1/
One pre-cleanse pipeline per MGA
One pipeline for central process
Combined outputs
Recommended option

MGA Pipelines

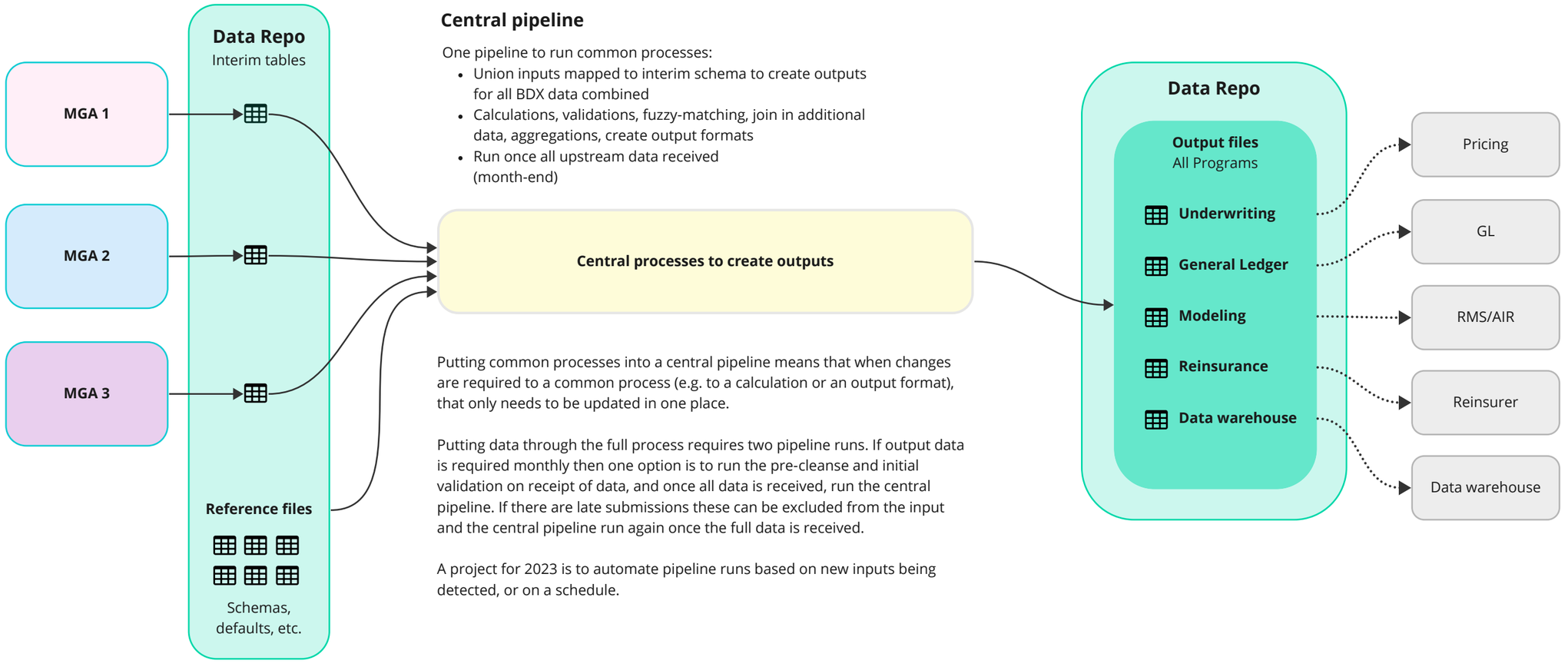
One short pipeline per MGA:

- Get data into single header format
- Map to interim schema
- Initial data quality validation
- Run on receipt of data or at month end

Central pipeline

One pipeline to run common processes:

- Union inputs mapped to interim schema to create outputs for all BDX data combined
- Calculations, validations, fuzzy-matching, join in additional data, aggregations, create output formats
- Run once all upstream data received (month-end)



Putting common processes into a central pipeline means that when changes are required to a common process (e.g. to a calculation or an output format), that only needs to be updated in one place.

Putting data through the full process requires two pipeline runs. If output data is required monthly then one option is to run the pre-cleanse and initial validation on receipt of data, and once all data is received, run the central pipeline. If there are late submissions these can be excluded from the input and the central pipeline run again once the full data is received.

A project for 2023 is to automate pipeline runs based on new inputs being detected, or on a schedule.

2/

One pre-cleanse pipeline per MGA
One pipeline for central process
Separate outputs
Alternative recommended option

MGA Pipelines

One short pipeline per MGA:

- Get data into single header format
- Map to interim schema
- Initial data quality validation
- Run on receipt of data

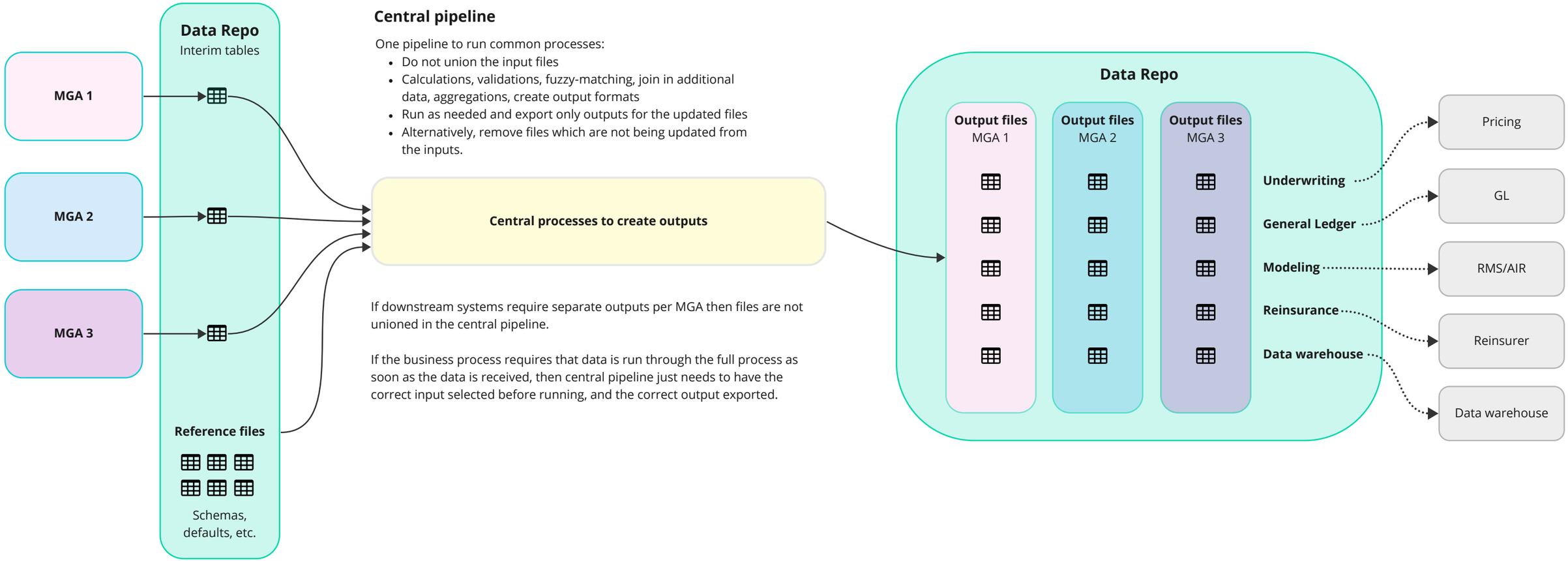
Central pipeline

One pipeline to run common processes:

- Do not union the input files
- Calculations, validations, fuzzy-matching, join in additional data, aggregations, create output formats
- Run as needed and export only outputs for the updated files
- Alternatively, remove files which are not being updated from the inputs.

If downstream systems require separate outputs per MGA then files are not unioned in the central pipeline.

If the business process requires that data is run through the full process as soon as the data is received, then central pipeline just needs to have the correct input selected before running, and the correct output exported.



3/

One full pipeline per MGA process
Separate outputs

MGA Stages

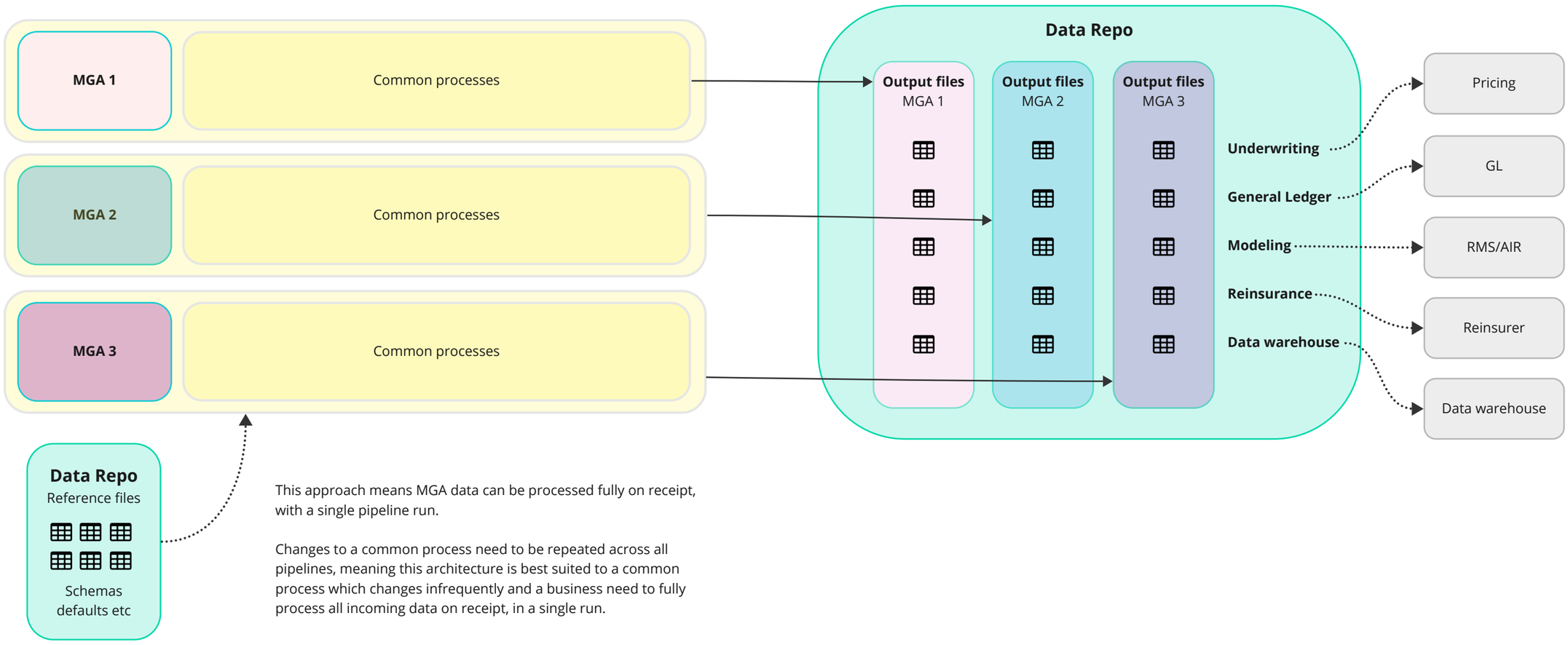
Specific stages per MGA:

- Get data into single header format
- Map to interim schema
- Initial data quality validation

Common stages

Identical stages to run common processes:

- Calculations, validations, fuzzy-matching, join in additional data, aggregations, create output formats
- Run and export as needed



This approach means MGA data can be processed fully on receipt, with a single pipeline run.

Changes to a common process need to be repeated across all pipelines, meaning this architecture is best suited to a common process which changes infrequently and a business need to fully process all incoming data on receipt, in a single run.

4/

One full pipeline per user

Suitable for high frequency SoV data

Input Stages

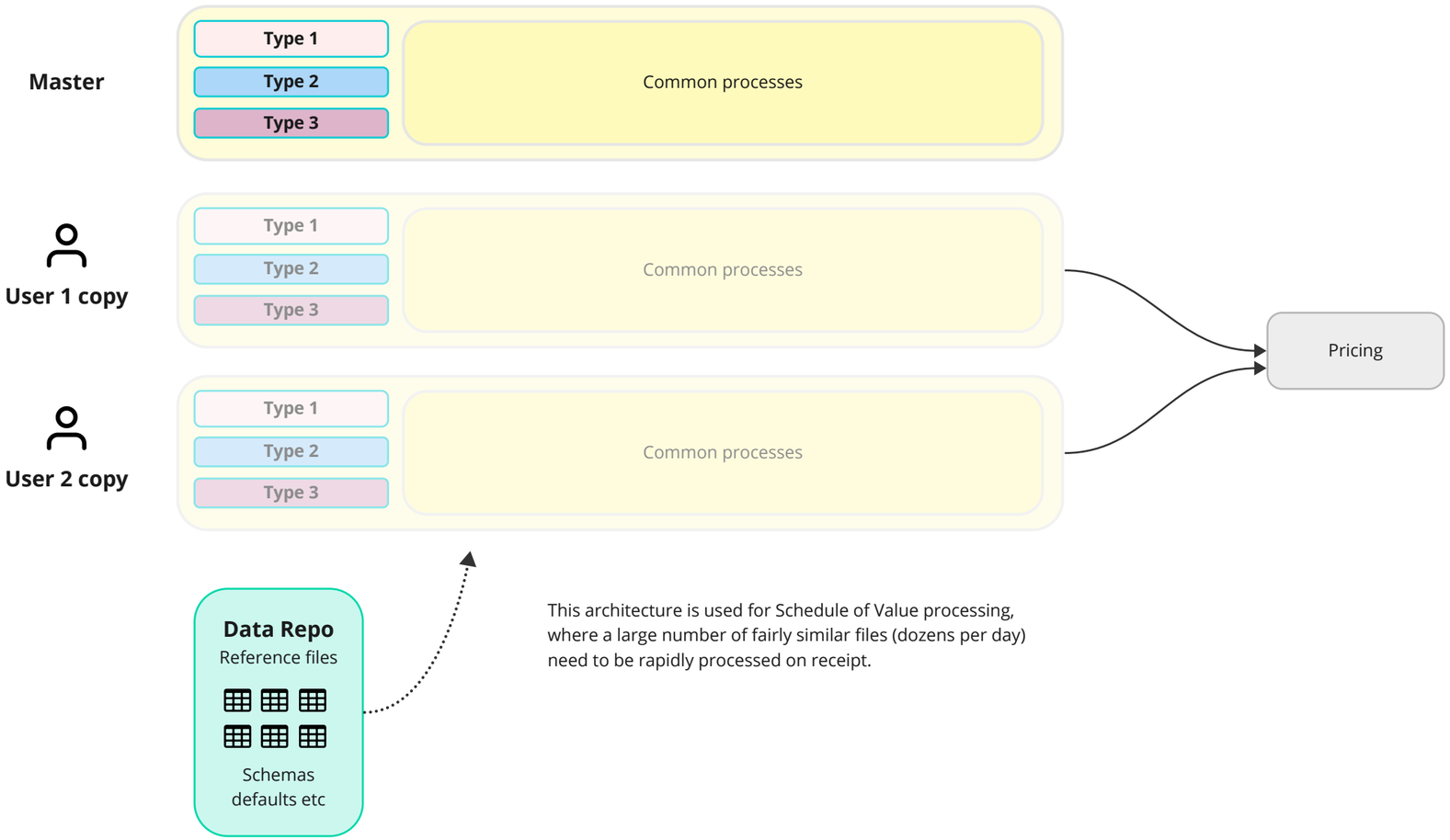
Flexible stages to address common input formats

- Get data into single header format
- Map to interim schema
- Initial data quality validation

Common stages

Identical stages to run common processes:

- Calculations, validations, fuzzy-matching, join in additional data, aggregations, create output formats
- Run and export as needed



This architecture is used for Schedule of Value processing, where a large number of fairly similar files (dozens per day) need to be rapidly processed on receipt.