# Web Data Collection Considerations:

# Use of Proxy Servers

The vast majority ofentities that access websites through an automated device (“*Data Collectors*”) do so via proxy servers.

Proxy servers consolidate Data Collector requests and communicate those requests to external servers on behalf of the Data Collector. One of the primary benefits proxy servers provide Data Collectors is added network security. Additional benefits proxy servers facilitate include fulfilling business needs such as communicating white-listed IP addresses to site owners and scaling commercial operations. A necessary byproduct of using proxy servers is that web traffic is routed and displayed as coming from the proxy servers’ IP addresses and therefore not the identifying IP address(es) of the Data Collectors. While there have not historically been any generally accepted principles concerning the use of proxy servers, the FISD Web Data Collection Group recommends that firms seek to follow the guidelines discussed below, unless circumstances merit otherwise.

**Whenever possible, site owners should be provided with sufficient information that would allow them to contact Data Collectors, either directly or through third parties operating at the direction of the Data Collector.** The provision of contact information may be accomplished by providing the site owner with metadata about the Data Collector itself or that of its proxy server. The following practices should be followed when providing contact information to site owners The Data Collector should seek to ensure that its proxy server relays any site owner communication to the Data Collector without undue delay.

* Similarly, a Data Collector should seek to ensure that its third-party web crawling vendor provides identifying metadata of itself or its proxy servers to site owners and inform the Data Collector of any site owner communication that it receives.
* Site owner communications received by Data Collectors, either directly or indirectly, should be addressed and resolved with the site owner (or third-party web crawling vendor) in a timely manner.

**How and where to provide contact details**. There are various ways by which the above guidelines may be met, and the preferred means may differ based upon a variety of factors, e.g., use case, crawling methodology, website configuration. For example:

* A Data Collector, proxy server or third-party vendor may identify itself in any of the following ways: company name, website URL, email address, physical address, hyperlink to a contact form, etc.
* Similarly, there are many locations where this identifying information can be presented to site owners or system administrators, e.g., USER\_AGENT, crawler header.