

Sharing data science artifacts across teams using Crane

Lucianos Lionakis*

Résumé (max 300 mots)

Does your organisation have to share many data science artifacts across teams? This is a problem for many data science organisations and can now be solved using a novel open source product Crane (<https://craneserver.net/>). Crane hosts data science artifacts such as data analysis reports, documentation sites, or even packages and libraries. All of these data science artifacts are kept under strict authentication and authorization using modern protocols (OIDC). In this talk, we walk you through the different features of Crane and provide a live demo to explain the concepts. We will discuss its configuration file and demonstrate that authentication (and authorization) in Crane is fully declarative and allows for fine-grained configuration (at user-level, group-level, network-level or using custom expressions (SpEL)) while still using an intuitive hierarchical tree that corresponds to the directory structure of the data. Next, we will show how artifacts can be accessed from or uploaded into Crane using the Crane API from R (e.g. to automate report updates, use data science artifacts in CI/CD) or using its customizable UI (e.g. for manual tasks and browsing). Further, we zoom in on audit logs to track operations on all files (e.g. for GxP purposes) and detail the different storage backends (a.o. S3 and local file system). To ensure Crane can perform in high security settings the code base has been tested using integration tests reaching a high code coverage of more than 70%. With this talk we want to teach any R user and developer the essentials of Crane and how it can be used to share his/her data science artifacts.

Mots-clefs (3 à 5) : Ingénieure - Data - Package - Reproductibilité

Références

<https://craneserver.net/>

*Open Analytics, lucianos.lionakis@openanalytics.eu