

Managing questionnaire metadata using Archivist

Becky Oldroyd
September 2022

Scan the QR
code to visit
our Archivist
GitHub page



What is Archivist?



Open-source web
application



Multiple users
can enter and edit
questionnaires



Link questions to
variables



Map questions/variables
to topics/concepts



Exports DDI
Lifecycle 3.2



Archivist →
CLOSER
Discovery

Why was Archivist developed?

About

Search

Explore

Lists 0



Explore the content of UK
longitudinal studies



Economic
and Social
Research Council

New to CLOSER Discovery?

Search and browse questionnaires and data from the UK's leading longitudinal studies.

If you are new to CLOSER Discovery we recommend you follow the quick [tour](#).

Read more [about](#) CLOSER Discovery or take a look at the [FAQs](#) or [How-to guides](#) to get started.

Not all of the information you need may be included in Discovery yet, please see the [content](#) page for an up to date list. The studies also collect genetic data, which isn't listed within CLOSER Discovery yet. For more information please see the summary [genetics table](#).

We encourage you to help shape the site to best meet your needs by providing [feedback](#).

Variable and question search



Search by type



490
Datasets



190,021
Variables



337
Questionnaires



49,683
Questions

Explore by topic

Explore the questions and variables by topic. Topics are indexed using social science (HASSET) and medical (MeSH) terms. Find out more about the [funders](#), [data collectors](#) and study [populations](#).



COVID-19 variables and questions

Want to know more about the studies?

Overview of the studies and their sweeps:

- [Avon Longitudinal Study of Parents and Children](#)
- [1970 British Cohort Study](#)
- [Health and Employment After Fifty](#)
- [Hertfordshire Cohort Study](#)
- [Millennium Cohort Study](#)
- [1958 National Child Development Study](#)
- [Next Steps](#)
- [MRC National Survey of Health and Development](#)
- [Southampton Women's Survey](#)
- [Understanding Society](#)
- [Wirral Child Health and Development Study](#)

LIFE AT SCHOOL

B1. Have you had any changes of school since you left your Primary/Prep School?

(JB1) Tick one box

NO change (same school)
all the time) ☐

YES one change ☐

YES two changes ☐

YES three changes ☐

YES four or more changes ☐

If YES, please answer 1(a) and (b) below:

1(a) Give name and address of the last school you went to before the present one
_____ (JB1A)* _____

1(b) Give dates that you attended that school
From _____ (JB1B)* _____ till _____



LIFE AT SCHOOL ^

qc_B1

Have you had any changes of school since you left your Primary/Prep school?

Tick one box

- ☒ 1 = NO change (same school) all the time)
- ☒ 2 = YES one change
- ☒ 3 = YES two changes
- ☒ 4 = YES three changes
- ☒ 5 = YES four or more changes

Min Responses : 1 Max Responses : 1

Variables

■ JB1

c_qB1

If YES, please answer 1(a) and (b) below:

qc_B1 >= 2 && qc_B1 <= 5

^

True

^

qc_B1_a

Give name and address of the last school you went to before the present one:

Tt Generic text (255 maximum length)

qc_B1_b

Give dates that you attended that school: From ... till

📅 Generic date () Date

📅 Generic date 2 () Date

 **B1**


Have you had any changes of school since you left your Primary/Prep school?



- | | |
|---|---------------------------------------|
| 1 | NO change (same school) all the time) |
| 2 | YES one change |
| 3 | YES two changes |
| 4 | YES three changes |
| 5 | YES four or more changes |

If YES, please answer 1(a) and (b) below:




 **B1 a**

Give name and address of the last school you went to before the present one:



Generic text

 **B1 b**

Give dates that you attended that school: From ... till



Generic date

Generic date 2

What can we document in Archivist?

- Questionnaire title
- Questions
 - Question items and grids
 - Instructions
 - Interviewee
- Response domains
- Code lists
 - Cardinality
- Routing (i.e., conditions)
- Loops
- Sequences
- Statements

[Code Lists](#)

Categories

451

Code Lists

122

[ResponseDomains](#)

Datetime Answers

Numeric Answers

Text Answers

[Questions](#)

Question Items

377

Question Grids

0

Instructions

20

[Constructs](#)

[Conditions](#)

[Loops](#)

[Questions](#)

[Sequences](#)

[Statements](#)



Things For You to Do

Statement

Thank you for filling this in. Children of the 90s loves to look at the things you draw!

Question name/label

A1.

Do you ever have a headache?

Question text

yes, quite often

1

yes, sometimes

2

yes, I had one once

3

no, never

4

Code list

→ If no, go to question A3 below

Routing/Condition

View

Instruments > alspac_99_ttd

ALSPAC - Things for you to do

BUILD

MAP

alspac_99_ttd ^

s_intro_i

Things For You to Do

s_intro_ii

Thank you for filling this in. Children of the 90s loves to look at the things you draw!

qc_A1

Interviewee : Cohort/sample member

Do you ever have a headache?

☒ 1 = yes, quite often

☒ 2 = yes, sometimes

☒ 3 = yes, I had one once

☒ 4 = no, never

Min Responses : 1 Max Responses : 1

Variables

c_caf100

c_A1

If no, go to question A3 below
qc_A1 == 4

Statement

Question text

Code list

Routing/condition

Flowchart

Summary

Details

Start

alspac_99_ttd

Things For You to Do

Thank you for filling this in. Children of the 90s loves to look at the things you draw!

Statement

Question name/label

A1

Do you ever have a headache?

Question text

- | | |
|---|---------------------|
| 1 | yes, quite often |
| 2 | yes, sometimes |
| 3 | yes, I had one once |
| 4 | no, never |

Code list

Condition text

If no, go to question A3 below

Condition logic

qc_A1 == 4

Routing/Condition

Who can use Archivist?

- Simplifies questionnaire entry
 - No DDI or XML knowledge required
 - Questionnaire entry is carried out by trained but not technical staff
- Can be done remotely



Sign in

LOG IN

or [signup](#) here

Quality control features of Archivist

Quality control features of Archivist

- Re-using code lists/response domains within questionnaires

Label *

cs_Male_Female

☒ Response Domain

Min Responses

1

Max Responses

1

Codes

ID	Order	Value	Label	Actions
427400	↓	Value 1	<div>Label</div> <div>Male</div>	
427401	↑	Value 2	<div>Label</div> <div>Female</div>	

RESET

SUBMIT

DELETE

Used By

ID	Question Type	Label
207453	QuestionItem	CMSEX
207470	QuestionItem	SEX_i
207480	QuestionItem	SEX_ii

Quality control features of Archivist

- Asterisks to highlight the minimum amount of metadata required

Label *

Instruction

Literal *

Response Domains

Type and Label

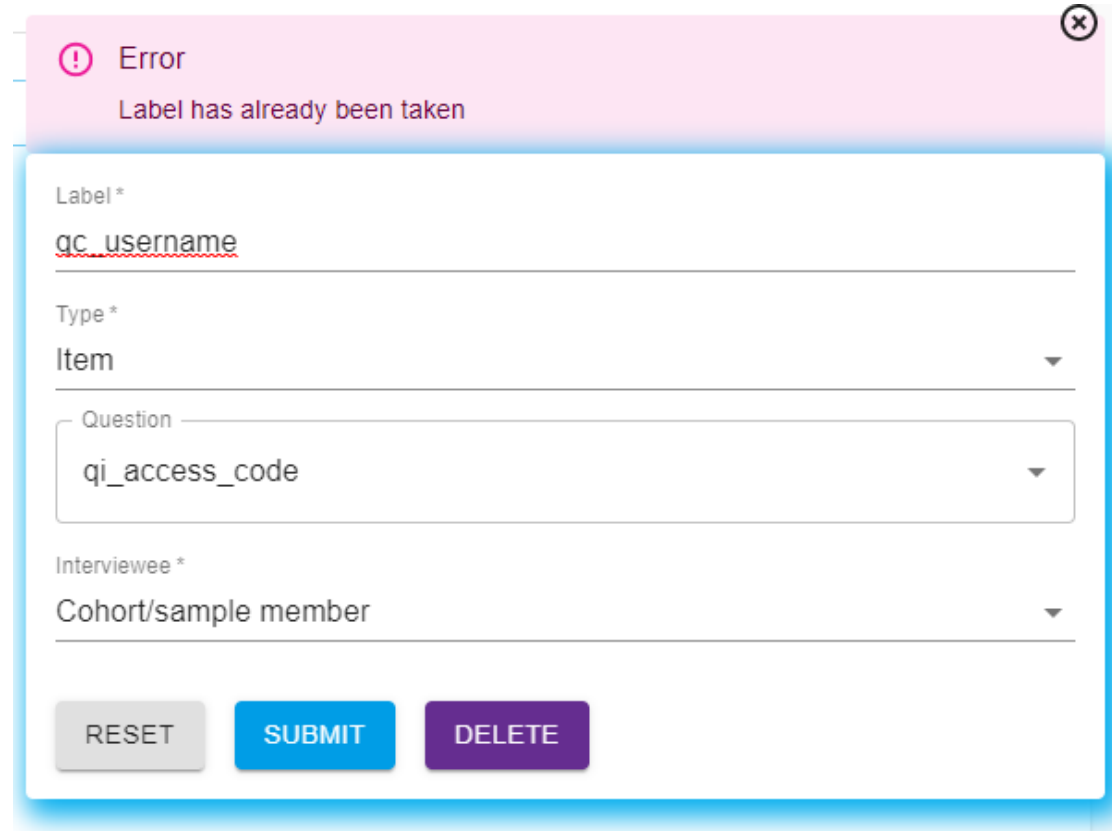
Actions

RESET

SUBMIT

Quality control features of Archivist

- Automated checks to flag errors and inconsistencies



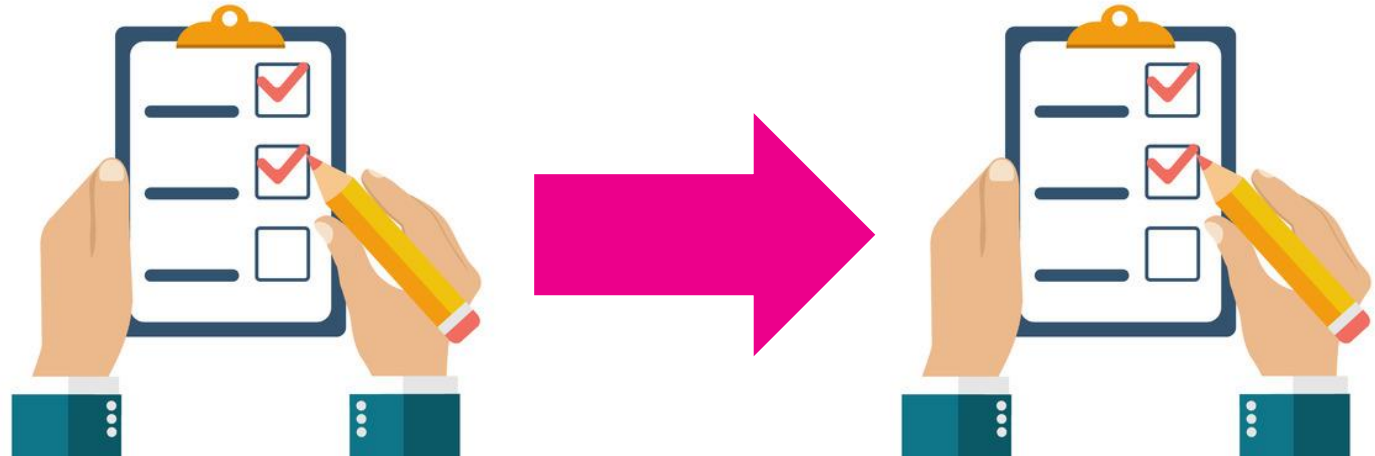
The screenshot displays a web interface for data quality control. At the top, a pink error banner with a close button (X) contains the text: "Error" and "Label has already been taken". Below this, a form is visible with the following fields:

- Label ***: A text input field containing "qc_username", which is underlined with a red dashed line.
- Type ***: A dropdown menu currently showing "Item".
- Question**: A dropdown menu currently showing "qi_access_code".
- Interviewee ***: A dropdown menu currently showing "Cohort/sample member".

At the bottom of the form, there are three buttons: "RESET" (grey), "SUBMIT" (blue), and "DELETE" (purple).

Quality control features of Archivist

- Copy completed questionnaires rather than starting from scratch



CLOSER's controlled vocabulary

Confluence

Spaces

Search

Log in

Controlled Vocabularies

Analysis Unit (Question Cons

Analysis Unit (Study)

Biological Sample

Collection Methods

Commonality Type

Interviewee (Response Unit)

Kind Of Data

Life Stage

MIMEType

Topics

Type of Material

Use Condition

Variable Source

4	104	Mental health and mental processes	Mental health = A person's condition with regard to their psychological and emotional well-being. Mental process = A series of actions or steps taken in order to achieve a particular end, relating to the mind. (Oxford Online)	4.1	10401	Mental disorders	Mental disorders comprise a broad range of problems, with different symptoms. However, are generally characterized by some combination of abnormal thoughts, emotions, behaviour and relationships with others. Examples are schizophrenia, depression, intellectual disability and disorders due to drug abuse. Most of these disorders can be successfully treated. (WHO)
				4.2	10402	Personality Temperament	Personality = behaviour -response patterns that characterize the individual. Temperament = Predisposition to react to one environment in a certain way; usually refers to mood change (MeSH)

CLOSER's controlled vocabulary

The screenshot displays the CLOSER Discovery website. On the left is a dark sidebar with navigation links: About, Search, Explore, Lists (0), and Help. The main header features the CLOSER Discovery logo and the tagline 'Explore the content of UK longitudinal studies', alongside the UKRI Economic and Social Research Council logo. A central tree of topics is highlighted with a pink rounded rectangle, listing categories such as Demographics, Housing and local environment, Physical health, Mental health and mental processes, Health care, Health behaviour, Family and social networks, Education, Employment and income, Expectations, attitudes and beliefs, Child development, Life events, Omics, Pregnancy, Administration, and COVID-19. Each item in the tree has a plus icon to its left. Above the tree are three icons: a bar chart, a question mark, and a funnel. A light blue welcome banner with a 'Tour this Page' button is positioned above the 'Browse Topics' section. The 'Browse Topics' section includes instructions on using the tree and a list of item types: Variables (the chart icon) and Questions (the ? icon). The 'Refine your results' section includes the instruction 'Filter (the funnel icon)'.

closer Discovery Explore the content of UK longitudinal studies

UKRI Economic and Social Research Council

About
Search
Explore
Lists 0
Help

+ Demographics
+ Housing and local environment
+ Physical health
+ Mental health and mental processes
+ Health care
+ Health behaviour
+ Family and social networks
+ Education
+ Employment and income
+ Expectations, attitudes and beliefs
+ Child development
+ Life events
+ Omics
+ Pregnancy
Administration
+ COVID-19

Welcome. This appears to be your first visit. Click the button below to take a quick tour.
[Tour this Page](#)



Browse Topics
Use the tree on the left hand side to choose the topic you are interested in. The (+) sign indicates there are sub topics.




Select item type

- Variables (the chart icon)
- Questions (the ? icon)


Refine your results
Filter (the funnel icon)

CLOSER's controlled vocabulary

 Maps 





16: Health

Topic  ^

INTRO_ji


I would now like to ask you about your health and the use you make of health services.


Variables 

Topic 

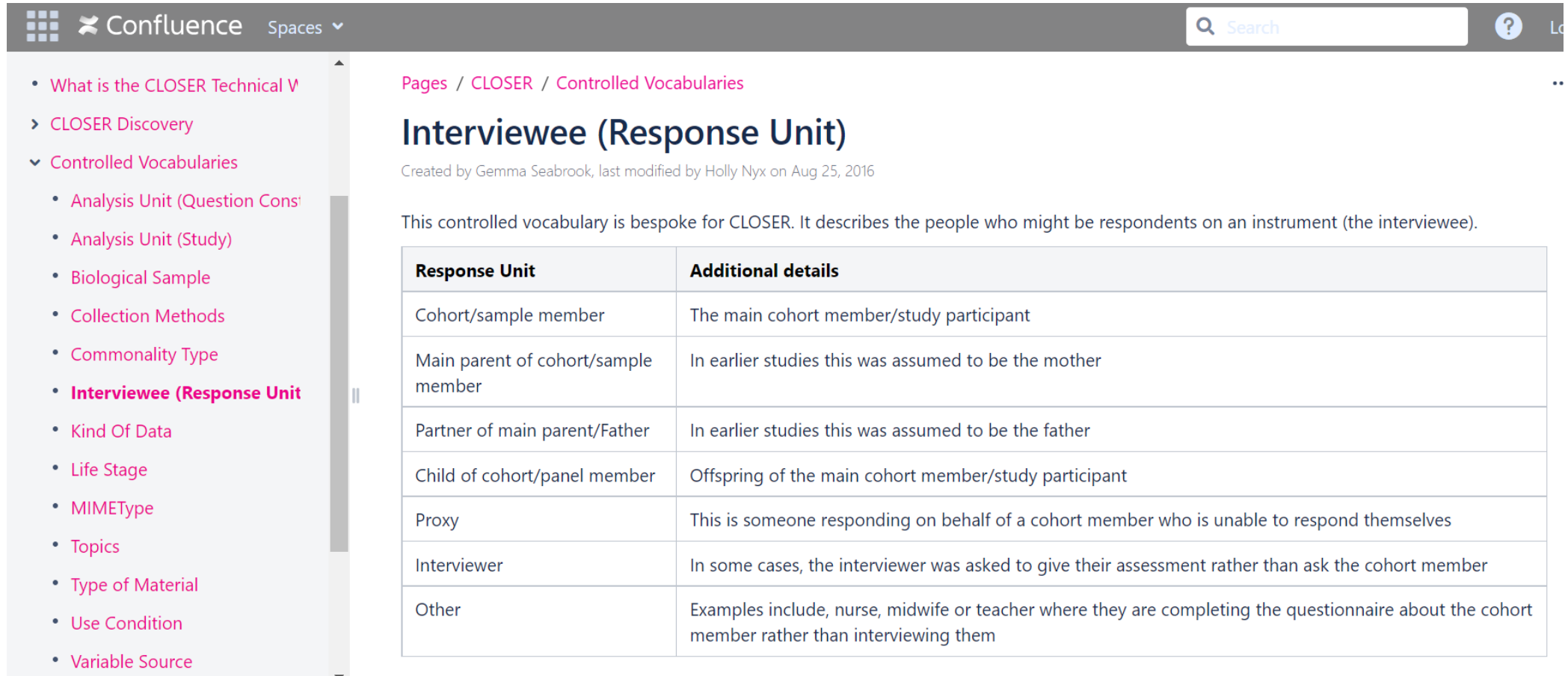
KHLSTAT

Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been... READ OUT...

Variables 

Topic 

CLOSER's controlled vocabulary



The screenshot shows a Confluence page with a sidebar on the left containing a list of links. The main content area has a breadcrumb trail 'Pages / CLOSER / Controlled Vocabularies', a title 'Interviewee (Response Unit)', and a description. Below the description is a table with two columns: 'Response Unit' and 'Additional details'.

Confluence Spaces

Search

Pages / CLOSER / Controlled Vocabularies

Interviewee (Response Unit)

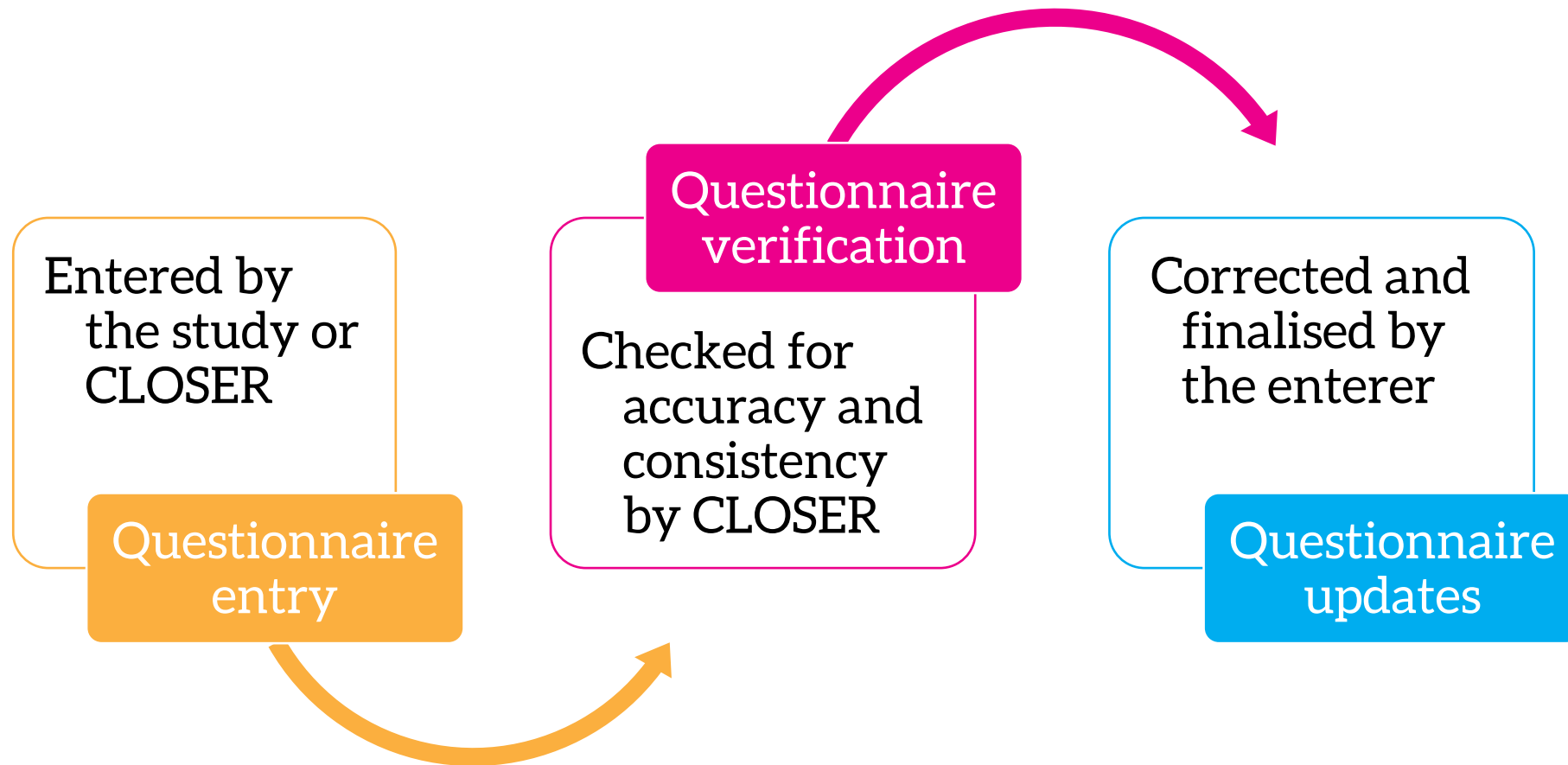
Created by Gemma Seabrook, last modified by Holly Nyx on Aug 25, 2016

This controlled vocabulary is bespoke for CLOSER. It describes the people who might be respondents on an instrument (the interviewee).

Response Unit	Additional details
Cohort/sample member	The main cohort member/study participant
Main parent of cohort/sample member	In earlier studies this was assumed to be the mother
Partner of main parent/Father	In earlier studies this was assumed to be the father
Child of cohort/panel member	Offspring of the main cohort member/study participant
Proxy	This is someone responding on behalf of a cohort member who is unable to respond themselves
Interviewer	In some cases, the interviewer was asked to give their assessment rather than ask the cohort member
Other	Examples include, nurse, midwife or teacher where they are completing the questionnaire about the cohort member rather than interviewing them

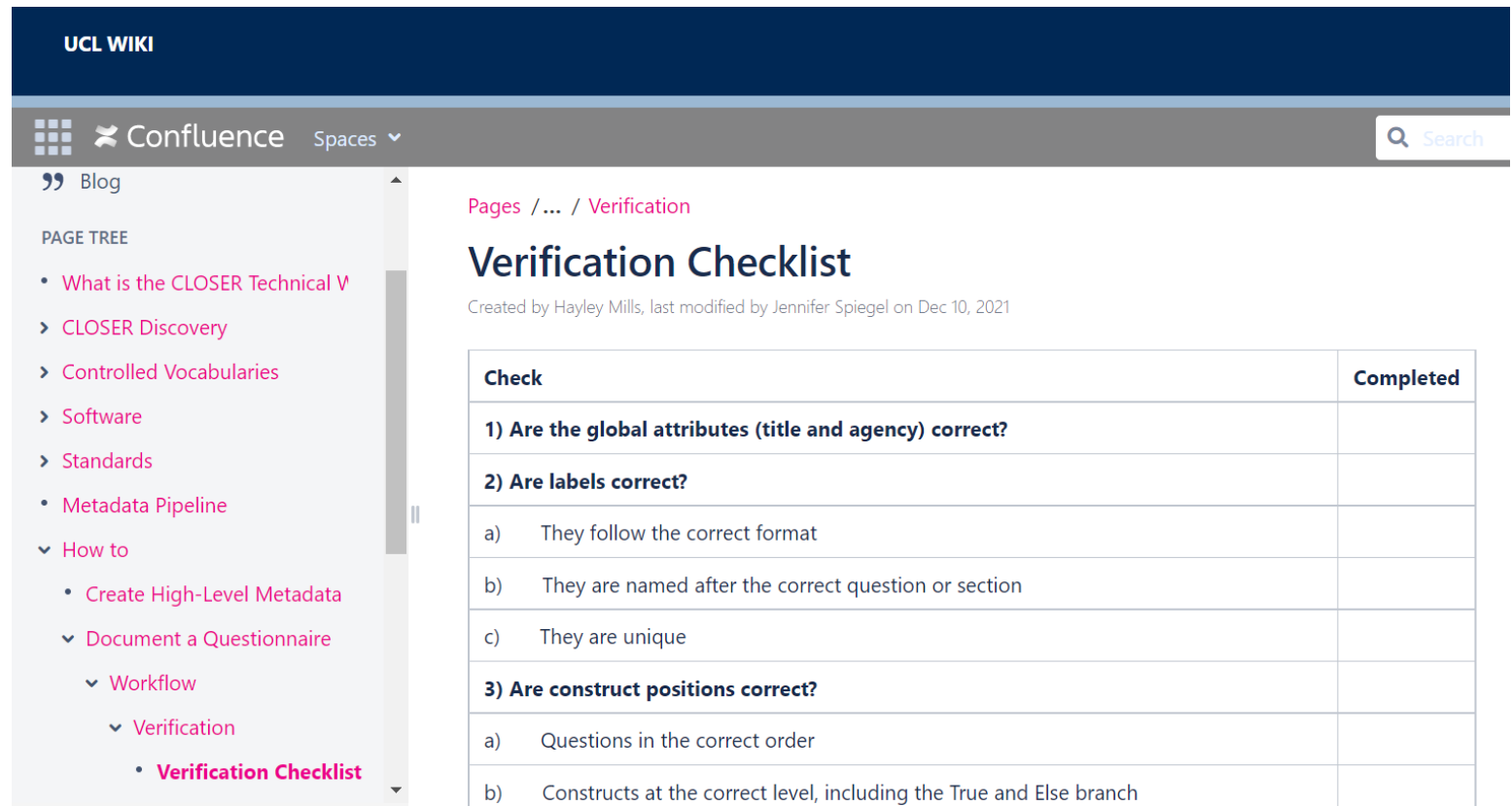
- What is the CLOSER Technical V
- CLOSER Discovery
- Controlled Vocabularies
 - Analysis Unit (Question Cons
 - Analysis Unit (Study)
 - Biological Sample
 - Collection Methods
 - Commonality Type
 - Interviewee (Response Unit)**
 - Kind Of Data
 - Life Stage
 - MIMETYPE
 - Topics
 - Type of Material
 - Use Condition
 - Variable Source

Quality control throughout the metadata workflow



Questionnaire verification

- Stringent verification process
 - Checklist
 - Summary tables
 - Listening

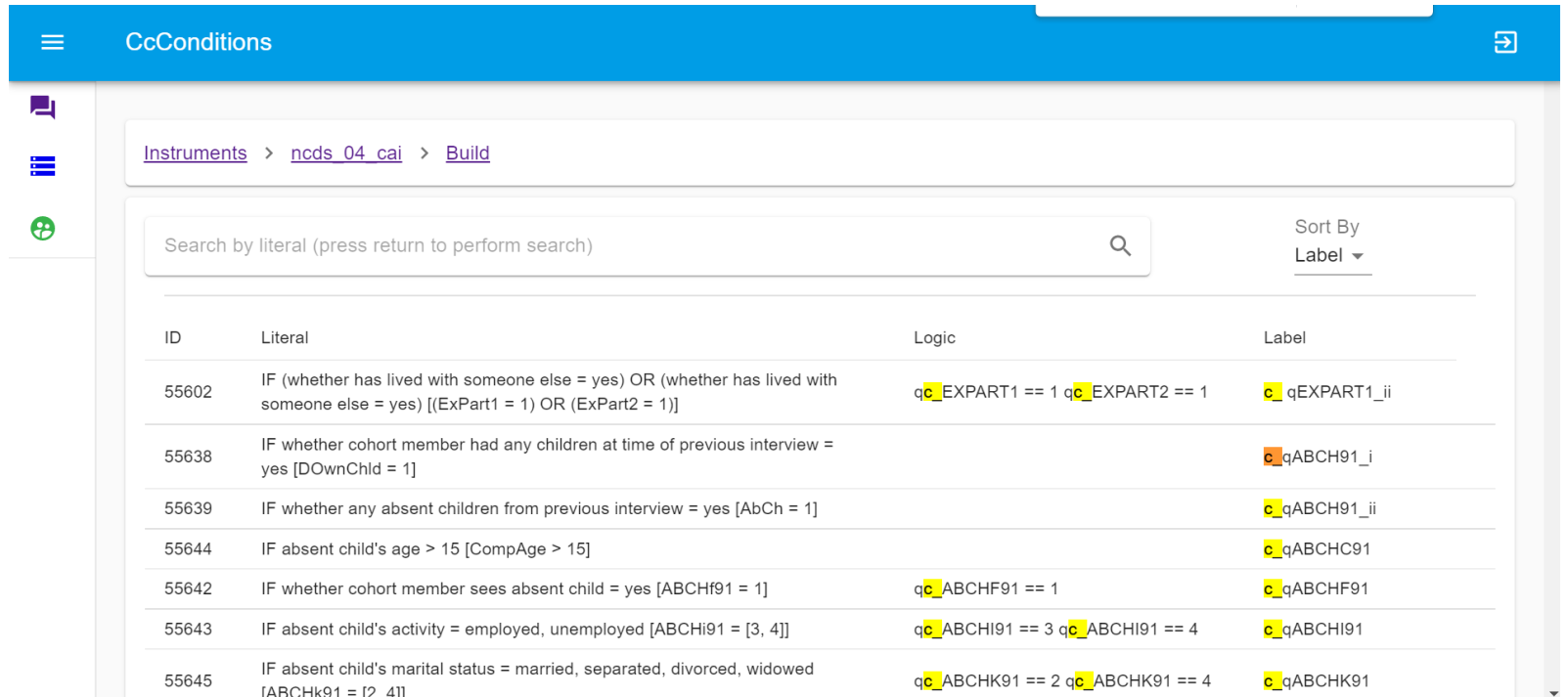


The screenshot shows a Confluence page titled "Verification Checklist" under the "UCL WIKI" header. The left sidebar contains a "PAGE TREE" with a hierarchy: "What is the CLOSER Technical V", "CLOSER Discovery", "Controlled Vocabularies", "Software", "Standards", "Metadata Pipeline", "How to" (expanded), "Create High-Level Metadata", "Document a Questionnaire" (expanded), "Workflow", "Verification" (expanded), and "Verification Checklist" (selected). The main content area shows the page title "Verification Checklist", its creation/modification history, and a table with two columns: "Check" and "Completed".

Check	Completed
1) Are the global attributes (title and agency) correct?	
2) Are labels correct?	
a) They follow the correct format	
b) They are named after the correct question or section	
c) They are unique	
3) Are construct positions correct?	
a) Questions in the correct order	
b) Constructs at the correct level, including the True and Else branch	

Questionnaire verification

- Stringent verification process
 - Checklist
 - **Summary tables**
 - Listening

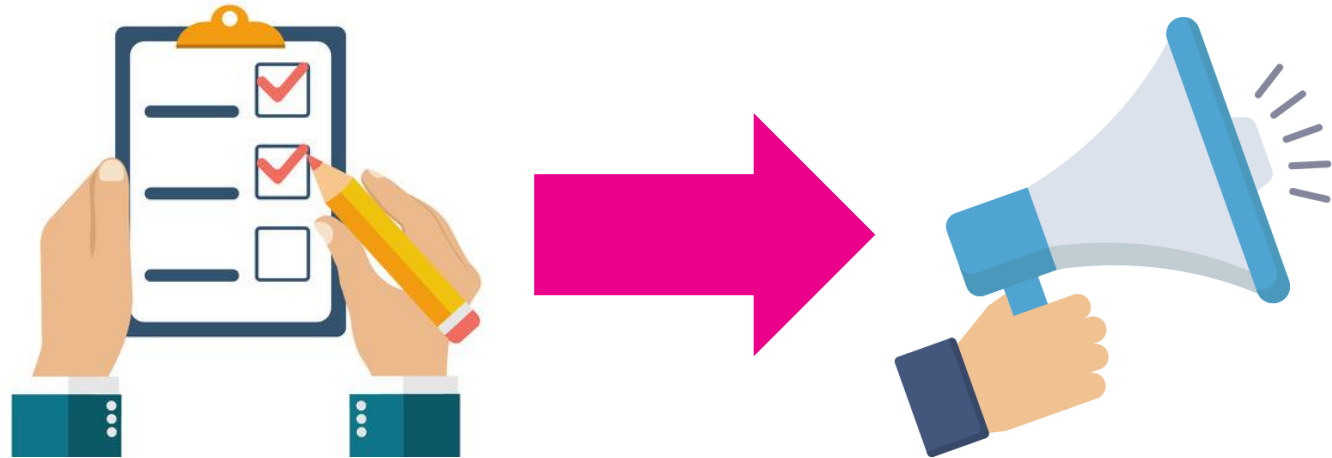


The screenshot displays the 'CcConditions' application interface. At the top, a blue header bar contains a menu icon, the text 'CcConditions', and a share icon. Below the header, a breadcrumb trail reads 'Instruments > ncds_04_cai > Build'. A search bar with the placeholder 'Search by literal (press return to perform search)' and a magnifying glass icon is positioned on the right. To the right of the search bar is a 'Sort By' dropdown menu currently set to 'Label'. The main content area features a table with four columns: 'ID', 'Literal', 'Logic', and 'Label'. The table lists seven conditions, each with a unique ID, a descriptive literal, a logical expression, and a corresponding label. The labels are color-coded: yellow for 'q' and orange for 'c'.

ID	Literal	Logic	Label
55602	IF (whether has lived with someone else = yes) OR (whether has lived with someone else = yes) [(ExPart1 = 1) OR (ExPart2 = 1)]	qc_EXPART1 == 1 qc_EXPART2 == 1	c_qEXPART1_ii
55638	IF whether cohort member had any children at time of previous interview = yes [DownChld = 1]		c_qABCH91_i
55639	IF whether any absent children from previous interview = yes [AbCh = 1]		c_qABCH91_ii
55644	IF absent child's age > 15 [CompAge > 15]		c_qABCHC91
55642	IF whether cohort member sees absent child = yes [ABCHF91 = 1]	qc_ABCHF91 == 1	c_qABCHF91
55643	IF absent child's activity = employed, unemployed [ABCHI91 = [3, 4]]	qc_ABCHI91 == 3 qc_ABCHI91 == 4	c_qABCHI91
55645	IF absent child's marital status = married, separated, divorced, widowed [ABCHK91 = [2, 4]]	qc_ABCHK91 == 2 qc_ABCHK91 == 4	c_qABCHK91

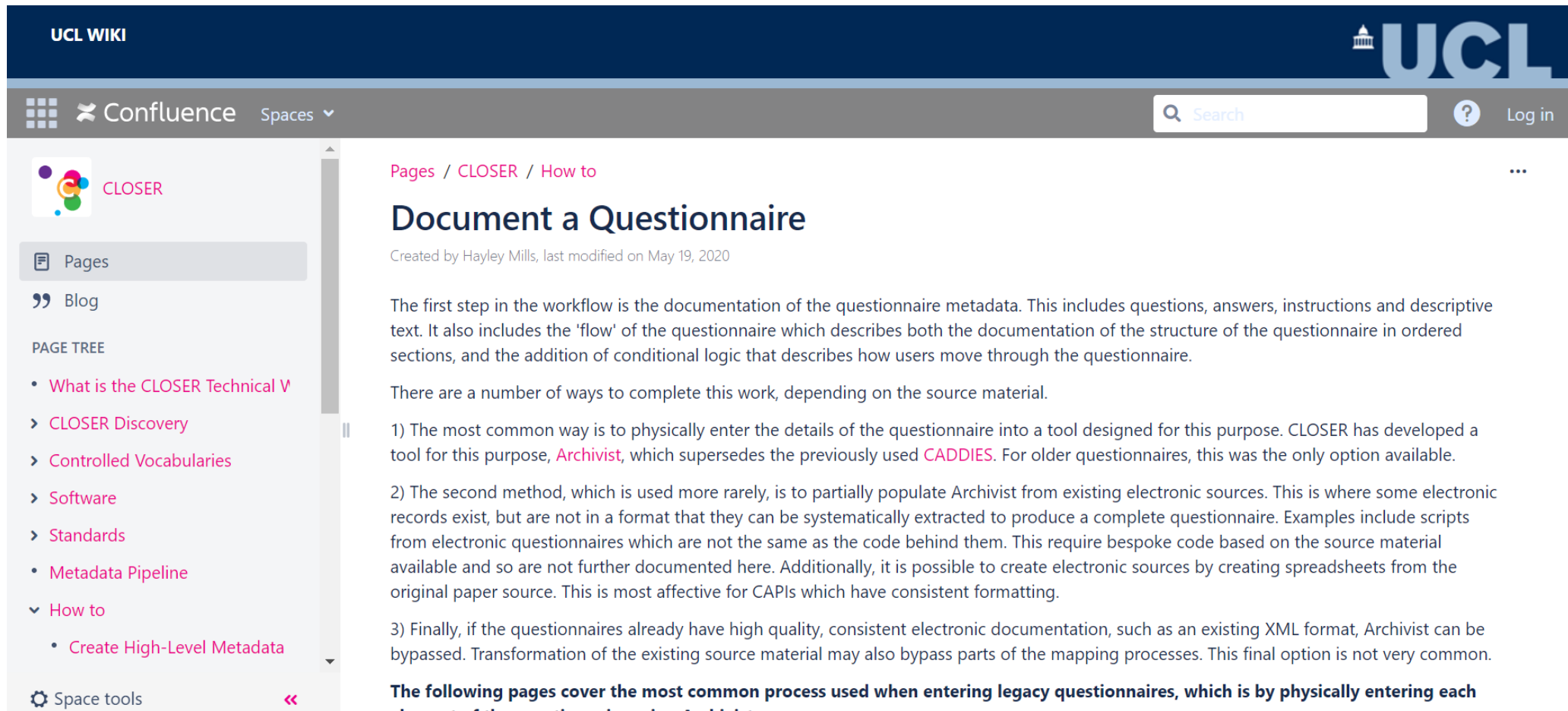
Questionnaire verification

- Stringent verification process
 - Checklist
 - Summary tables
 - **Listening**



CLOSER Technical Wiki

<https://wiki.ucl.ac.uk/display/CLOS/CLOSER>



The screenshot shows the CLOSER Technical Wiki interface. The top navigation bar includes 'UCL WIKI' and the UCL logo. Below this is a Confluence header with a search bar and a 'Log in' link. The left sidebar contains the 'CLOSER' logo and a 'PAGE TREE' with links to 'What is the CLOSER Technical V', 'CLOSER Discovery', 'Controlled Vocabularies', 'Software', 'Standards', 'Metadata Pipeline', and 'How to' (which is expanded to show 'Create High-Level Metadata'). The main content area displays the breadcrumb 'Pages / CLOSER / How to' and the title 'Document a Questionnaire', created by Hayley Mills on May 19, 2020. The text describes the first step in the workflow: documenting questionnaire metadata, including questions, answers, instructions, and descriptive text. It also mentions the 'flow' of the questionnaire and conditional logic. A list of three methods for completing this work is provided: 1) Physically entering details into a tool like Archivist, 2) Partially populating Archivist from existing electronic sources, and 3) Bypassing the process if high-quality electronic documentation is already available. The page concludes by stating that the following pages cover the most common process for entering legacy questionnaires.

UCL WIKI

Confluence Spaces

Search

Log in

CLOSER

Pages

Blog

PAGE TREE

- What is the CLOSER Technical V
- CLOSER Discovery
- Controlled Vocabularies
- Software
- Standards
- Metadata Pipeline
- How to
 - Create High-Level Metadata

Space tools

Pages / CLOSER / How to

Document a Questionnaire

Created by Hayley Mills, last modified on May 19, 2020

The first step in the workflow is the documentation of the questionnaire metadata. This includes questions, answers, instructions and descriptive text. It also includes the 'flow' of the questionnaire which describes both the documentation of the structure of the questionnaire in ordered sections, and the addition of conditional logic that describes how users move through the questionnaire.

There are a number of ways to complete this work, depending on the source material.

- 1) The most common way is to physically enter the details of the questionnaire into a tool designed for this purpose. CLOSER has developed a tool for this purpose, [Archivist](#), which supersedes the previously used [CADDIES](#). For older questionnaires, this was the only option available.
- 2) The second method, which is used more rarely, is to partially populate Archivist from existing electronic sources. This is where some electronic records exist, but are not in a format that they can be systematically extracted to produce a complete questionnaire. Examples include scripts from electronic questionnaires which are not the same as the code behind them. This requires bespoke code based on the source material available and so are not further documented here. Additionally, it is possible to create electronic sources by creating spreadsheets from the original paper source. This is most effective for CAPIs which have consistent formatting.
- 3) Finally, if the questionnaires already have high quality, consistent electronic documentation, such as an existing XML format, Archivist can be bypassed. Transformation of the existing source material may also bypass parts of the mapping processes. This final option is not very common.

The following pages cover the most common process used when entering legacy questionnaires, which is by physically entering each element of the questionnaire using Archivist

Take home points



Re-use metadata
where possible



Automated checks to
highlight errors



Integrate quality
control into your
metadata workflow



Document your
metadata entry
process



Thank you 😊

Scan the QR
code to visit
our Archivist
GitHub page

