transparent sharing of anonymized OMICs knowledge bases.

## POP-Java

The POP-Java programming language, a Java language extension for distributed computing, allows programmers to transparently distribute Java objects over the network. In our new model, which we implement as an extension of POP-Java, we want to not only abstract the distribution of the object, but also allow the creation of application specific collaborative networks that can share computing resources in a secure way.

The organization of such a network is done in a way resembling social networks, with only people being able to join that are authorized by one of its members.

Through TFC we offer the developer (in addition to the POP-Java feature):

- Automatic resource discovery inside a secure network
- Secure connections using SSL
- Resource usage overview for statistics or billing

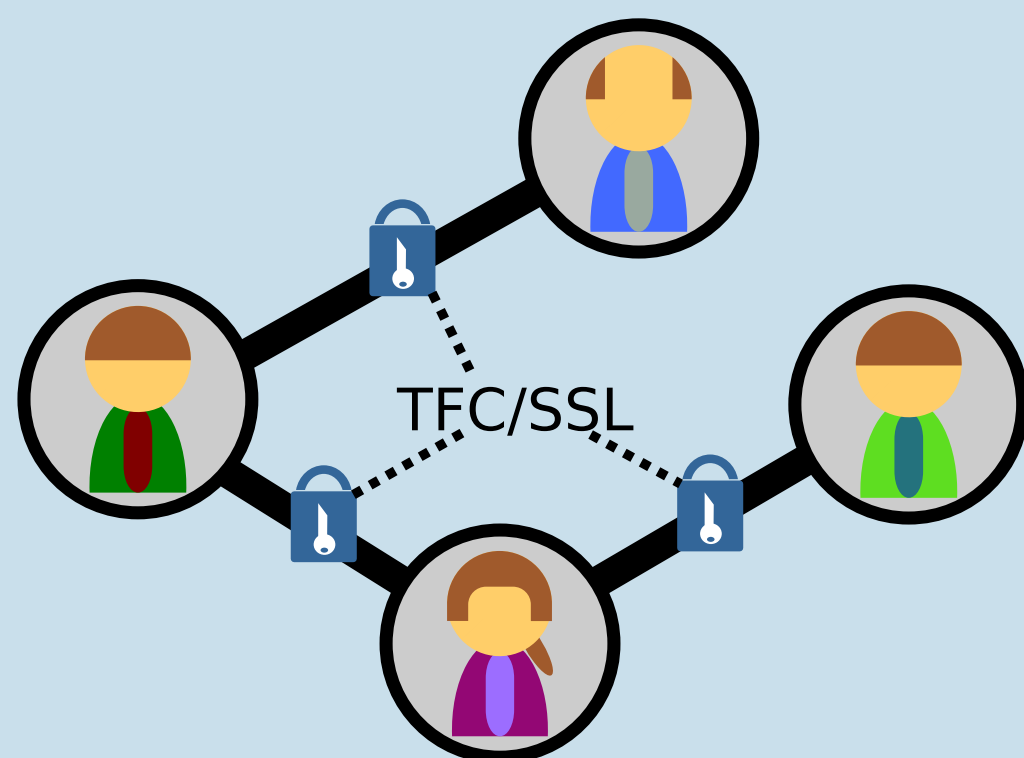


Figure 1: Example layout of a TFC network using secured connections

Those features (and more) are part of an upcoming open-source version of POP-Java.

## GensearchNGS

Using the TFC technology, we augment the NGS data analysis software GensearchNGS developed by Phenosystems SA to allow for various features that take advantage of this secure distributed network. Our initial focus is on the sharing of variant data across multiple laboratories, where partners can check the existence of a variant in the participating laboratories. This initial proof of concept will lead to further features to be implemented, such as distributed alignment and sharing of other types of OMICs data (CNV, etc.).

The access to the OMICs data by members of the secure network can be configured by the participants of the network. But it is to note, that the implemented features will not send sensitive data over the network, but only overview statistics, such as the amount of times a variant has been seen by a certain laboratory, or its classification. This will allow for global statistics as well as anonymized GWAS analyses across laboratories.

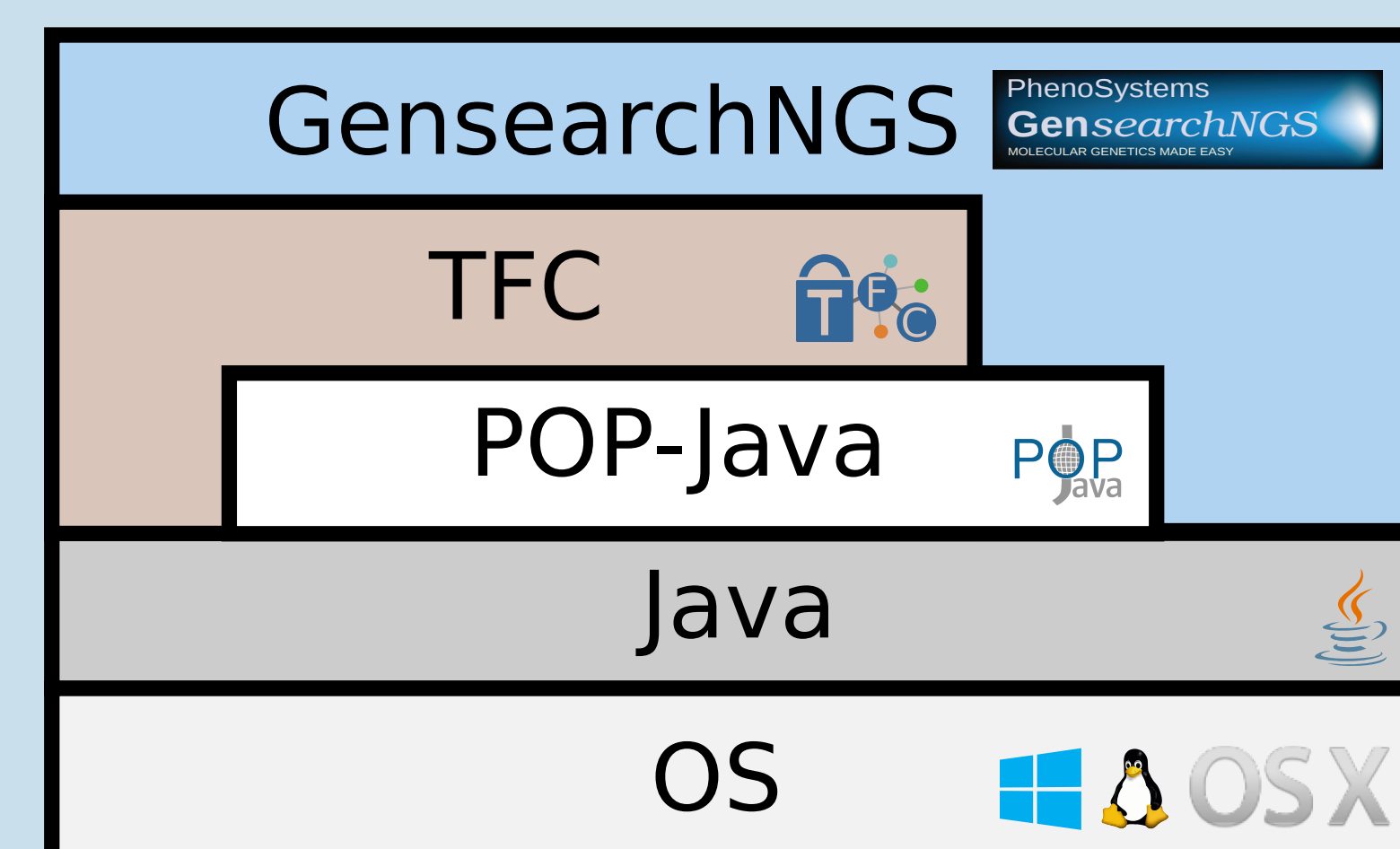


Figure 2: Architectural layers of GensearchNGS using the TFC functionality

## Conclusion

We propose a new method to securely share OMICs knowledge sources as well as computing resources. To achieve those goals, we integrated the TFC (Trusted friend computing) model into POP-Java.

Using the improved distributed programming language, we are prototyping distributed computing functionalities inside GensearchNGS to make OMICs data analysis easier.

Those features include:

- Distributed OMICs knowledge bases (Variants, CNVs, etc.)
- Distributed calculations (Alignment, protein folding, etc.)

The underlying technology, POP-Java and TFC are going to be freely available, with the commercial NGS data analysis software GensearchNGS implementing those features using that technology.

Grants : CTI no. 18781.1 PFES-ES

Contact: Beat wolf, beat.wolf@hefr.ch

Website POP-Java (TFC):  
<https://github.com/pop-team/pop-java>  
Website GensearchNGS:  
<http://www.phenosystems.com/>

