



# Reusable research data analysis platform

Code	IT7664
Programme	FCT
Department	IT
Responsible	28854 - Dr. Timothy John Smith

## Title

Reusable research data analysis platform

## Description

The experimental particle physics data resulting from collisions of accelerated particles in particle detectors is collected and analysed by large collaborations of physicists. The results of individual physics analyses are usually published in the form of scientific papers. In order to test new theories that may appear in the future, it is often wanted to study new models on the existing data even many years after the data taking ended.

The future reuse and reinterpretation of research data analyses requires having access to the original computing environment, the input datasets, the analysis software and the workflow steps that were taken by the original analyst to produce the original scientific results in the first place.

The research data team in the CERN IT department is developing a set of tools that support particle physics researchers in preserving the knowledge around individual physics analysis so that capturing, sharing, and reusing data would become easier.

The applicant will work as part of the research data team on the platform that permits to instantiate research data analyses on the containerised computing cloud. The system allows researchers to structure their analysis in a reusable manner, to submit their analysis workflows on the computing cloud and to interact with running jobs through live user interfaces.

The developed source code is part of free, open source software projects REusable ANALyses <https://github.com/reanahub>, CERN Analysis Preservation <https://github.com/cernanalysispreservation>, CERN Open Data <https://github.com/cernopendata>, and Invenio digital repository framework <https://github.com/inveniosoftware>.

## Skills

Databases: PostgreSQL.

Information Technologies: Building web applications (e.g. with jQuery, HTML5), Developing with virtualised infrastructure (e.g. Openstack), Using software development tools (e.g. Git, Jira, Trac), Using web application frameworks (e.g. Django, GWT, APEX).

Programming Languages: Python, Shell Script

Programming (Python, shell), databases (PostgreSQL), source code management tools (Git), web application frameworks (Flask, REST API), message queues (RabbitMQ), container technologies (Docker, Kubernetes, Mesos), distributed object storage systems (Ceph, EOS), access control and authentication (OAuth).

## Disciplines

Information Technologies