

UNISTATS output file information

The UNISTATS output file is available to download in an XML format only.

What is XML?

XML stands for eXtensible Mark-up Language and can be used to structure, store, and transport data.

XML is now the international standard for data transfer, enabling hierarchical data structures to be transferred in a single file.

What does the file look like?

When looking at an XML file you will notice that each element is contained within angle brackets – commonly referred to as ‘tags’. Where the angle brackets exist without a back-slash (<>) this denotes the opening tag of an element. The value of the element (or field) will then precede the closing bracket, followed by the closing tag which always contains a back-slash (</>).

Using the output file

It is strongly advised that the UNISTATS output file is analysed through a database. Whilst technically XML files can be imported into software such as Microsoft Excel it is ill-advised due to the size and complexity of the file. Moreover, sound data management practices also suggest that storing and/or analysing data in spread sheets is strongly inadvisable.



The KIS output file is likely to be a very large file and therefore you are advised to take this into consideration when opening, importing or analysing the data.

Understanding the file

The UNISTATS output file has a hierarchal structure formed of entities and attributes.. Entities have attributes and (optionally) entities as parts that belong to them - these are sometimes referred to as ‘Parent elements’ and ‘Child elements’ respectively. For example the entity ‘Institution’ has attributes (UKPRN, ACCOMURL, INSTBEDS etc), as well as the entity ‘KIS course’ belonging to it.

To take the example further, the ‘KIS course’ entity again has fields that belong to it, as well as a number of other entities defining different aspects of the KIS course e.g. ‘Accreditation’, ‘Course stage’ etc. When using the output file it is important to maintain the integrity of the

link between entities i.e. ensure that child elements do not become detached from the parent element.

```
<KIS>
<INSTITUTION>
<UKPRN>99999999</UKPRN>
<ACCOMURL>http://www.hesa.ac.uk</ACCOMURL>
<INSTBEDS>3000</INSTBEDS>
<INSTLOWER>3200</INSTLOWER>
<INSTUPPER>7000</INSTUPPER>
<PRIVATELOWER>2500</PRIVATELOWER>
<PRIVATEUPPER>5400</PRIVATEUPPER>
<COUNTRY>XF</COUNTRY>
<NSP>1</NSP>
<KISCOURSE>
<KISCOURSEID>BSCPRODUCT</KISCOURSEID>
<KISTYPE>1</KISTYPE>
<TITLE>BSc (Hons) Product Design Technology</TITLE>
<CRSEURL>http://www.hesa.ac.uk/productdesign</CRSEURL>
...
```

Null data within the file

When analysing data you might notice that some entities or fields exist for some KIS courses but not for others. For example, Poppleton University might have courses that have professional accreditation thus the ‘Accreditation’ entity exists for such KIS courses, however they also run courses with no accreditation and therefore the entity does not exist for those KIS courses. Another example might be where an institution does not have owned or sponsored beds and therefore the INSTBEDS and INSTLOWER/UPPER fields will not exist for them.

The value zero

Care should be taken when analysing the data to account for the distinction between nulls and zeros (‘0’). Where the data is null (i.e. not applicable such as the accreditation example) then the field or entity will often not exist in the UNISTATS output file. However, where a zero (0) is returned within a data field then this would be

representative of the true value e.g. where the PRACTICAL field contains a zero this would indicate that there is no practical assessment for the course stage.



Give further consideration to the fact that the file will contain the number zero, as well as zeros for which the meaning is defined by the data specification e.g. NSP=0 'Not participating in NSP'. Valid entries for every field can be viewed within the KIS coding manual on the HESA website.

Structure of the file

The structure of the file is defined in the **XML Schema Definition (XSD)** file. This defines the elements, their optionality and their structure. Specifically:

- The XSD defines the **minimum and maximum occurrences** for each element. If the minimum is one then the element has to be present in every case; a value of zero implies that there are cases where this element does not occur - these cases will be controlled through validation rules.
- The XSD defines the nesting of elements - in this case defining which fields belong to which entity, as well as defining the hierarchical structure of the entities.
- The XSD defines the data types. The XSD includes defining lists of valid entries, character sets for name fields, as well as defining which fields are date or numeric format.
- The XSD defines the order in which elements must appear within submitted files. This can be different to the order that they are presented within the coding manual.

The XSD is essential when developing systems that will import or interrogate the UNISTATS output file. The XSD can be viewed at:

<https://www.hesa.ac.uk/collection/c17061/download/UnistatsOutputSchema.xsd>