

Le CERN passe le LHC en mode Open Data

Le **LHC** (*Large Hadron Collider*) est le plus grand accélérateur de particules de la planète, dont la taille et les capacités ouvrent la voie à un nouvel éventail de possibilités dans le monde de la recherche scientifique.

Le **CERN** franchit aujourd'hui une autre étape dans le développement de ce projet, avec la mise en ligne d'un portail dédié à la **publication de jeux de données en Open Data**. « *La mise en ligne du portail Open Data est une étape importante, mais la fin de la route n'est pas encore atteinte. L'Open Science représente beaucoup plus que la somme d'actions 'ouvertes' : c'est un idéal, et pour nous, au CERN, un retour à nos racines,* » explique l'organisation dans son communiqué.

Des téraoctets de données...

Accessible depuis l'adresse opendata.cern.ch, l'offre Open Data du CERN propose des ensembles de données issus des quatre grands détecteurs installés au sein du LHC : **Alice, Atlas, CMS et LHCb**.

Les données extraites de l'expérimentation [CMS](#) (*Compact Muon Solenoid*) sont de loin les plus imposantes. Certains *datasets* dépassent ainsi les **2,5 To**. Des jeux de données plus compacts et dédiés à des domaines d'analyse plus restreints sont toutefois proposés aux visiteurs.

L'expérimentation [Alice](#) a permis de fournir des ensembles de données pouvant dépasser les **600 Mo**. Là encore, des sous-ensembles plus compacts, et donc plus faciles à exploiter, sont livrés au public. De premiers éléments pour les capteurs [Atlas](#) et [LHCb](#) sont également publiés. Notez enfin que divers frameworks et outils sont de la partie.

... issus de pétaoctets de mesures

L'ensemble d'archives proposé aujourd'hui par le CERN est toute à fait impressionnant par sa taille. Il ne sera toutefois probablement exploitable que par un public d'initiés, du fait de la complexité de mise en œuvre de ces jeux de données.

*« Ce sont des données issues d'événements réels de collision. Il ne faut pas sous-estimer leur complexité, **ni le temps et les efforts requis par les nouveaux arrivants** pour maîtriser les outils et les techniques nécessaires pour les interpréter, précise le CERN. Au côté des données de bas niveau proposées sur notre portail, nous publions des ensembles de données plus restreints, qui permettront de faciliter la manipulation et l'analyse des données. »*

Le portail Open Data est l'aboutissement d'une collaboration de grande ampleur. Avec les équipes des quatre expériences du LHC, d'un côté, et de l'autre, les experts des bibliothèques numériques, de la conservation des données et de l'application de métadonnées.

« Cela représente également le rapprochement entre deux domaines distincts : les bibliothèques numériques et la gestion des données massives. » Un secteur dont le CERN est devenu l'un des pionniers, suite à

l'ouverture du LHC. Ce dernier génère en effet environ **15 pétaoctets de données** à traiter chaque année.

Sur le même thème

[Red Hat devient le socle des applications critiques du CERN](#)

[L'INPI livre 4,2 millions de documents en Open Data](#)

[Open Data : le gouvernement libère les données publiques](#)