



Alternative Data Standards

"Learning Point"

Data Lineage

Alternative Data Standards "Learning Point" Data Lineage

In a recent [FISD Alternative Data webinar](#) that discussed the risks of managing alternative data, an audience poll asked: Which of the customer's compliance requirements are the most difficult to meet? The answer provided by a resounding 90% of the audience was: "proving data lineage" was the most difficult compliance requirement to meet.

What is data lineage and why is it so difficult? Data lineage identifies the source of a data set and describes the transfers, reviews and transformations applied to the data before it arrives at the customer. Today's complex digital world of mobile applications, websites, IoT sensors and SAAS platforms, among other sources, create data content. There is a rich ecosystem of data vendors, resellers, marketplaces, product and analytics packagers licensing data to buy-side firms. In the market for financial information, then, it can be challenging to sort through the agreements to identify the origin of the data and ascertain whether a particular data vendor or a data buyer is dealing with has the authority to license the data set.

The data lineage also describes the processing steps the data undergoes. Key steps that comprise the data lineage include delivery methods, formatting, validation, standardization and loading of the data set as well as accurate logging of both when the steps occurred and responsible party/process for each.

Example questions buy-side customers may ask about the data lineage include: does the data vendor backfill data to create a point-in-time data set, are missing values interpolated, are values changed to align with the schema types, are outliers removed? Each transformation made to the data set needs to be understood and evaluated before the data buyer can make an informed purchase decision.

The [FISD Alternative Data Council](#) recently published the best practice [Due Diligence Questionnaire](#) for evaluation of the compliance and legal risks within these new alternative data sets. Included are key questions that seek to ascertain the origin of the data set.

Additional information about data set specifications for financial firms are outlined in the [FISD Data Vendor Tear Sheet](#).