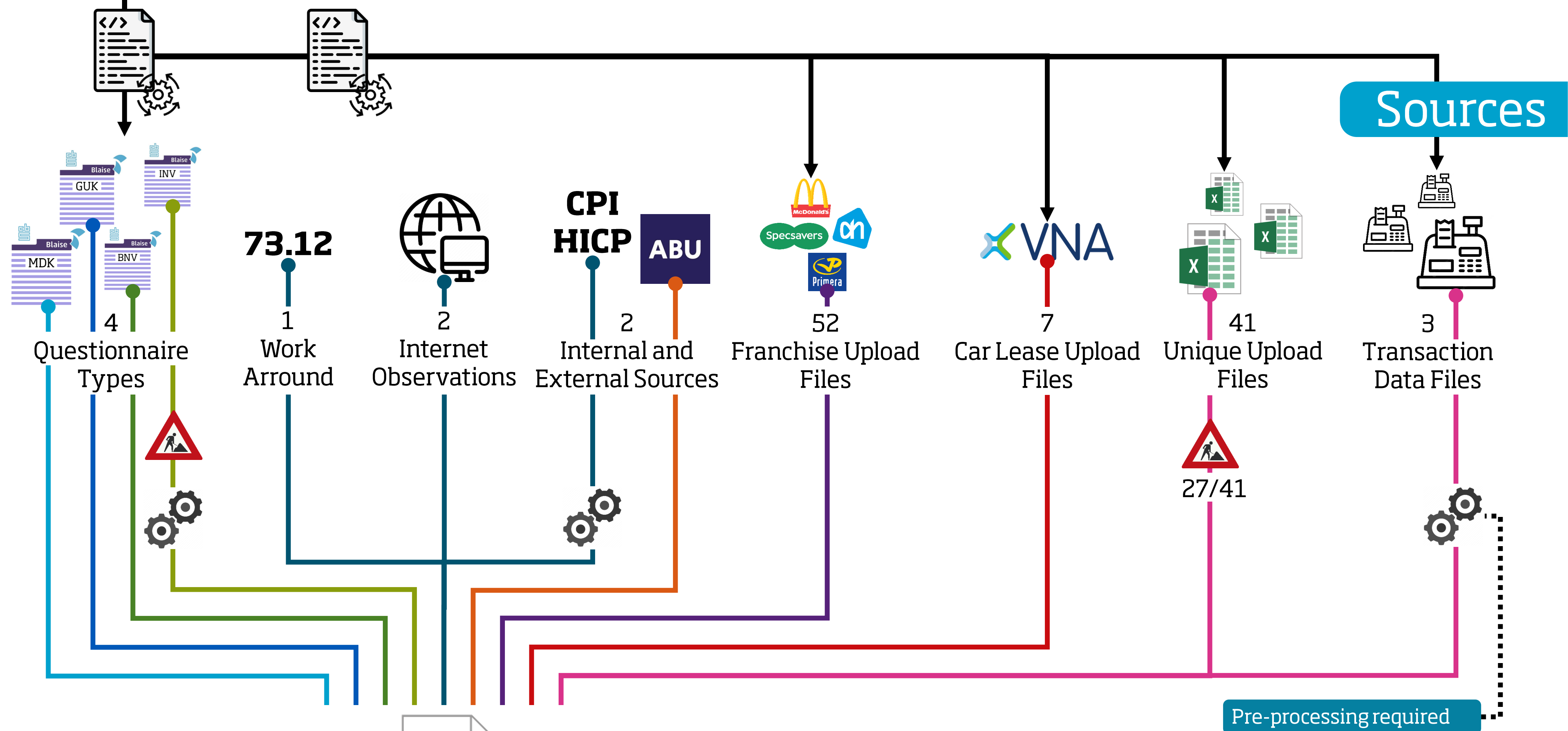


Streamlining pipelines and IT tooling for price statistics



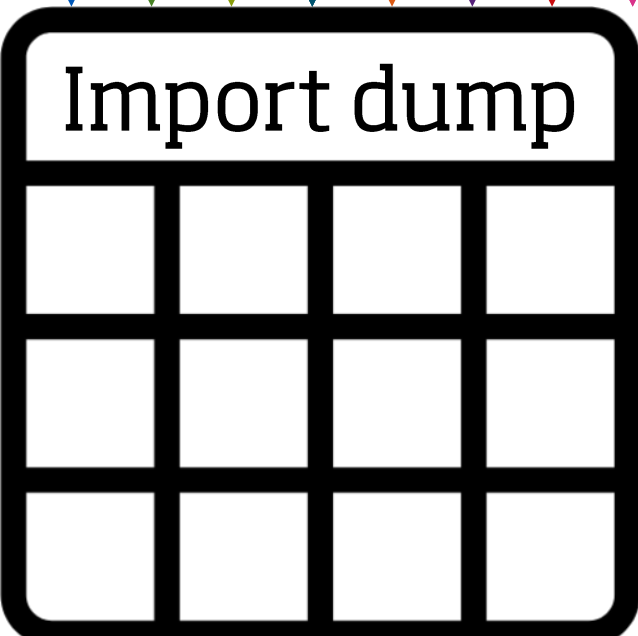
Objectives

1. Automatically retrieve, sort, and read new responses
2. Minimize manual data entry
3. Eliminate custom solutions
4. Minimize data preprocessing before importing
5. Being able to implement annual data model adjustments without modifying the software



Supported file types

1. Excel
2. CVS
3. ZIP
4. BST



File normalization

0	1	2	3	4	5
PRD_15797	150	PERIOD_VAN	Higer wages	4920	EUR
PRD_36858	36	RESPONDENT_KEY	Low water on the Rhine	5041	USD

Linking table

Example linking table

Dynamic SQL	Target table	Target field	Group	#
EOMONTH(CONCAT(SUBSTRING(filename,6,4),'-',SUBSTRING(filename,10,2),'-1'),+2)	PriceStaging	PERIOD_VAN		
SUBSTRING(filename,19,8)	PriceStaging	RESPONDENT_KEY		
IPS_0000	PriceStaging	PRODUCT_KEY		
MAX(CAST(IPS_0001 AS FLOAT))	PriceStaging	PRICE		
CASE WHEN IPS_0005<>' THEN 'EUR' ELSE IPS_0005 END	PriceStaging	CURRENCY		
IPS_0002	CharacteristicStaging	TEXT	Respondent	4
IPS_0003	CharacteristicStaging	TEXT	Service	200
CAST(IPS_0004 AS FLOAT)	CharacteristicStaging	AMOUNT	Service	12

112 Files
9 linking tables
2500 Respondents
11000 Services
45000 Prices per year

Result

Analysts are involved and can make adjustments themselves for importing sources.

Future goals

1. Making other unique upload files 'LUCI proof'
2. Making uploaded files less susceptible to (unintentional) sabotage by respondents
3. Run some manual scripts automatically.
4. Setting up a separate linking table for the transaction data flows, eliminating the need for preprocessing.
5. Import other data (for example weights) and via a linking table



Centraal Bureau voor de Statistiek