Knowing what XML is and how it works is key to understanding how to exploit it. XML is a format “language” sometimes used by webpages to transport data between client and server. This is usually always done via a library of platform API

XML - extensible markup language

XML entities: XML entities are a way of representing an item of data within an XML document, instead of using the data itself. Various entities are built in to the specification of the XML language. For example, the entities &lt; and &gt; represent the characters < and >. These are metacharacters used to denote XML tags, and so must generally be represented using their entities when they appear within data.

DTD - The XML document type definition (DTD) contains declarations that can define the structure of an XML document, the types of data values it can contain, and other items. The DTD is declared within the optional DOCTYPE element at the start of the XML document. The DTD can be fully self-contained within the document itself (known as an "internal DTD") or can be loaded from elsewhere (known as an "external DTD") or can be hybrid of the two.

XML Custom Entities: XML allows custom entities to be defined within the DTD. For example:

<!DOCTYPE foo [ <!ENTITY myentity "my entity value" > ]>

This definition means that any usage of the entity reference &myentity; within the XML document will be replaced with the defined value: "my entity value".

External Entities: XML external entities are a type of custom entity whose definition is located outside of the DTD where they are declared.

The declaration of an external entity uses the SYSTEM keyword and must specify a URL from which the value of the entity should be loaded. For example:

<!DOCTYPE foo [ <!ENTITY ext SYSTEM "http://normal-website.com" > ]>

The URL can use the file:// protocol, and so external entities can be loaded from file. For example:

<!DOCTYPE foo [ <!ENTITY ext SYSTEM "file:///path/to/file" > ]>

XML external entities provide the primary means by which [XML external entity attacks](https://portswigger.net/web-security/xxe) arise.