# Web Data Collection Considerations:

# Interacting with Robots.txt

The Robots Exclusion Protocol, often called “robots.txt,” is a file that website operator might place in the top level directory of a website to request that cooperating web crawlers and other automated visitors (“*Bots*”) do not access all or part of the website. Despite having been created in 1994, robots.txt has not been adopted as a formal standard, nor are there established norms for how to interact with it outside of the search engine industry. Courts have articulated, under various legal theories, that robots.txt is not a binding directive, and there is currently no legally binding requirement to adhere to the preferences stated in a robots.txt file or even to check for the existence of the file on a given website. The mere posting of the file by a website operator does not create a binding contract with site visitors or any other duty towards the website operator, and there is no indication that collecting data from a website having a robots.txt file, without more, is deceptive.

In the absence of any legal requirements or generally accepted principles for non-search engine Bots (used by “*Data Collectors*”), the FISD Alternative Data Council recommends that firms consider the approach discussed below, unless circumstances merit otherwise.

**In the absence of other legal or contractual obligations, we believe Data Collectors retain discretion to choose whether to comply with robots.txt files.** This includes the choice of whether to search for the presence of a robots.txt file as an initial matter and whether to honor the request in the file. Data Collectors may decide to honor the request in certain circumstances or not at all. The presence or absence of robots.txt on a website is only one among multiple considerations in assessing claims against Data Collectors, as instructed by the examples, below:

* Data Collector Alice is not obligated to adhere to www.bob.com’s robots.txt file, which contains a particular crawl delay parameter. Nonetheless, in designing a Bot, Alice should still take reasonable measures to avoid collection activities that have a significant adverse impact on the performance of bob.com, similar to those the robots.txt sought to accomplish and as discussed in the other FISD Alternative Data Council Web Data Collection Considerations.
* If Alice’s Bot is programmed to identify robots.txt but not follow it, then Alice’s awareness of Bob’s Bot preferences could be taken into consideration when making a decision about certain legal claims. Alice’s knowledge of the robots.txt file would not likely determine, on its own, the outcome of any legal claim being made by Bob; however, Bob might point to Alice’s awareness of the Robots.txt file to argue that her conduct was willful or knowing.
* If Alice’s Bot is programmed to look for robots.txt and follow it, then this approach may be used to support an argument that Alice respected Bob’s explicit wishes with regard to web data collection.