

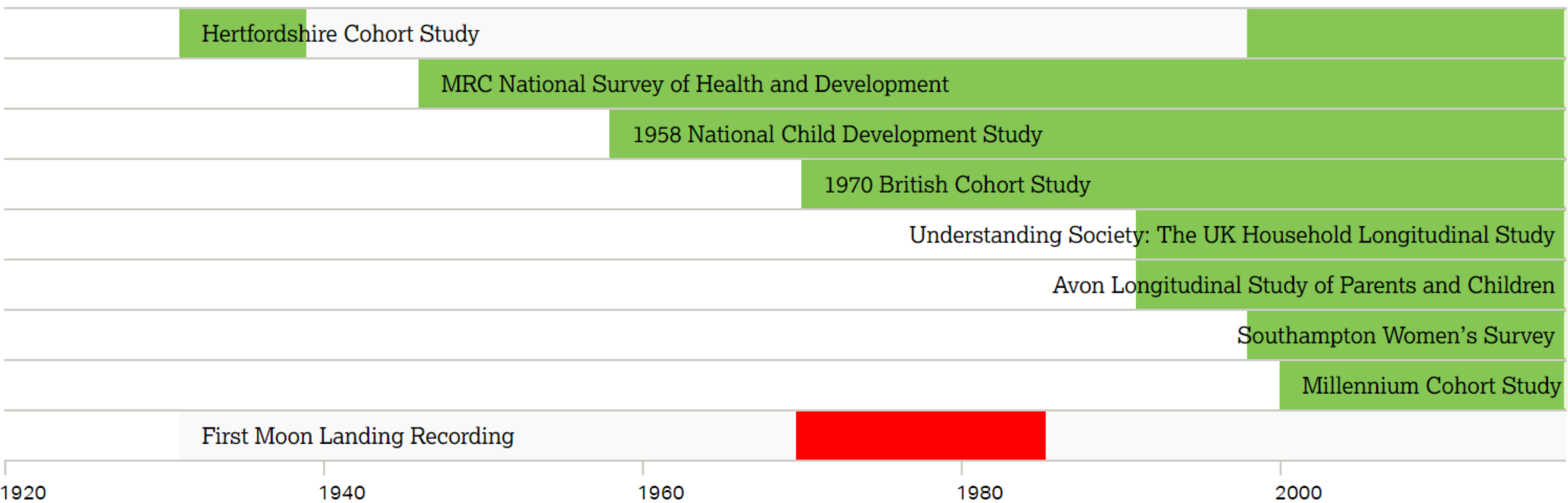
CLOSER Repository

Modernising Longitudinal Study Management

Will Poynter
Technical Manager, CLOSER
20th July 2017

@willpoynter, #esra17

15 second video of first moon landing





Big Data Problem

“Big Data is a loosely defined term used to describe data sets so large and complex that they become awkward to work with using standard statistical software.”

- Snijders, C.; Matzat, U.; Reips, U.-D. (2012)



Big Data Problem

“The dataset has become too large to be collected, analysed or managed efficiently, or a combination of all three.”



Big Data Problem

Over:

- 70 years
- 180,000 participants
- 120,000 variables



Big Data Problem

Interview > Face To Face > CAPI/CAMI

Interview > Face To Face > PAPI

Interview > Telephone > CATI

Interview > Email

Interview > Web Based

Self Administered Questionnaire > Email

Self Administered Questionnaire > Paper

Self Administered Questionnaire > SMS or MMS

Self Administered Questionnaire > Web Based

Self Administered Questionnaire > Computer Assisted

Focus Group > Face To Face

Focus Group > Telephone

Focus Group > Online

Self Administered Writings And Diaries > Email

Self Administered Writings And Diaries > Paper

Self Administered Writings And Diaries > Web Based

Observation > Field > Participant

Observation > Field > Nonparticipant

Observation > Laboratory > Participant

Observation > Laboratory > Nonparticipant

Observation > Computer Based

Experiment > Laboratory

Experiment > Field Intervention

Experiment > Web Based

Recording

Content Coding

Transcription

Compilation Synthesis

Aggregation

Simulation

Measurements And Tests > Educational

Measurements And Tests > Physical

Measurements And Tests > Psychological



What do we want?

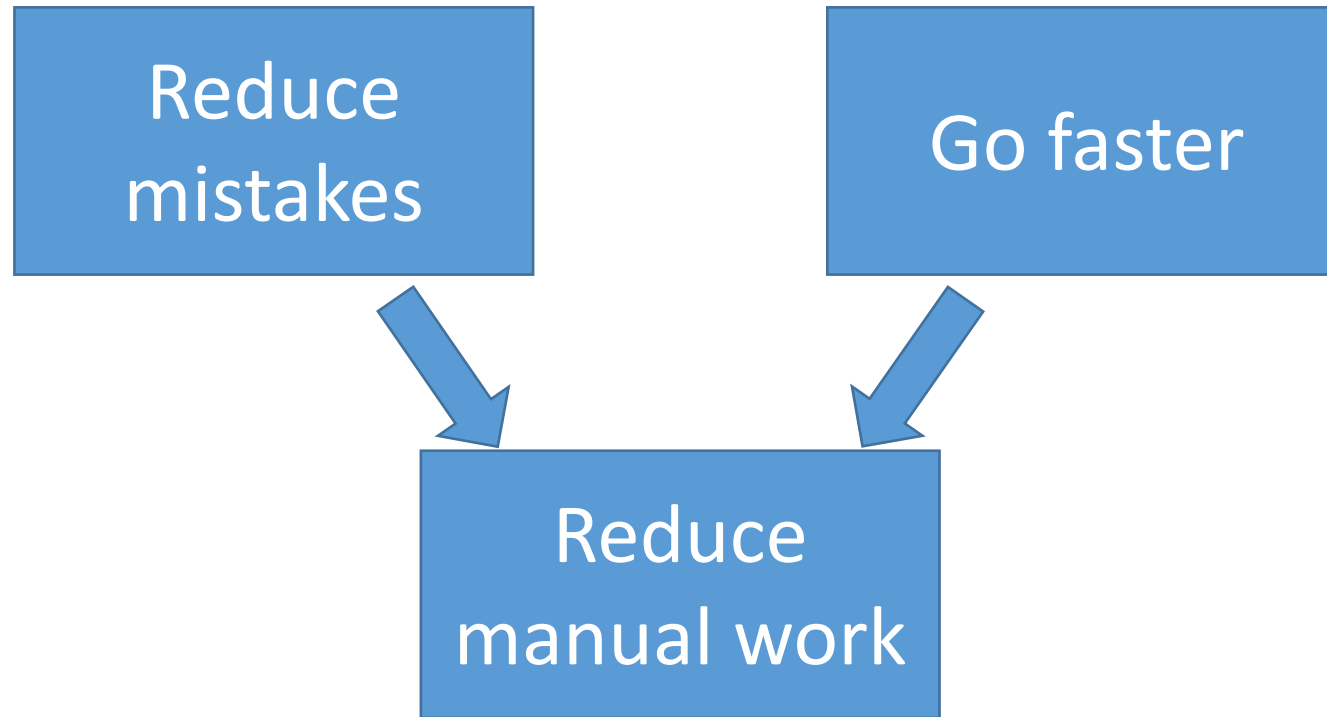
Reduce
mistakes

Go faster

Reduce
manual work



What do we want?



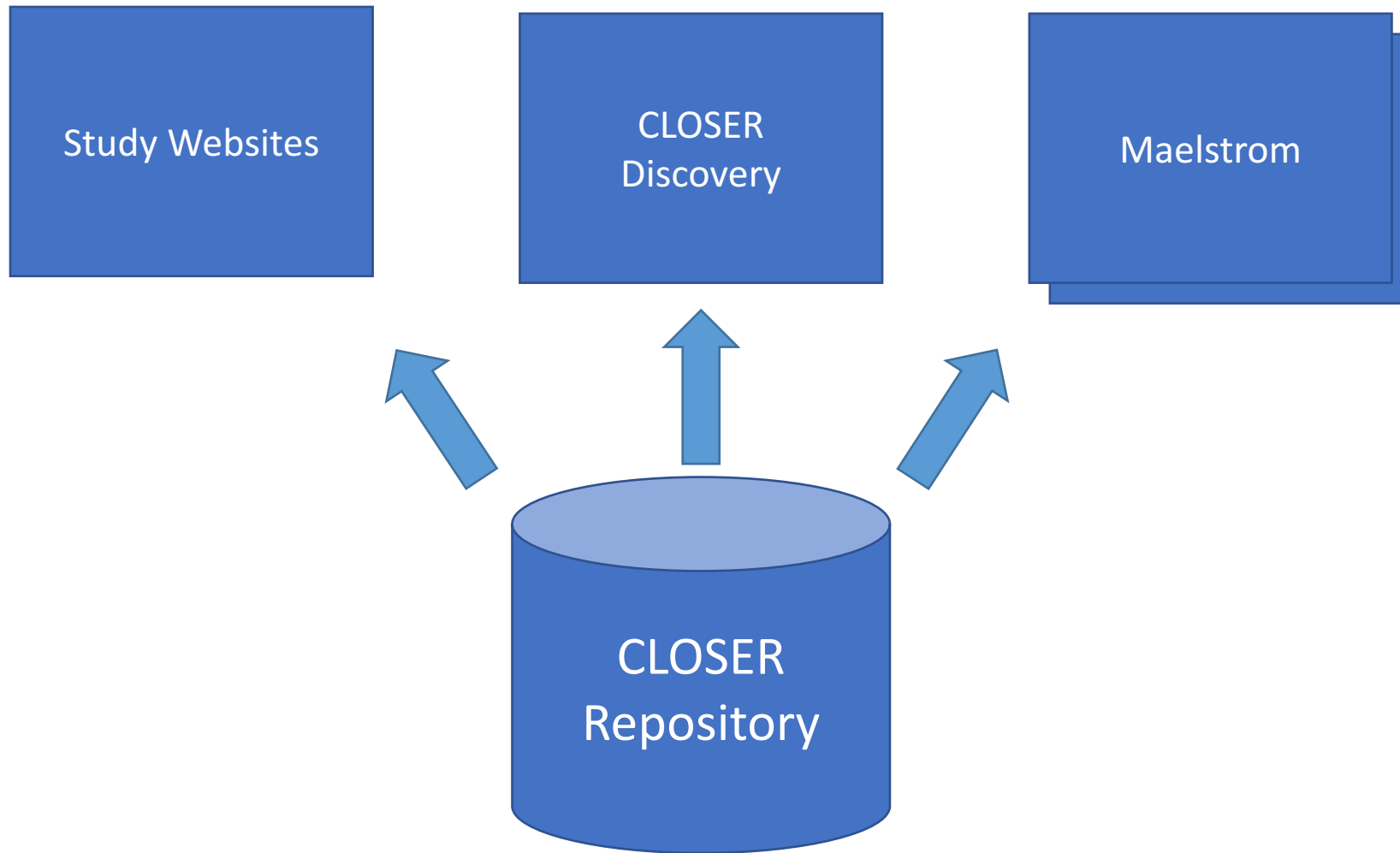


How to reduce work?

- DRY your work
 - Don't Repeat Yourself



Write once, read many



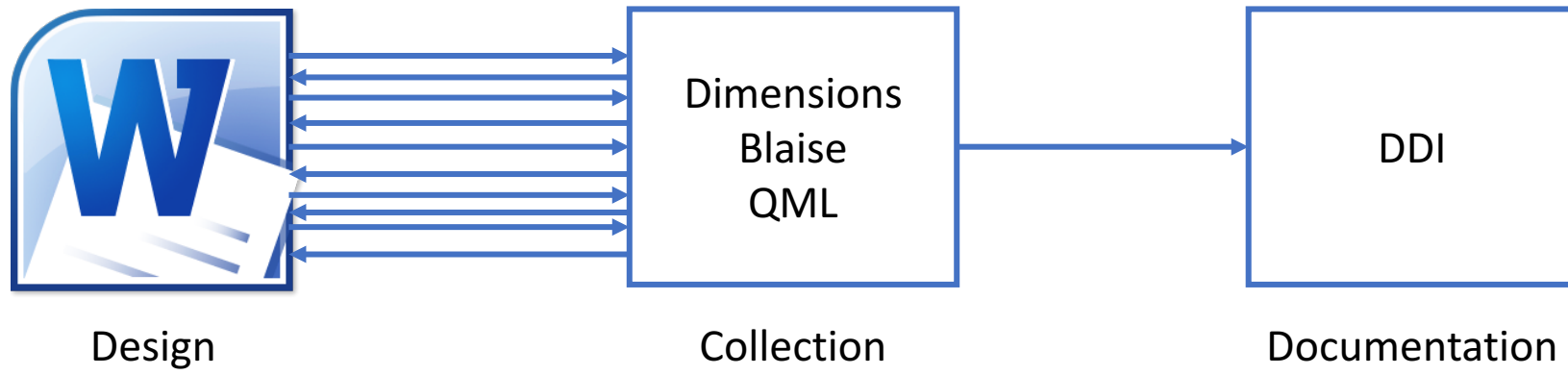


How to reduce work?

- DRY your work
 - Don't Repeat Yourself
- Remove format and software barriers

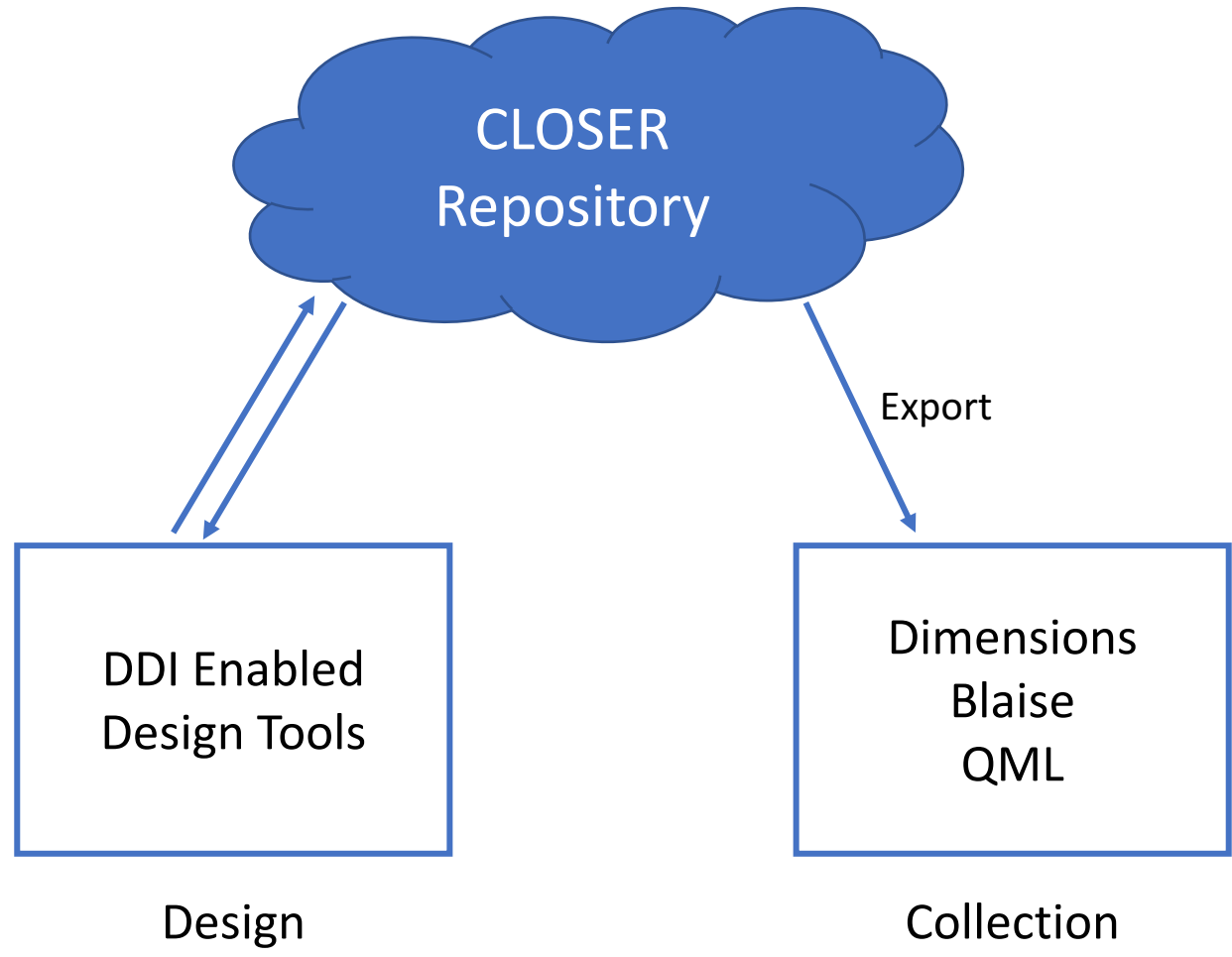


The outlook is cloudy





The outlook is **cloudy**





How to reduce work?

- DRY your work
 - Don't Repeat Yourself
- Remove format and software barriers
- Robots!



Automate

- Create PDFs
- Update variable catalogue
- Apply DDI Concepts (CLOSER Topics)
- Data cleaning?
- Read PDFs?



Thank you

- @willpoynter
- w.poynter@ucl.ac.uk
- @CLOSER_UK
- closer@ucl.ac.uk
- closer.ac.uk
- discovery.closer.ac.uk
- wiki.ucl.ac.uk/display/CLOS/CLOSER
- <https://youtu.be/F-vzi3YlIPQ>