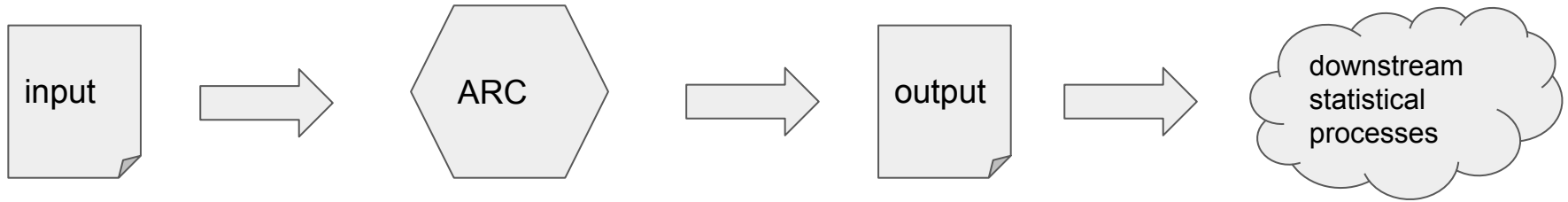# ARC and Relais integration

## I3S Roma meeting, may 2019

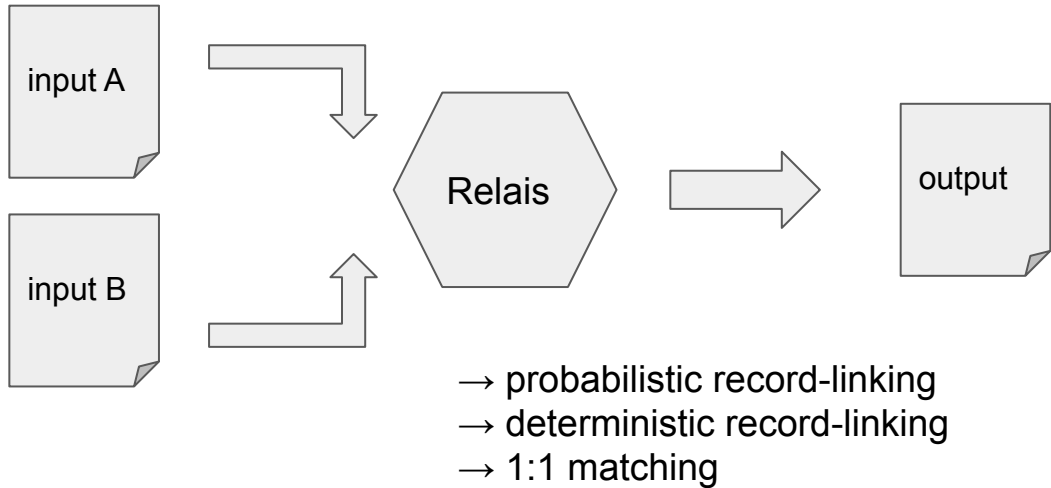# What is ARC ?

ARC is a "*service implements extended functionalities for integrating administrative data into statistical processes*".

input → ARC → output → downstream statistical processes

→ integrate data sources
→ classify and code
→ normalize
→ user defined rules for control, edit, impute.

# What is Relais ?

*"RELAIS (REcord Linkage At IStat) is a toolkit providing a set of techniques for dealing with record linkage projects."*

input A

input B

Relais

output

→ probabilistic record-linking
→ deterministic record-linking
→ 1:1 matching

# Both processes are deeply related, how to integrate them ?
(Paris meeting)
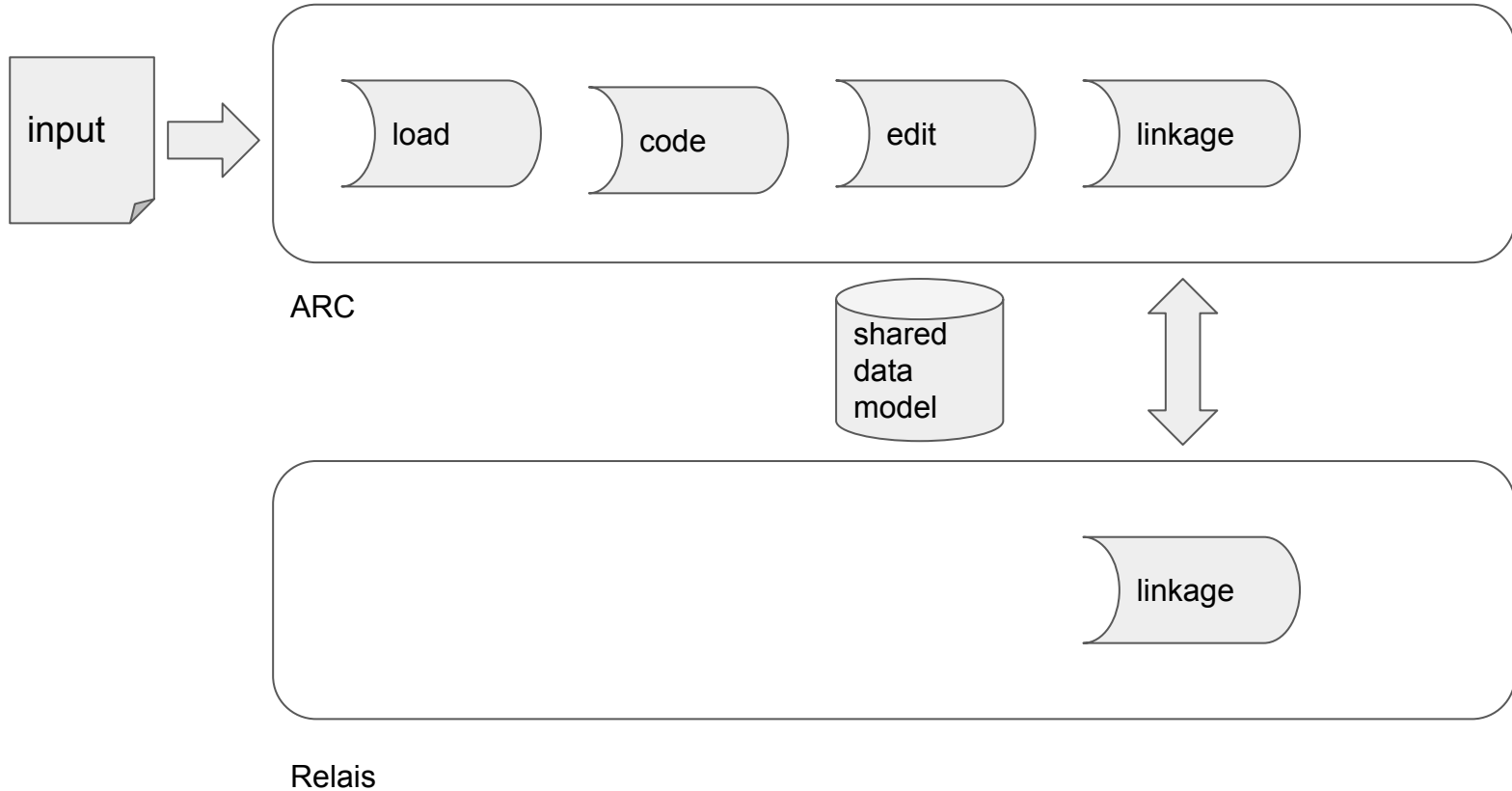
# Integration points

At the business level :

- during the integration phase of ARC, use Relais to link two sources
- at the end of the main ARC process, after validation and editing

At the technical level :

- ARC calls a Relais service through the wire for performing the linkage
- ARC uses a set of Relais methods via Java modules

# Selected use case

# Integration, work hypothesis

Data exchange between ARC and RELAIS performed by translator data model service(s). Following this approach, we could combine and harmonize the following process steps and related modules of ARC and RELAIS

- Data loading [ARC]
- Data coding [ARC]
- Data editing and imputation [ARC]
- Derive new variables and units (temporary simple data model) [ARC]
- Probabilistic record linkage [Relais]
  - search space reduction
  - selection of thresholds
  - definition of match, possible match, unmatch subsets
- Derive new variables and units (complex final data model) [ARC]

# Integration, work hypothesis
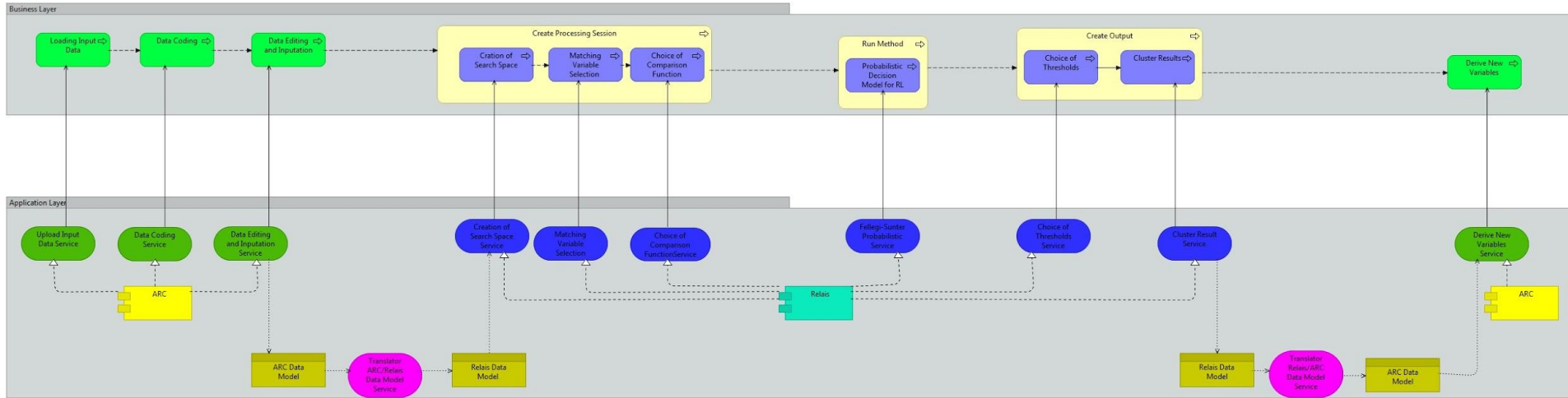
# Pending questions

- is ARC the main UI, with Relais as a supporting service ? Or should we use two different applications, sharing the same data model / same database ?
    - in case of ARC being the main UI, should it carry the UI for configuring the linkage ?
- at least the data models should be unified whatever the technical integration.
    - a potential complexity in the ARC data model
- could we realistically send large administrative files from one service to another ? In the case of two different apps, their should share the same data layer (link to the CSPA data architecture ?).
- another use case is the sequential use of both applications
    - in that case, only the data model need to be shared ?