Data scientist position

Within the context of the project SciQLOP, the Laboratory of Plasma Physics (LPP) is hiring an engineer for computing high level data products from multiple spacecraft mission databases.

Context:

The SciQLOP project, started in 2015, aims at providing scientists with a flexible, efficient and intuitive tool to analyze space plasma data from multiple in situ spacecraft mission. The project is composed of five main parts:

- a **multi-mission multi-server data tree**, targeting servers of specific spacecraft missions and space agencies and providing high level data products ready for science without requiring researchers to handle file formats, server specific details etc.
- a **plot engine,** enabling plotting hundreds of thousands of points, real time computing, multiple visualization modes, and strong interactivity
- a **python terminal**, providing users with limitless options to analyze data with community or user developed python toolkits
- a **catalog module**, for users to save data intervals associated to specific events, publications, etc. associated with metadata, with possibilities to build catalogs in collaboration with other users.
- A **machine learning engine**, enabling scientists to automatically detect structures in data, augmenting visualization with rich metadata, add events to catalogs and build models.

SciQLOP is a graphical user interface program, open source and mainly written in C++ using the framework Qt. The team is composed of researchers and engineers working in close collaboration. The project is funded by Universite Paris-Saclay, through the projects Center For Data Science 1.0 (CDS 1.0), and the project SPACEOBS.

Mission:

Your mission consists in participating to the development of the multi-mission data tree. In the SciQLOP application, each spacecraft mission is registered via a *mission plugin*, providing the data tree with all of the mission's products. When a user drags & drops a high level product (distribution functions in various basis, 2D slices, projections or 3D, particle and field spectrograms, etc.) from the data gate to the plot area, the associated mission plugin retrieves all raw data necessary (either via download on the appropriate server, or from the cache) and must construct the required product before sending it to the plot engine for visualization. You will work on implementing routines that compute high level data products from raw data products. Some parts of the code will be highly tuned to details associated with specific spacecraft instruments, some other parts will have to be abstracted and used across multiple missions sharing similar data types.

Required skills, experience and knowledge:

The successful candidate will necessarily demonstrate the following points:

- working knowledge in C++ (preferably) or C, with at least a year of experience
- working knowledge in interpreted languages such as python
- several years of experience in data analysis in physics
- be able to autonomously read, understand and use scientific articles describing spacecraft instrumentation and data products
- to be a team player and be able to autonomously discuss with scientists and engineers specialized in specific missions and data products
- to be curious, rigorous and strongly motivated

Desirable skills, experience and knowledge:

All points below are strong assets:

- having experience in data analysis of in situ space plasma data, either as a engineer, or phd/postdoc
- good knowledge in python packages such as numpy, matplotlib, scipy, spacepy etc.

Position:

The position is a full time 1 year fixed term contract. The salary will be fixed depending on the candidate's experience. The work location will be at Ecole Polytechnique, Palaiseau, France. Please contact sciqlop@lpp.polytechnique.fr for any question regarding this position. Applications should be sent to sciqlop@lpp.polytechnique.fr, and enclose

- a CV,
- a short (1 page max) *personal* motivation letter,
- possible list of publications if relevant,
- a description (and URL if possible) to previous projects and codes related to data analysis.
- contact information of previous supervisor(s)

Selected applicants will be contacted for interviews. The start date is as soon as possible but the offer is open until the successful candidate has been found.