# Principles for Archiving Metadata

*To protect the overall standard by which the metadata is documented*

## 0. Documentation must make sense

**Principle 0** must be maintained at all times. Ensuring the documentation makes sense should be the basis for all decisions made based on the principles below. Note that this does not refer to correcting mistakes in the questionnaire, but choosing the best method in which the questionnaire is documented.

For example, when there are two options for condition text, the text which refers to the true branch is used, and the alternative is usually added as a statement. However, in cases were the second statement only contains the direction (with or without an arrow) it does not have any meaning. To use the example on the right, 'Go to Section D on page 30' would be ignored, because it is not clear which answer it refers to without the arrow, and does not make sense on its own.

> C23. Is there anyone else who regularly does these things with your daughter?
>
> Yes [ ] 1    No [ ] 2 ———→ Go to section D on page 30
>
> If yes, who ........................................................

## 1. Maintain and do not alter the semantic meaning of the questionnaire

**Principle 1** must also be maintained at all times. CLOSER intends that the metadata documented is capable of being shared with other DDI compliant organisations, hence *principle 1* ensures that CLOSER produces consistent and comparable metadata.

The practice of keeping to *principle 1* requires decisions to be made as to what questionnaire elements provide meaning. For example it was decided that bold font provided no semantic meaning and therefore CLOSER is not documenting the weight of the font. The order of multiple choice options was deemed relevant and therefore special care is taken to preserve the order within the documentation.

The three following principles were conceived purposefully to give structure and guidance as to how the first principle should be followed at all times and in all situations.

For example we can see that "Does the mother care for children at home … ?" is in bold. We do not document this, however we do document that the order of the code list is Yes, No, No known.

> **8 Does the mother care for children at home** (including adopted and foster children) **?**
>
> Yes ................................................. 1
> No .................................................. 2
> Not known ...................................... 0

## 2. Do not correct the questionnaire

**Principle 2** should only be broken when doing so maintains *principles 0* and *1*. It is fairly common to find what seem like mistakes in the questionnaire design; these can range from typos (e.g. 'martial status' instead of 'marital status') to impossible condition logic.

Any mistakes within the questionnaire could have altered the data being collected, and therefore it is important to avoid correcting the metadata. Also, what seems like a mistake always has the potential of being done purposefully.

In the case of a typo, the misspelt word can be aliased within the search engine to allow effective searching (e.g. searches looking for the 'marital', would also find aliased questions with the word 'martial').

There are rare situations where *principle 2* must be violated in order to follow *principle 1* and *principle 0*.

For example when a code value is accidently printed with the wrong value. There is no easy way to alias code values and CLOSER's documentation would suggest that a single category has two distinctly different codes. In this case, documenting the mistake would mislead the user and violate *principle 1*.

> 9. (a) Do you usually bring up any phlegm (spit from the chest) first thing in the morning in the winter?
> PHLT71    Yes ........ 1    No ........ 0
>
> (b) Do you usually bring up any phlegm (spit from the chest) during the day or at night in the winter?
> PHLD71    Yes ........ 1    No ........ 0
>
> *If 'yes' to either question 9(a) or (b)*
>
> (c) Do you bring up phlegm (spit from the chest) on most days for as much as three months each winter?
> PHLM71    Yes ........ 1    No ........ 0

## 3. Only document what is contained within the questionnaire

**Principle 3** should only be broken when doing so maintains *principles 0, 1* and *2*. There are situations where the questionnaire does not provide all of the information to document meaningfully or to generate valid DDI. However, it is important to refrain from adding additional information that is not within the questionnaire.

For example when a questionnaire asks: *3.2 'how many?'*

Often this question and similar questions can be found within a condition, immediately following a question like: *3.1 'Do you own a car?'*

Question 3.2 is largely meaningless without being able to see question 3.1, which makes it tempting to concatenate question 3.1's text to question 3.2 creating *'Do you own a car? how many?'*

Doing this violates *principle 1*. The solution to maintain context for question 3.2 is by accurately routing using conditional constructs. For example question 3.2 will be within the condition 'If yes to 3.2'

An example where the metadata must be added to in order to maintain *principle 1* is when a questionnaire uses an arrow to denote a condition. It is impossible to document an arrow literally and leaving it out of the documentation altogether changes the logical flow of the questionnaire and violates *principle 0* of making sense. Therefore, text representing the arrow's meaning has to be added.

> **11 Does this child have Special Educational Needs (SEN)?**
> TICK **ONE** BOX ONLY
>
> Yes [ ] → Go to **12**
> No [ ] → Go to **14**

## 4. Do not allow the data to inform what is recorded

**Principle 4** is the least significant principle. Whilst documenting the structure, flow and intent of a questionnaire, it may seem harmless to consult the collected data in order to better understand the questionnaire being documented. This practice, however, should also be avoided.

The aim of the ingest programme is to record the instruments used for data collection as accurately as possible. Using information that was created after the collection event can alter the perception and understanding of the instrument.

For example when a multiple choice question has an 'other (specify)' option. If the same response is often given within that option then it is relatively common to code that answer in addition to the multiple choice options (e.g. Consultant). It is important not to document this additional coded answer, because it was not offered to the respondents and therefore potentially had an effect on the collected data.

The most common situation where breaking *principle 4* is valid is when *principles 2* or *3* must be broken. For example, in the situation where it is most appropriate to correct the questionnaire it is obviously vital to check whether the mistake was intentional or not and if the mistake had a distinct effect on the collected data.

> Who undertook the antenatal care?
> Hospital medical staff ... A0193A 1
> L.H.A. M.O. ................. A0193B 2
> G.P. ............................. A0193C 3
> Midwife
>   Hospital .................... A0193D 4
>   G.P. Unit ................... A0193E 5
>   Domiciliary ............... A0193F 6
> Other (specify) ............ A0193H 7
> CONSULTANT ............ A0193G