**Working towards building a strong Data Quality Culture**

Data is fundamental to effective, evidence-based decision-making. It underpins everything from major policy decisions to routine operational process. Often, however, our data is of unknown or questionable quality. This presents huge challenges. Poor or unknown quality data weakens evidence, undermines trust, and ultimately leads to poor outcomes. It makes organisations less efficient, and impedes effective decision-making. To make better decisions, we need better quality data.

Data quality is about fitness for purpose and is an output of better data management. Data quality requirements are defined as accuracy, completeness, currency, precision, privacy, reasonableness, integrity, timeliness, uniqueness and validity. These metrics will be core for us at St Mary’s university as we work towards developing a strong data culture.

Data maybe used by several different users and each for different purposes. Communicating the quality of data to users gives them a fuller picture of your data and its journey and mitigates against people using the data for the wrong purposes.

Data quality Principles have been defined to support organisations to create a data quality culture and Data quality Dimensions which should be used to make assessments of data quality and to identify data quality issues.

**Data Quality Principles**

|  |
| --- |
| **Commit to data quality**Data trustees must create a sense of accountability for data quality across their team(s) and make a commitment to the ongoing assessment, improvement and reporting of data quality. |
| **Know your users and their needs**Understanding user needs is essential to ensuring that data is fit for purpose. Data owners should research and understand users’ needs, prioritising efforts on the data which is most critical. |
| **Assess quality throughout the data lifecycle**Data should be managed across its lifecycle, paying close attention to quality measures and assurance at each stage. |
| **Communicate data quality clearly and effectively**Communicate quality to users regularly and clearly to ensure data is used appropriately. |
| **Anticipate changes affecting data quality**Not all future problems can be predicted. Where possible, anticipate and prevent future data quality issues through good communication, effective management of change and addressing quality issues at source. |

Dimensions of Data Quality

 **6 Dimensions of Data Quality**

|  |
| --- |
| **Accuracy**The data recorded is correct and free of errors. |
| **Completeness**The data should have all the relevant information to meet business goals and no missing information. |
| **Consistency**Everyone is collecting data in the same way and to have same understanding of what to fill in on data collection forms as part of a processes. |
| **Timeliness**Data should be available at the time it is required. |
| **Validity**Information is in a specific format and follows business rules. |
| **Uniqueness**Data is not duplicated and only one instance in which the information appears in the database. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Quality Dimension** | **What it means** | **Example of good practice** | **Example of bad practice** |
| Accuracy | The information your data contains corresponds to reality, is correct and free of errors. | Student course enrolment detail and student programme route codes match for all students enrolled within a particular academic year resulting in no errors for HESA/HESES returns and internal reports. | The programme route code held on student course enrolment and student programme route do not match for the student. This also includes mismatches in award route codes.  |
| Completeness | Does the data have all the relevant information to meet the business goals and doesn’t contain missing information. This means everyone on your team is reporting a full set of data. | Term Time Postcode is recorded accurately on student course enrolment. | There are a large number of records with the Term Time postcode missing currently on student course enrolment.  |
| Consistency | Everyone is collecting data in the same way. The goal here is for everyone who collects data to have the same understanding of what to fill in on data collection forms and as part of processes.  | Consistent end dates for students who have been withdrawn or on Leave of Absence across Course enrolment records including student Absence records.  | There is inconsistency among end dates being recorded for withdrawn and Leave of Absence students across student enrolment screens along with inconsistent student absence records making it challenging to calculate full-time equivalence of the student for the reporting year (STULOAD) for HESA. |
| Timeliness | Data should be available at the time it is required. | Closing of student records either by completing or withdrawing of all by the end of an academic year. | Records were not closed on time nor progressed on time at the end of the reporting year. |
| Validity | Information is in a specific format and follows business rules. | Start date on student course enrolment for an academic year should co-relate to standard academic year date that has been agreed for that particular year and shouldn’t be any date earlier than this agreed date. | There are currently records which the student course enrolment start date has been set to earlier than the agreed start date for the academic year. |
| Uniqueness | Data is not duplicated and only one instance in which the information appears in the database. | No duplicate student course enrolment records for the same programme and same block, same occurrence for students within the academic year should exist unless repeating the year. | There are duplicate student course enrolment records for students and that have no module records attached.  |