Matching food consumption and food composition data:
the challenge of the food linkage

Sandrine Carrillo / Carine Dubuisson (Methodology and survey studies unit)
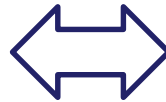Gloria Calamassi / Laure Du Chaffaut / Marine OSEREDCZUK / Martial Ledoux (Food Observatory)
Food agency for food, environmental and occupational health and safety

Richfields Workshop on Dietary Assessment and Food Matching Tools,
8th april 2016, Brussels

Food linkage between 2 datasets

Food consumption data		Food composition table
INCA3 survey (Epicsoft)	⟺	**Table Ciqual**

Food linkage between 2 datasets

INCA3 food consumption data

~3000 food items (FOODNUM)

332 600 lines
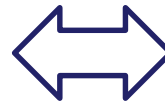
→ 51 125 unique combinations of FOODNUM*facetstring

INTERV

Ciqual food composition table

~ 1800 food items (FOODCIQUAL)

⟷

anses

How to avoid coding all these lines by hand?

Step 1 : Starting from the consumption study to build a linkage table

For each FOODNUM used in INCA3
→ How many item(s) from the composition table?

- **None** (20%, >50% with low frequency)

- **One**, if only one is available (64%)

- **More**, if several codes are available (16%)
  - in this case, add facets & descriptors to the foodnum to describe it more precisely and make it match with the composition food item

anses

Step 1 : Starting from the consumption study to build a linkage table

LINKA3

FOODNUM Name	Ciqual name	FAC1	DESC1	FAC2	DESC2	FAC3	DESC3
potato	potato, baked in oven, without peel	04		03	34	13	01
potato	potato, boiled/cooked in water	04		03	03	13	
potato	potato, steamed, with peel	04		03	04	13	02
potato	potato, canned, drained	04	01	03		13	
potato	potato, steamed, vacuum packed	04	10	03	04	13	
potato	potato, roasted/baked in oven	04		03	34	13	
potato	potato, stir fried, frozen, cooked	04	02	03	94	13	
potato	potato, boiled/cooked in water, without peel	04		03	03	13	01
potato croquette							
potato dauphine	Potato dauphine, frozen, cooked						

No proper food item in the composition table

Only 1 food item in the composition table

If implicit facets in the ciqual name, description with facets/descriptors as well

Several food items in the composition table : description with facet/descriptors

anses

Why ? -> example of the vacuum packed potato, roasted

INCA3 food consumption data

Potato, vacuum packed, roasted

?

Ciqual food composition table

Potato, vacuum packed

Potato, roasted

Step 2 : Prioritisation of facets

Why ? -> example of the vacuum packed potato, roasted

INCA3 food consumption data

Potato, vacuum packed, roasted

Ciqual food composition table

Prioritisation of facets

PR1 = FA 04 (preservation method)

PR2 = FA 03 (cooking methode)

Potato, vacuum packed

Potato, roasted

Name_INCA3	Name_Ciqual	1st Priority		2nd Priority		3rd Priority	
		FAC1	DESC1	FAC2	DESC2	FAC3	DESC3
Potato	Potato, steamed, vacuum packed	04	10	03	04	13	
Potato	potato, roasted/baked in oven	04		03	34	13	

Step 2 : A data extraction of LINKA3 table

A set of « priorities of facets&descriptors » is defined for each pair of FOODNUM/CIQUAL code, to link with the study data.

FOODNUM	NAME	CIQUAL Code		CIQUAL Name	1st Priority		2nd Priority		3rd Priority	
					Fac1	Desc1	Fac2	Desc2	Fac3	Desc3
45	red tomato	20047		Tomato, raw	04	*	03	99	13	
45	red tomato	20048		Tomato, peeled, canned, drained	04	01	03		13	01
45	red tomato	20154		Tomato, cooked in oil	04		03	09	13	
45	red tomato	20242		Tomato, with skin, boiled/cooked in water	04		03	03	13	02

* In the string of priorities, the descriptor of a facet is empty when its presence isn't necessary for the matching

Fac 04:
Preservation method
<u>Desc:</u>
01 = Canned

Fac 03:
Food cooking method
<u>Desc:</u>
99 = raw
09 = cooked with fat
03 = boiled

Fac13:
Skin consumed
<u>Desc:</u>
01 = without skin
02 = with skin

anses

Why ? -> example of the raspberry pie

INCA3 food consumption data

Fruit pie, raspberry

?

Ciqual food composition table

Strawberry pie

Berries pie

Why ? -> example of the raspberry pie

INCA3 food consumption data

Fruit pie, raspberry

Ciqual food composition table

Strawberry pie

Berries pie

FACET 06 = characteristic ingredient
Berries
  Strawberry
  Raspberry
  Blueberry
  Redcurrant
  Blackcurrant

anses

Example of facet 03 : COOKING METHOD

GROUPING

| FACET CODE	DESCRIPTOR CODE	DESCRIPTOR NAME	DESCRIPTOR CODE	DESCRIPTOR NAME	DESCRIPTOR CODE	DESCRIPTOR NAME
3	99	raw, unprocessed				
3	0	cooking method undefined				
3	94	cooked n.s.	34	roasted/baked in oven n.s.		
3	94	cooked n.s.	40	stewed n.s.		
3	94	cooked n.s.	42	stir fried/pan fried n.s.		
3	94	cooked n.s.	44	pressure cooked n.s.		
3	94	cooked n.s.	95	cooked n.s. with fat	8	pan fried/deep fried n.s.
3	94	cooked n.s.	95	cooked n.s. with fat	9	stir fried/pan fried with fat
3	94	cooked n.s.	95	cooked n.s. with fat	10	deep fried
3	94	cooked n.s.	95	cooked n.s. with fat	18	breaded and pan fried/deep fried n.s.
3	94	cooked n.s.	95	cooked n.s. with fat	26	in flour and pan fried/deep fried n.s.
3	94	cooked n.s.	95	cooked n.s. with fat	36	deep fried actifry
3	94	cooked n.s.	95	cooked n.s. with fat	52	breaded and deep fried
3	94	cooked n.s.	95	cooked n.s. with fat	53	breaded and pan fried
3	94	cooked n.s.	95	cooked n.s. with fat	56	in flour and deep fried
3	94	cooked n.s.	95	cooked n.s. with fat	57	in flour and pan fried
3	94	cooked n.s.	96	cooked n.s. without fat	3	boiled in water
3	94	cooked n.s.	96	cooked n.s. without fat	4	steamed
3	94	cooked n.s.	96	cooked n.s. without fat	5	blanched
3	94	cooked n.s.	96	cooked n.s. without fat	6	poached

Grand father

Father

Step 4 : Automatisation

Cross INTERV and LINKA3 tables

- with the « *joinby* » fonction (STATA)

→ all the food items available in LINKA3 for each foodnum will be linked to each occurrence of the foodnum in INTERV (whatever the facetstring)

INTERV	x	LINKA3
	JOINBY	

anses
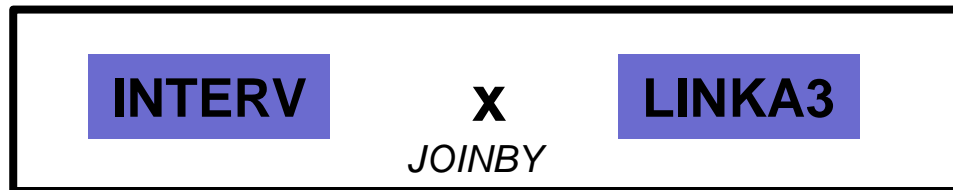
Step 4 : Automatisation

Ex : in LINKA3, 2 ciqual codes for the foodnum 3639 = fried potato
➜ Dans INTERV*LINKA3, all the foodnum=3639 are linked with 2 lines.

INTERV   **LINKA3**

NAME	FOODNUM	FACET_STR_CORR	ORIGFDCD	ORIGFDNM	FACPR1	DESCPR1	
fried potato	3639	1003,2503,0309,2004,1301	4015	fried potato	03	42	X
fried potato	3639	1003,2503,0309,2004,1301	4036	fried potato with duck fat	03	09	✓
fried potato	3639	1003,2503,0342,2005,1301	4036	fried potato with duck fat	03	09	X
fried potato	3639	1003,2503,0342,2005,1301	4015	fried potato	03	42	✓
fried potato	3639	1003,2503,0310,2006,1302	4036	fried potato with duck fat	03	09	
fried potato	3639	1003,2503,0310,2006,1302	4015	fried potato	03	42	

In this table, the facets in the facetstring of each line will be compared to the priority facets and the lines will be kept only if we find them ✓

Otherwise, we try the GROUPING process.

Step 4 : Automatisation

Is there at least 1 ciqual code for this foodnum?

No → No matching possible so far. To check by hand

Yes → For this foodnum, are there PR identified in the LINKA3 table?

No → Unique matching (not depending on the facets) — *33% of INTERV*

Yes → Can we find all the PR in the facetstring of the foodnum?

Yes → Perfect matching — *21% of INTERV*

No → Use the **GROUPING** table

Does the next level of descriptors allow to find them in the facetstring of the foodnum?

Yes → Imperfect matching

No → No matching possible so far. To check by hand

PR = priority facet

anses

Quality and monitoring

First round : conservative

In the LINKA3 table, the associations are as precise as possible (even for the 1:1 matching)

→ Thus, we can check by hand what the program didn't manage to code by itself

→ At this stage, imperfect matching can only come from the GROUPING step

Second round : broader

By checking the blanks, we may implement new rules in the LINKA3 table.

Ex : if there is only one ciqual code for one foodnum and we consider that it would be suitable for all the appearances of the foodnum whatever its facets, we can remove the PR from the LINKA3 table and make the ciqual code match for all the appearances of the foodnum (a tag could be created then).

Third round : checking

Check the final linkage (especially for the more consumed food items) and fill in the blanks

Traps and limitations

- **In the LINKA3 table**

Multiple-choice facets
Very complicated to handle !
Ex : facet 06 on characteristic ingredients, sometimes many descriptors
-> remove spices and herbs?

When brandname can't be overlooked
For specific food groups, brandnames have to be taken into account -> specific approach for:
- Baby food
- Enriched food such as breakfast cereals
- Mineral waters

Traps and limitations

- ## In the GROUPING table

Ex : someone ate a « fried plaice »

In the food composition table we only have « Cooked plaice » for this FOODNUM.

After the GROUPING, « fried plaice » will become « cooked plaice » and it will match.

However, « fried sole » exists in the composition table and it may have been a better choice from the food composition point of view.

- ## In the FINAL table, filling the blanks

- Copy another food item composition, or calculate a mean, etc.
- Deal with food items and descriptors « unspecified » (What is the food « milk n.s. »?, What is a potato « cooked n.s. »?)

Thank you for your attention

anses