



StarQube

INVESTMENT DECISION IN A BOX

StarQube ESG #2

Collecte, nettoyage, hébergement, transformation
des données ESG brutes



Collecte, nettoyage, hébergement, transformation des données ESG brutes

Dans cette série d'articles, nous soulignons les défis pratiques auxquels les asset managers sont confrontés face à la multiplication des données ESG.

Notre billet précédent était consacré à la gestion des tables de correspondance – notamment entre émetteurs et instruments financiers. Nous nous pencherons ici sur les enjeux en termes de collecte, de nettoyage, d'hébergement et de transformation des données ESG brutes.

A l'échelle d'un portefeuille de 200 titres et avec un fournisseur de données ESG unique, l'intégration des données ESG peut sembler simple et pouvoir être gérée manuellement sur une feuille Excel. Mais c'est rarement la réalité des asset managers – qui sont rapidement confrontés à des enjeux en termes de volumétrie nécessitant la mise en place d'un processus ESG industriel.

Prenons l'exemple d'un asset manager qui souhaite avoir une couverture ESG aussi exhaustive que possible sur son univers d'investissement global et cross-asset. Sa « volumétrie ESG » sera fonction des paramètres suivants (les chiffres qui suivent sont des ordres de grandeur) :

Couverture émetteurs / instruments : ~10.000 émetteurs / 500.000 instruments

- 8.500 émetteurs dans l'univers ESG de MSCI, 11.000 dans l'univers ESG de S&P.
- MSCI indique couvrir 680.000 instruments financiers.

Nombre de champs de données par émetteur : 50+ indicateurs (voire 1.000+ champs)

- *En s'appuyant sur les scores ESG de fournisseurs externes* : 1 score ESG composite + 3 scores par piliers (E, S et G) + 10 scores par thèmes + 35 scores par sujets-clés pour MSCI soit a minima 50 indicateurs par émetteur x nombre de fournisseurs de données ESG auxquels l'asset manager souscrit.
- *Pour une collecte directe des données auprès des émetteurs* : ce sont plus de 1.000 champs de données que MSCI ou S&P indiquent collecter auprès des émetteurs.
- Les obligations de transparence des asset managers et leur souci d'avoir une perception fine des risques ESG qui pèsent sur les émetteurs les poussent de plus en plus à collecter les données ESG brutes plutôt qu'à se contenter de scores externes.

Profondeur historique : ~15-20 années + nécessité d'augmenter la fréquence de collecte

- 17 années de profondeur historique pour MSCI.
- Si les scores sont aujourd'hui construits sur une fréquence annuelle, il est vraisemblable qu'à l'avenir les investisseurs souhaiteront disposer de scores beaucoup plus dynamiques, qui s'ajustent en cours d'année avec le newsflow ; dans un prochain papier, nous expliquerons pourquoi il est indispensable, d'ores et déjà, d'organiser ses données ESG afin de construire des indicateurs dynamiques dans le temps et horodatés (« point-in-time »).

Et bien évidemment, tous ces scores (a minima les scores composites) doivent pouvoir être distribués dans l'ensemble des portefeuilles de la société de gestion à des fins de gestion de portefeuille et de reporting (interne, commercial, réglementaire).

Il est dès lors facile de comprendre qu'il est impossible de construire un processus ESG industriel sans fondations-données extrêmement solides, qui facilitent :

1. La collecte des données
 - Les mises à jour des données ESG, généralement en provenance de fournisseurs multiples et livrées sous des formats divers, doivent pouvoir être collectées de la manière la plus automatique qui soit.
 - Quand bien même les indicateurs ESG sont le plus souvent renseignés par les émetteurs sur une fréquence annuelle, tous les émetteurs ne communiquent pas de manière synchronisée, de sorte que les données ESG brutes (et les scores composites calculés par les agences) font l'objet de mises à jour tout au long de l'année.
2. Le nettoyage des données entrantes
 - Les fichiers transmis par les fournisseurs ne sont jamais propres à 100% : données manquantes, incohérentes (sauts de données par rapport à l'information précédente) ... Les sociétés de gestion doivent mettre en place des processus automatiques d'identification et de correction des erreurs les plus manifestes dans les données collectées.
3. L'hébergement des données
 - Les choix relatifs à l'infrastructure et à l'hébergement des données ESG sont très structurants quant à la capacité des équipes opérationnelles (analystes ESG, chercheurs quantitatifs, gérants de portefeuilles, reporting officers ...) d'accéder à celles-ci et de les exploiter efficacement.
4. La transformation et l'enrichissement des données
 - En l'absence d'un cadre ESG standard, les fournisseurs de scores ESG ont développé des nomenclatures propriétaires : les scores ESG de MSCI vont ainsi de AAA (pour les meilleurs émetteurs) à CCC (pour les plus mauvais), tandis que les scores S&P sont construits sur une échelle 0 - 100 (100 pour les meilleurs). Les sociétés de gestion doivent donc a minima normaliser les scores ESG de leurs fournisseurs pour construire des reportings cohérents ou bâtir des scores composites.
 - Mais la plupart des sociétés de gestion ne se contentent pas de collecter des scores ESG auprès de fournisseurs externes et préfèrent développer une méthodologie ESG propriétaire qui reflète leurs priorités environnementales, sociales et de gouvernance. Sans une infrastructure-données souple, la mise en place d'une méthodologie ESG propriétaire est vite complexe pour de larges univers d'investissement.
 - Les données doivent aussi pouvoir être enrichies – par exemple en remplaçant une information non renseignée sur un émetteur par une moyenne ou une médiane sectorielle.

CE QU'OFFRE STARQUBE

StarQube est optimisé pour la gestion (et l'exploitation) de gros volumes de données et offre les capacités suivantes :

1. *Collecte / Nettoyage*: StarQube dispose de connecteurs « sur étagères » avec la plupart des fournisseurs de données (financières ou extra-financières). Ces connecteurs permettent de récupérer au fil de l'eau les nouveaux fichiers mis à disposition (quels qu'en soient le format et le mode de livraison), de les mettre en forme et d'insérer tous les tests de complétude / cohérence / correction souhaités. De nouveaux connecteurs peuvent être construits très rapidement par StarQube au gré des besoins des utilisateurs, qui ont aussi toute liberté de construire leurs propres connecteurs en s'inspirant des modèles et de la documentation fournis. Les connecteurs sont « schedulés » pour récupérer automatiquement les données mises à jour, sans déclenchement manuel de la part des utilisateurs.
2. *Hébergement*: les données sont hébergées dans une base de données NoSQL optimisée pour les calculs lourds (backtests, optimisations de portefeuilles) sur de gros volumes de données, et nativement point-in-time (toutes les données sont horodatées – lire notre prochain article sur l'importance d'horodater ses données). La plupart des calculs effectués « à la volée » sont instantanés, les backtests les plus complexes pouvant prendre quelques dizaines de secondes, quelques minutes au maximum.
3. *Organisation*: les données sont organisées autour (1) d'un référentiel unique (identifiant unique par émetteur et par instrument) qui facilite la navigation entre jeux de données, et (2) d'une bibliothèque d'organigrammes de groupes qui permet de faire redescendre naturellement les scores ESG des émetteurs vers l'ensemble des instruments financiers émis par les sociétés du Groupe – lire notre article précédent relatif au « mapping ».
4. *Transformation / Enrichissement*: StarQube dispose d'un langage « low-code » qui offre une totale souplesse pour transformer les données ESG brutes en scores propriétaires ou pour enrichir les données en complétant les informations manquantes (à partir de moyennes ou médianes sectorielles/pays ...).

[LIEN VERS L'ARTICLE INTEGRAL](#)

A propos de StarQube

Fondée en 2013, StarQube développe une solution innovante et modulaire à destination des sociétés de gestion d'actifs qui s'articule autour de deux piliers. Le pilier **data management** permet d'industrialiser la collecte, la mise en qualité et l'organisation de tout type de données utiles pour le processus de gestion au sein d'une base de données NoSQL centralisée. Le pilier de **construction et gestion de portefeuilles** permet d'analyser l'univers de recherche, de construire des modèles de risques propriétaires, de créer des indices ou portefeuilles modèles, de modéliser et de backtester des stratégies d'investissement, d'optimiser et de rebalancer des portefeuilles. Des interfaces graphiques permettent de visualiser, d'analyser et de gérer des portefeuilles à l'aide d'écrans personnalisables afin d'afficher les informations pertinentes selon le style de gestion.

Contact

StarQube

Rue des Corps-Saints 4
1201 Genève

Contact commercial : Guillermo Albiñana Arias

Téléphone : +33 6 52 33 80 33

Mail : guillermo.albinana@starqube.com

Web : www.starqube.com

Linkedin : <https://www.linkedin.com/company/starqube>