

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



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:

 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	<input type="checkbox"/>
yes, sometimes	<input type="checkbox"/>
yes, I had one once	<input type="checkbox"/>
no, never	<input type="checkbox"/>

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

- o ccaf100

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

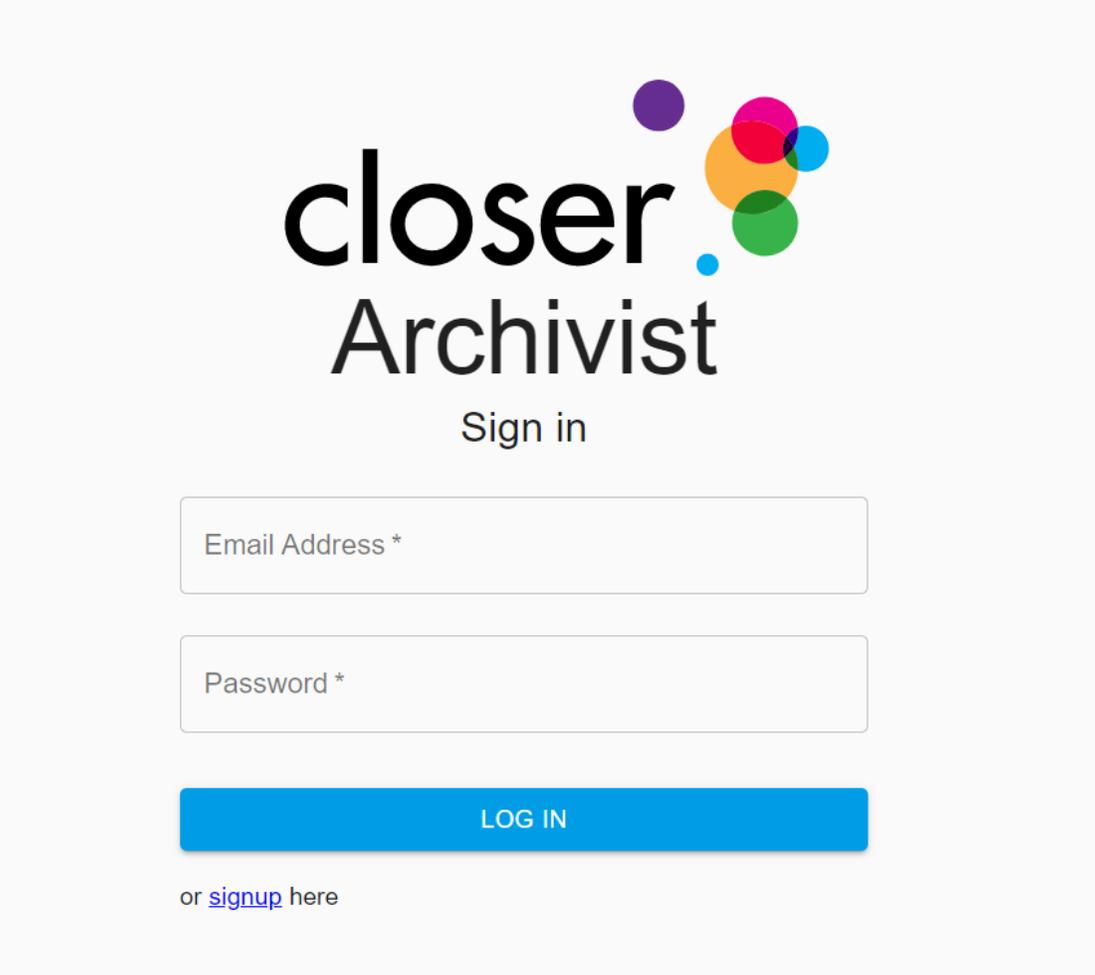
Condition logic

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



The screenshot shows the login interface for 'closer Archivist'. At the top, the logo features the word 'closer' in a bold, black, sans-serif font, followed by a cluster of five overlapping circles in purple, pink, orange, green, and blue. Below the logo, the word 'Archivist' is written in a smaller, black, sans-serif font. Underneath that, the text 'Sign in' is centered. There are two input fields: the first is labeled 'Email Address *' and the second is labeled 'Password *'. Below these fields is a prominent blue button with the text 'LOG IN' in white, uppercase letters. At the bottom of the form, there is a link that says '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	<input type="text" value="Male"/>	
427401	↑	Value 2	<input type="text" value="Female"/>	

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 *

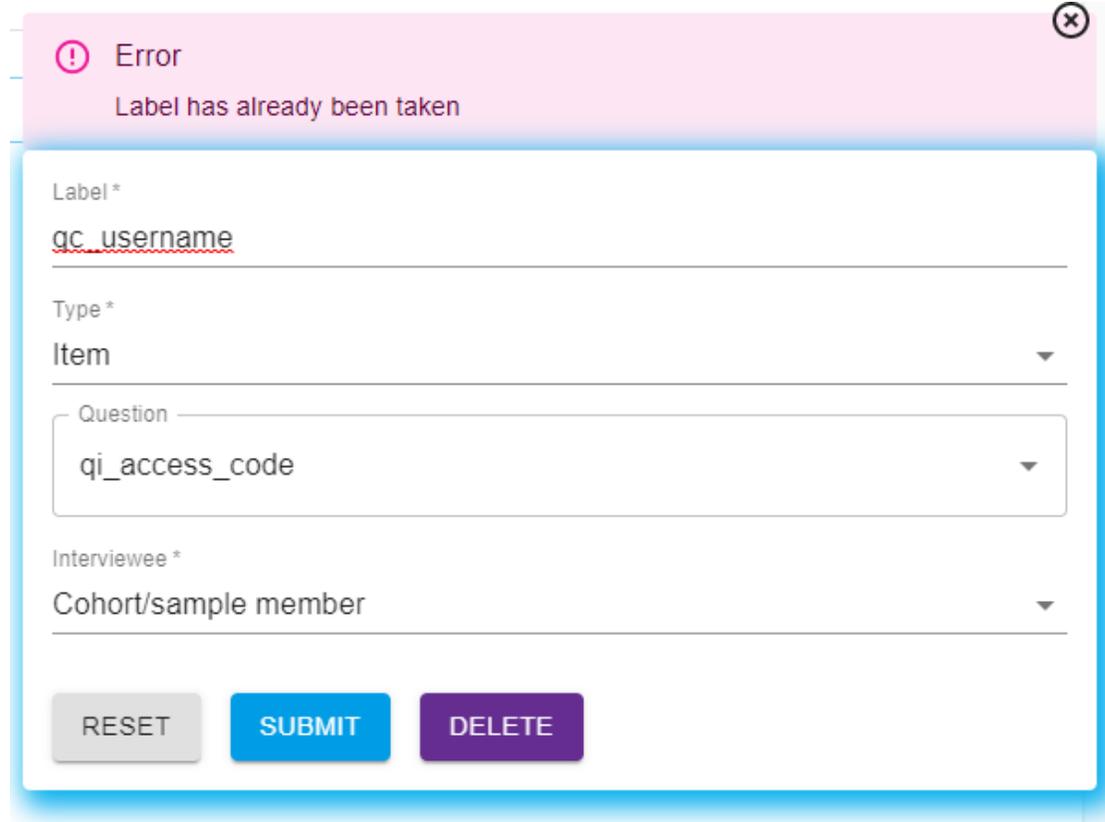
Response Domains [⊕]

Type and Label	Actions
----------------	---------

RESET SUBMIT

Quality control features of Archivist

- Automated checks to flag errors and inconsistencies



The screenshot shows a web form with a pink error banner at the top. The banner contains a red exclamation mark icon, the word "Error", and the message "Label has already been taken". Below the banner, the form has several fields: "Label*" with the value "qc_username" (underlined in red), "Type*" with a dropdown menu set to "Item", "Question" with a dropdown menu set to "qi_access_code", and "Interviewee*" with a dropdown menu set to "Cohort/sample member". At the bottom of the form are three buttons: "RESET" (grey), "SUBMIT" (blue), and "DELETE" (purple).

Error

Label has already been taken

Label*

qc_username

Type*

Item

Question

qi_access_code

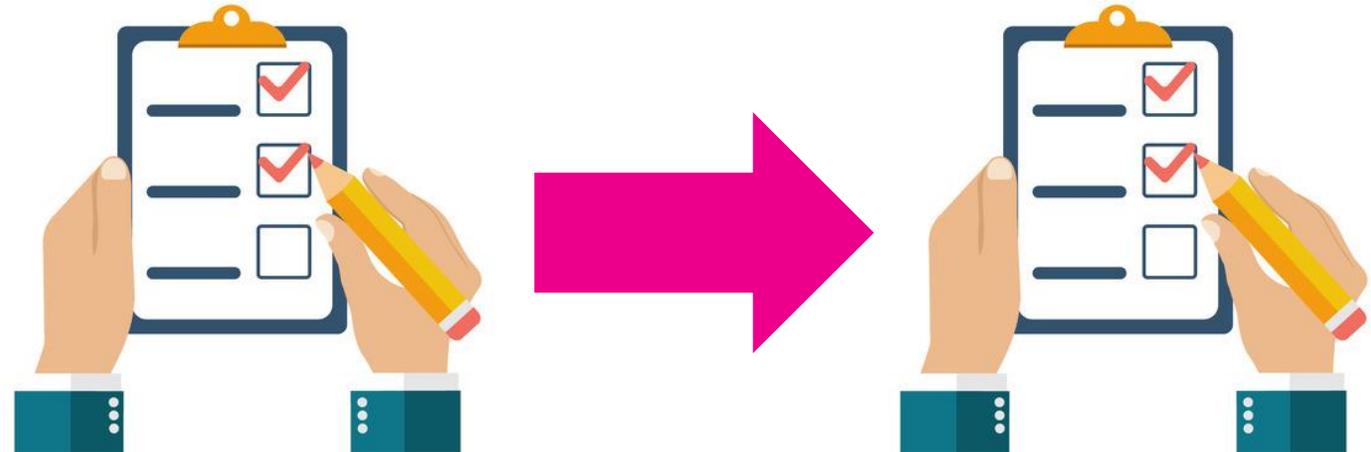
Interviewee*

Cohort/sample member

RESET SUBMIT DELETE

Quality control features of Archivist

- Copy completed questionnaires rather than starting from scratch



CLOSER's controlled vocabulary

Confluence Spaces		Search		Log in				
Controlled Vocabularies <ul style="list-style-type: none">Analysis Unit (Question Cons)Analysis Unit (Study)Biological SampleCollection MethodsCommonality TypeInterviewee (Response Unit)Kind Of DataLife StageMIMETypeTopicsType of MaterialUse ConditionVariable 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 changes (MeSH)

CLOSER's controlled vocabulary

The screenshot displays the CLOSER Discovery website interface. On the left is a dark sidebar with navigation links: About, Search, Explore, Lists (0), and Help (with a dropdown arrow). The main header features the CLOSER Discovery logo (with a colorful dot graphic) and the tagline "Explore the content of UK longitudinal studies", alongside the UKRI Economic and Social Research Council logo.

The central content area is divided into two main sections. On the left, a controlled vocabulary tree is shown, enclosed in a pink rounded rectangle. The tree lists various topics, each with a plus sign icon indicating sub-topics: Demographics, Housing and local environment, Physical health, Mental health and mental proci, Health care, Health behaviour, Family and social networks, Education, Employment and income, Expectations, attitudes and bel, Child development, Life events, Omics, Pregnancy, Administration, and COVID-19.

On the right, there are search and filter controls. At the top, there are icons for a bar chart (Variables), a question mark (Questions), and a funnel (Filter). Below these is a light blue notification box with the text "Welcome. This appears to be your first visit. Click the button below to take a quick tour." and a blue "Tour this Page" button. Further down, the "Browse Topics" section explains that the tree on the left is used to choose topics, with the (+) sign indicating sub-topics. The "Select item type" section lists "Variables (the chart icon)" and "Questions (the ? icon)". The "Refine your results" section lists "Filter (the funnel icon)".

CLOSER's controlled vocabulary

The screenshot displays the CLOSER interface for a 'Maps' section. The top navigation bar is blue with a hamburger menu icon on the left, the text 'Maps' in the center, and a share icon on the right. Below the navigation bar, the main content area is titled '16: Health' and features a 'Topic' dropdown menu. Two question cards are visible, each with a blue header button, a question text, and two dropdown menus for 'Variables' and 'Topic'.

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 dark header bar containing the Confluence logo, 'Spaces' dropdown, a search bar, and a help icon. A left sidebar lists navigation items, with 'Interviewee (Response Unit)' highlighted. The main content area shows the page title, creation/modification info, a descriptive paragraph, and a table of response units.

Confluence Spaces

Search

Pages / CLOSER / Controlled Vocabularies

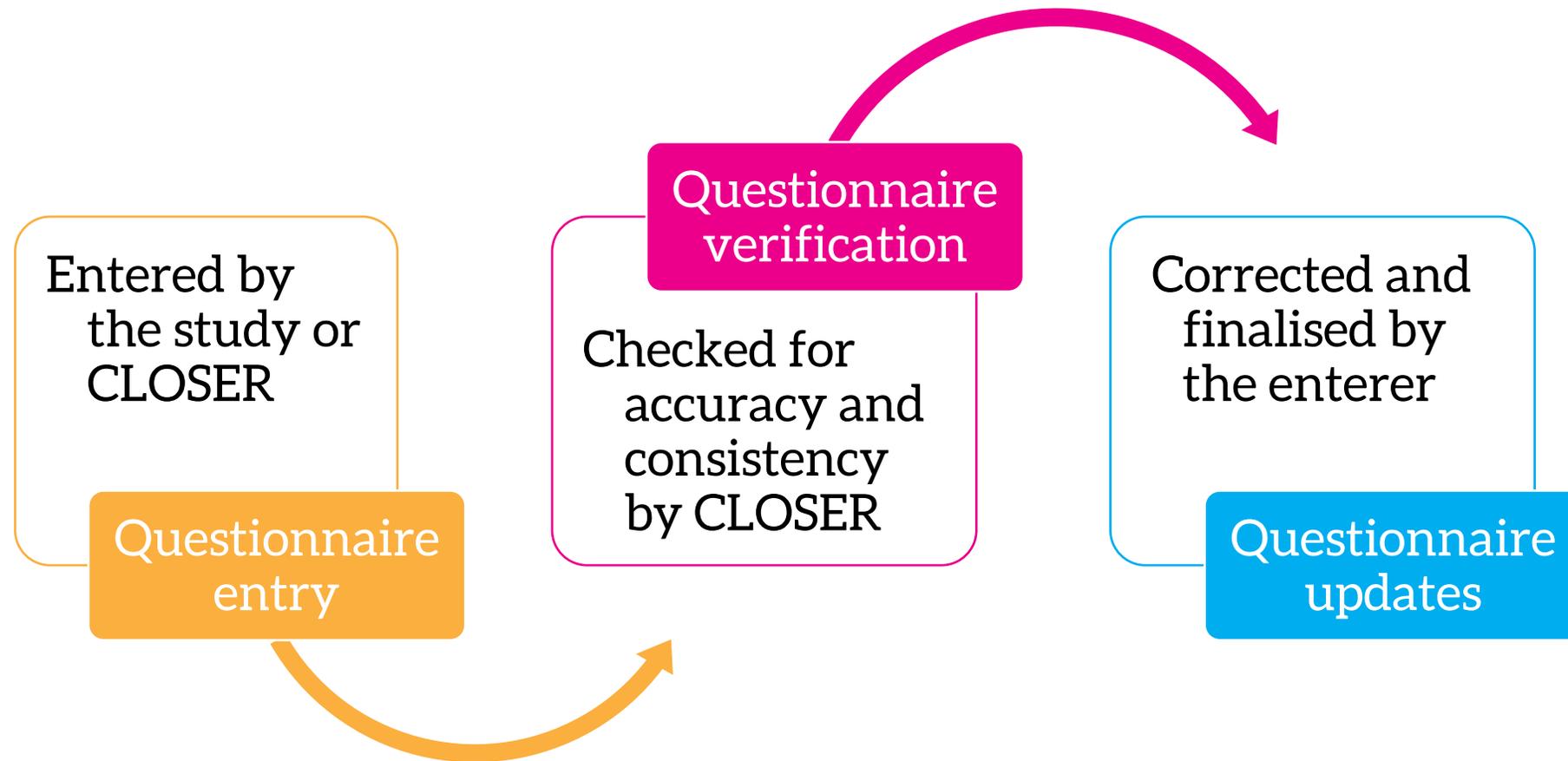
Interviewee (Response Unit)

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

Quality control throughout the metadata workflow



Questionnaire verification

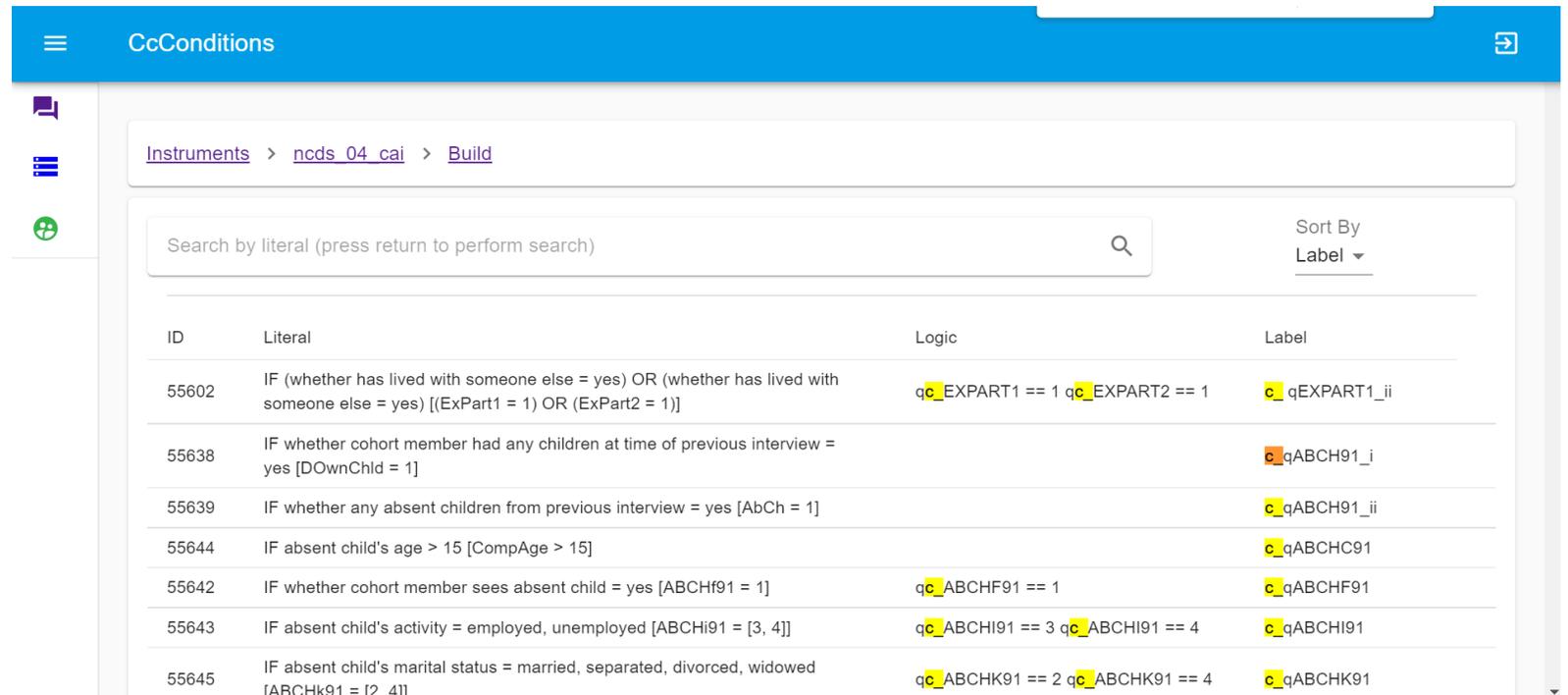
- Stringent verification process
 - Checklist
 - Summary tables
 - Listening

The screenshot shows a Confluence page titled "Verification Checklist" under the "UCL WIKI" header. The page is part of a "Verification" space. The left sidebar contains a "PAGE TREE" with the following items: "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" (expanded), "Verification" (expanded), and "Verification Checklist" (selected). The main content area shows the page title "Verification Checklist", the author "Created by Hayley Mills, last modified by Jennifer Spiegel on Dec 10, 2021", 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

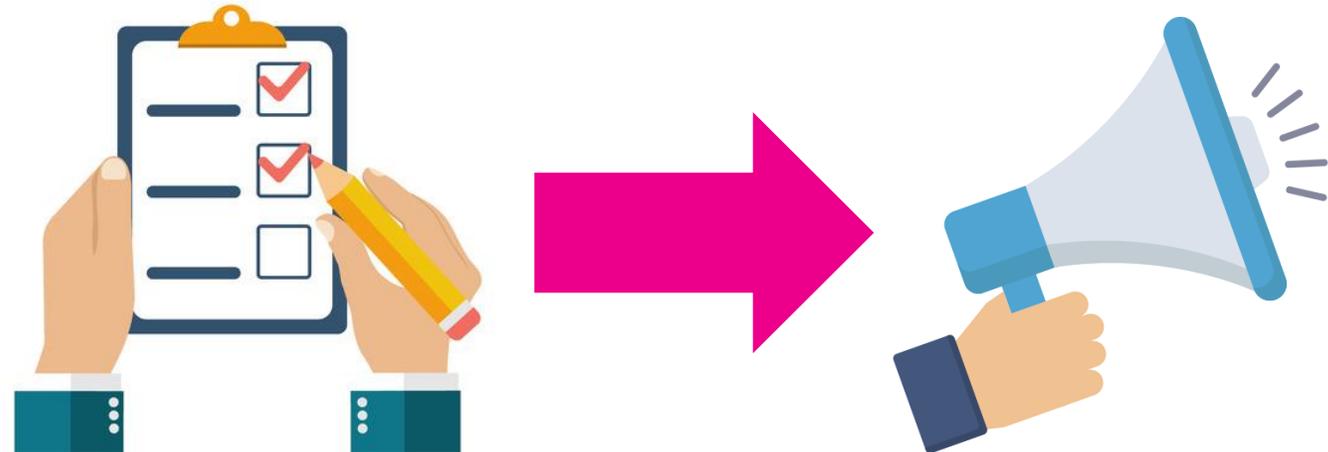


The screenshot shows the 'CcConditions' web application interface. The breadcrumb path is 'Instruments > ncds_04_cai > Build'. There is a search bar with the placeholder text 'Search by literal (press return to perform search)' and a 'Sort By Label' dropdown menu. The main content is a table with the following data:

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 page. The header includes 'UCL WIKI' and the UCL logo. Below the header is a navigation bar with 'Confluence Spaces', a search bar, and a 'Log in' button. The left sidebar contains the CLOSER logo, a 'Pages' section, a 'Blog' section, and a 'PAGE TREE' section. The 'PAGE TREE' section lists several pages, including 'What is the CLOSER Technical V', 'CLOSER Discovery', 'Controlled Vocabularies', 'Software', 'Standards', 'Metadata Pipeline', and 'How to'. The 'How to' section is expanded to show 'Create High-Level Metadata'. The main content area displays the page title 'Document a Questionnaire', the author 'Hayley Mills', and the creation date 'May 19, 2020'. The page content describes the first step in the workflow: documenting questionnaire metadata. It includes a paragraph explaining the process and a list of three methods to complete the work. The first method is physically entering details into a tool called 'Archivist'. The second method is partially populating 'Archivist' from existing electronic sources. The third method is bypassing 'Archivist' if the questionnaires already have high quality, consistent electronic documentation.

UCL WIKI 

Confluence Spaces ? 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 require 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 affective 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

