

Metadata in a Thousand Files: Bringing it all Together

Claude Gierl
Jon Johnson

Metadata in a Thousand Files: Bringing it all Together

Where we are at

The repository loader

The input files and the benefits of working with DDI-L

Some other lessons learned

The CLOSER Project

www.closer.ac.uk

Aims to maximise use of the UK longitudinal studies

Eight participating studies

100 000s variables

CLOSER Search Platform: discovery.closer.ac.uk

The CLOSER Search Platform Development

Capture
Metadata

Load
Repository

Develop
Web Portal

The Repository Loader

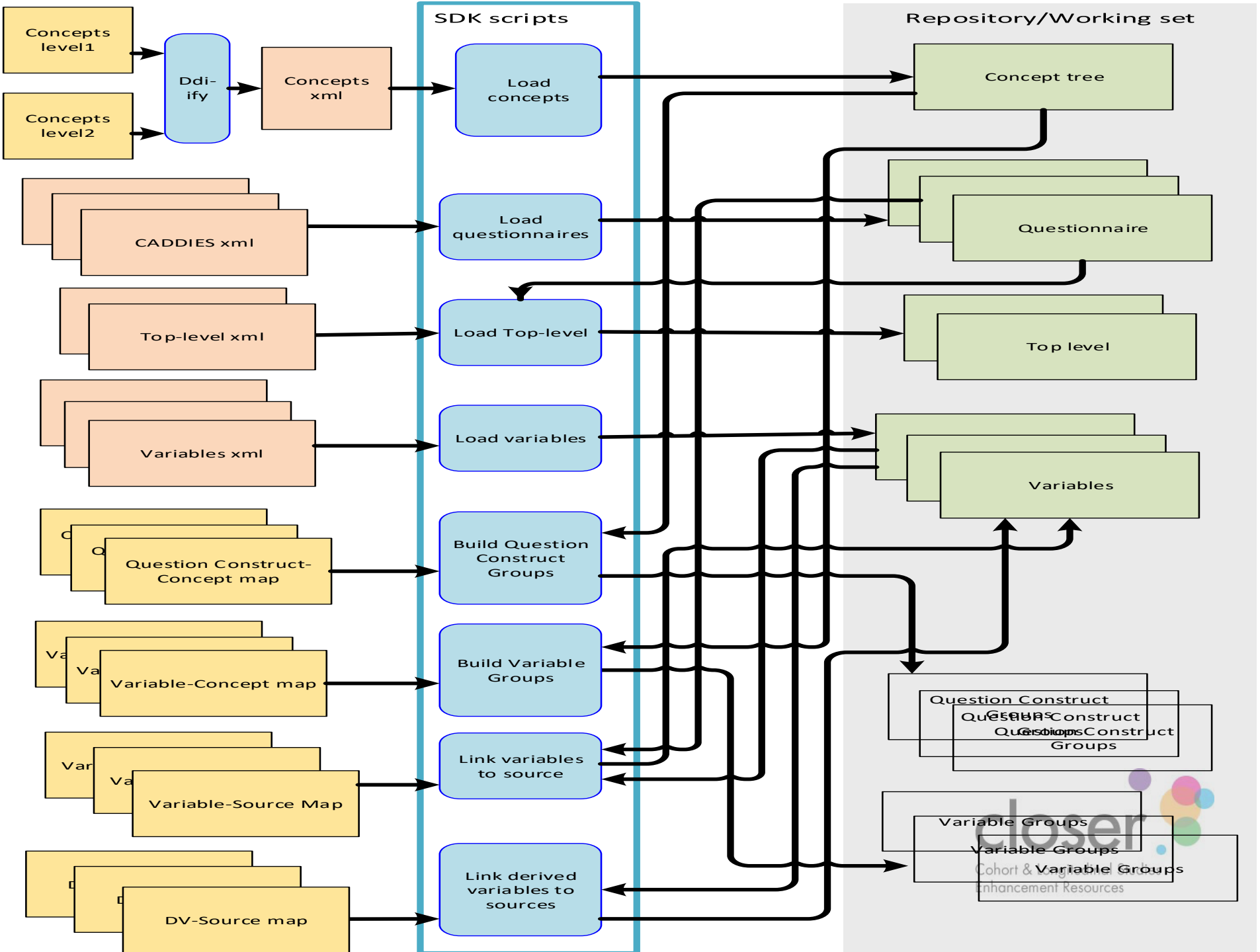
Loads DDI-L(3.2) input files and integrates the components
Creates links from the mapping files

Written in C#

Colectica SDK library

Microsoft SQL Server

In development phase



Loader Input

Concepts

Questionnaire Metadata

Variables Metadata

Top Level (Study) Metadata

Mapping and Linking Files

Concepts

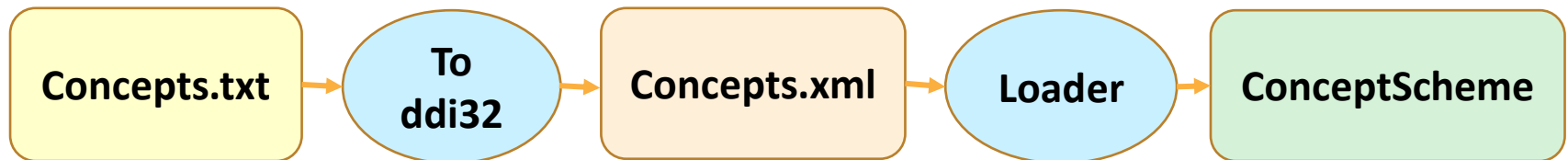
Used in the portal to provide one of the routes to variable discovery

Concepts as broad topics

Concept hierarchy

Level 1: e.g. *Education*

Level 2: e.g. *Schooling, Qualifications*

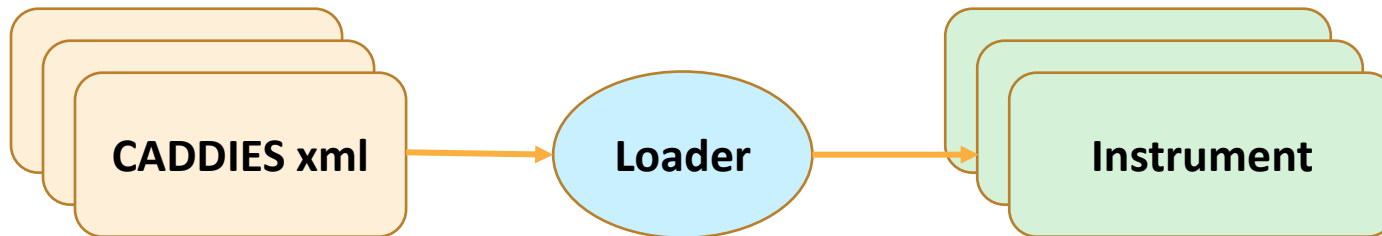


Questionnaires

Metadata team enters questionnaire into CADDIES

A few hundred questionnaires or instruments

Output is DDI-L(3.2) files, one per questionnaire

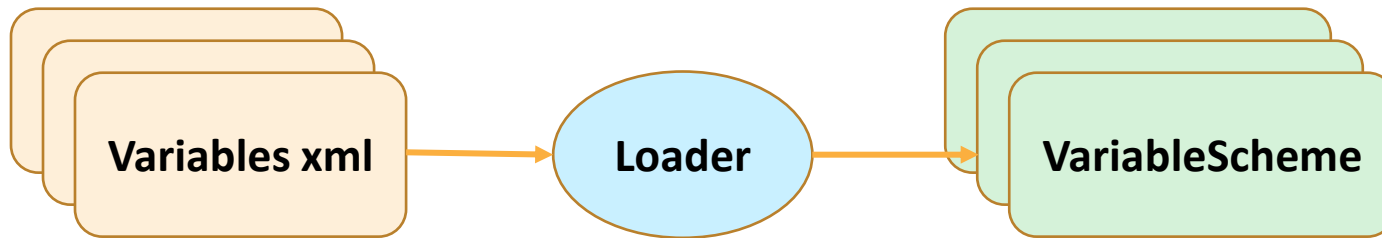


Variables

Parsed from the SPSS datafiles

SledgeHammer and Python scripts

Output is DDI-L(3.2) files, one per datafile

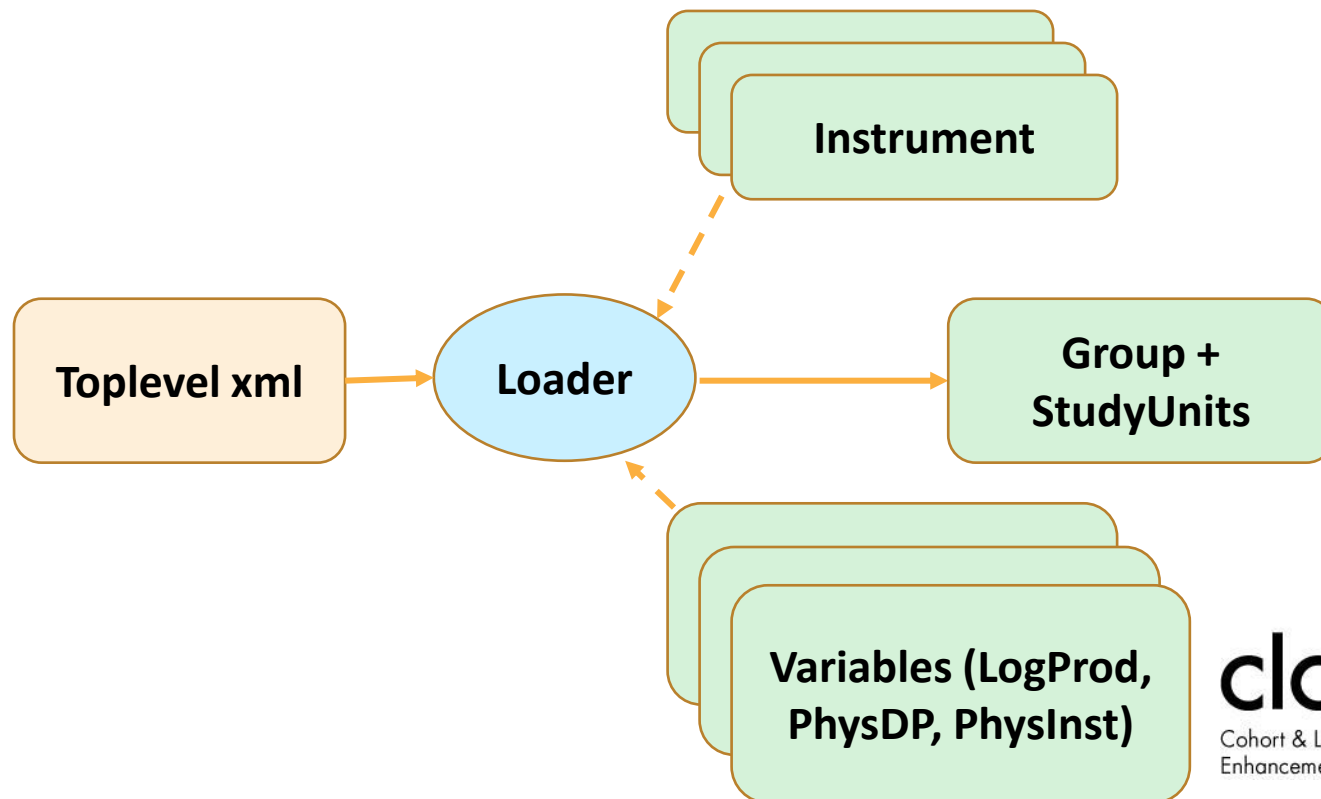


Toplevel

Metadata entered into Colectica Designer

Output is DDI-L(3.2) files, one per study

One study corresponds to multiple questionnaires and variable sets



Mapping Files

IDs, Names or Labels?

alspac_00_ays-qc-000003

qc_A2_b_i

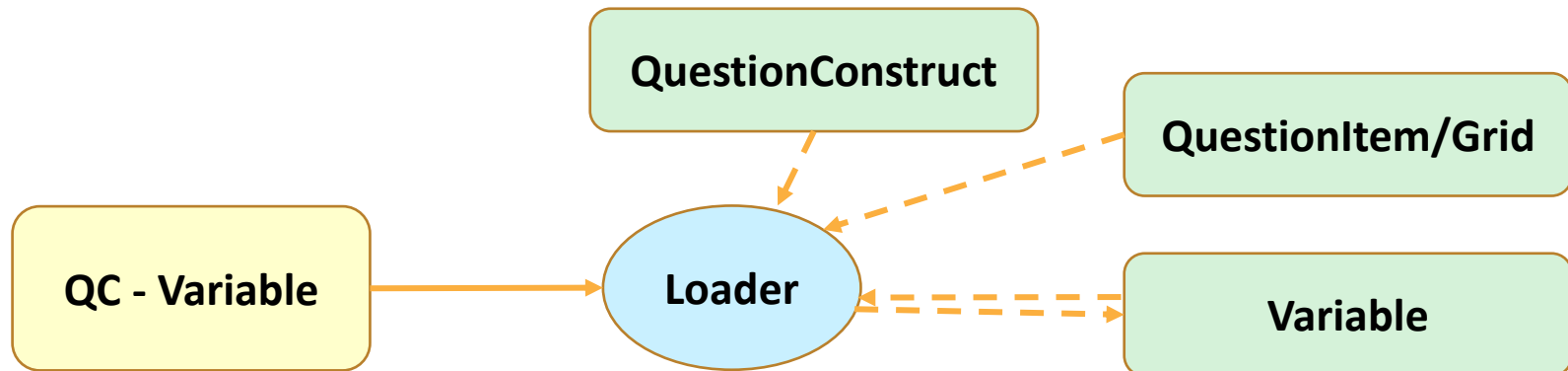
A2 b(i)

Variable Source Map

QuestionConstruct to Variable mapping file

Collected via in-house Mapper software

Loader sets Variable Source to QuestionItem or QuestionGrid



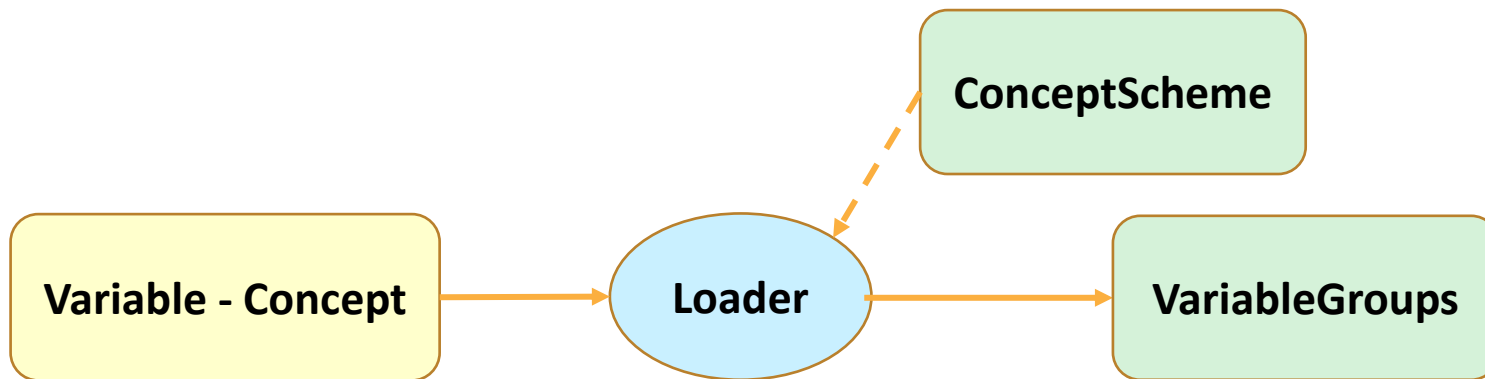
Variable/Question to Concept maps

Variable to Concept map and

QuestionConstruct to Concept map

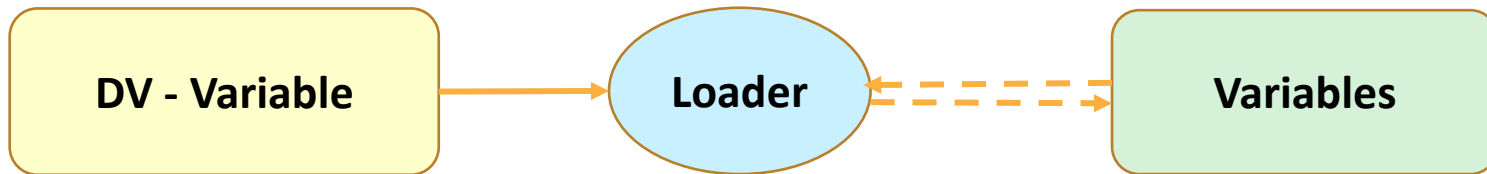
Collected via in-house Mapper software

Loader builds VariableGroups and QuestionConstructGroups

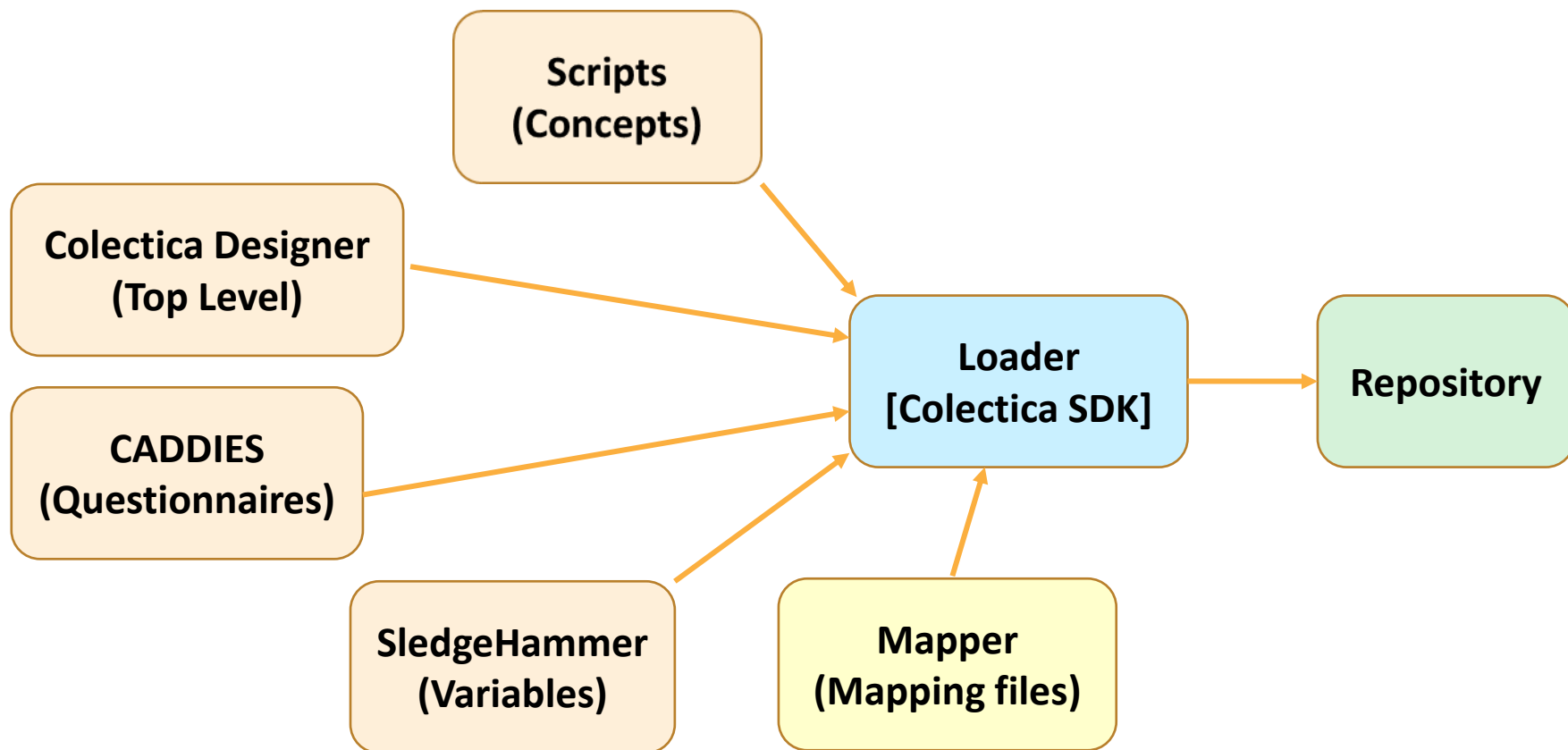


Derived Variable-Source Variable Map

DerivedVariable to Variable mapping file
Collected via in-house Mapper software
Loader sets SourceVariableReference



Conclusion: They all talk DDI-L(3.2)



More Conclusions

Consistent Inputs

Pragmatic Incremental Design

Knowing when to stop

discovery.closer.ac.uk

The screenshot shows the home page of the CLOSER Discovery website. The browser's address bar displays 'discovery.closer.ac.uk'. The website's navigation bar includes links for Home, Search, Explore, and Basket. A search bar is located in the top right corner. The main content area features a 'Welcome' section with a description of the resource and a 'System Status' indicator showing 'Beta testing'. Below this is a 'Search' section with a search input field and a magnifying glass icon. The page also displays five statistics cards: 8 Studies, 33 Sweeps, 27,302 Variables, 12,047 Questions, and 77 Instruments. Logos for the Medical Research Council (MRC) and the Economic & Social Research Council (ESRC) are visible in the bottom right corner.

Home - CLOSER Discovery

discovery.closer.ac.uk

tfi News IOE DDI 3.2 Colectica C# SDK GitHub Tickets caddies moocs Search Results: room b... UCL Library Services bbk lib LLC British Museum swim The Register other Trello

CLOSER Discovery Home Search Explore Basket

Home

Welcome

CLOSER Discovery is an online resource that enables researchers to [search](#) and [explore](#) the data from eight leading UK longitudinal studies. CLOSER Discovery is currently in **beta testing**. We need your [feedback](#) to help us shape this resource to best meet the needs of its users.

To find out more about CLOSER Discovery visit the [CLOSER website](#) or take a look at the [FAQs](#).

System Status: **Beta testing**

Search

Search

8 Studies

33 Sweeps

27,302 Variables

12,047 Questions

77 Instruments

MRC Medical Research Council

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

Copyright © 2015 CLOSER. [View licence agreement](#).