

Open Data : Yahoo offre un jeu de données de 13,5 To

Yahoo vient de mettre à disposition des chercheurs et développeurs [un jeu de données massif](#), comprenant environ **110 milliards d'enregistrements** liés à l'activité d'environ 20 millions d'utilisateurs. Le tout correspond à des relevés faits entre février et mai 2015. Le document proposé atteint les **13,5 To**, soit un fichier compressé de **1,5 To**.

Que contient cette archive ? Des données sur les interactions utilisateur liées **aux flux de news de Yahoo**. Des données anonymisées, bien entendu, mais suffisamment complètes pour permettre la tenue d'analyses poussées. Un résumé des actualités consultées est proposé et – côté utilisateurs – la ville, l'âge et le genre.

Les informations publiées sur la page de garde de Yahoo sont concernées, mais aussi celles publiées sur les services **Yahoo News, Sports, Finance, Movies** et Real Estate.

Des données à foison

Yahoo propose de nombreuses autres sources de données à destination des chercheurs. **55** sont accessibles [depuis cette page web](#).

Certaines sont particulièrement lourdes, en particulier celles liées à **Yahoo Answers** (166 Go ; 3,8 Go pour les questions en français), les jeux d'images de **Flickr** (ECM1, 83 Go ; Creative Commons, 14 Go), une sélection de **formulaires HTML** extraits de pages web publiques (plus de 50 Go), les **logs Hadoop** de la firme (8,8 Go), etc.

À lire aussi :

[Yahoo passe du spleen au split de ses activités](#)

[Yahoo confirme sa préférence pour Alibaba](#)

[Yahoo : résultats en baisse, titre en baisse](#)

Crédit photo : © 360b – Shutterstock