



## THE CODING OF JOB HISTORIES

### 1 JOB HISTORIES: WHY CODE THEM?

The collection of diverse data on occupational exposures is unique to Constances, which has no equivalent on this scale. Among the data collected is the professional history of volunteers.

When they are included in the Constances cohort, participants are asked to complete a questionnaire reconstructing their entire professional career since their first job. For each career episode, they are asked to clearly indicate in free text the profession exercised and the corresponding economic activity sector, as well as some additional information (start and end dates of the episode, employment status, type of contract, work time); where applicable, they must provide information on episodes of interruption of employment. An extract from the questionnaire is given in Annex 1.

Collecting and coding job histories are useful for two main types of research needs.

1- Evaluation of occupational exposure to various hazards by linkage with job-exposure matrices (JEMs). A JEM is made up of rows (jobs) and columns (hazards); the cells at the intersection of a row and a column contain indicators of exposure to the relevant hazard: exposure level indicators (which may vary according to the degree of precision: only yes-no, proportion of time of work with exposure, or numerical value) and probability of exposure. JEMs are usually historized (for the same job, there may be several dated periods which correspond to different values of the exposure indicators).

The jobs in a matrix are a combination of an occupation code and an economic activity sector code. In order to link a professional history and a JEM, it is therefore necessary to code the jobs according to the classification used by the matrix.

Researchers may have varying exposure data needs for a given nuisance:

- A lifetime exhibition flag (yes / no)
- Number of years of exposure
- Cumulative value for life or for periods of exposure (in years)

Calculations of cumulative exposure over a lifetime or over defined periods are the responsibility of the researchers, but it is necessary to provide them with the data that allow the calculations.

2- Reconstruction of professional trajectories (with chronology). Some research projects are interested in employment trajectories rather than in occupational exposures.

Job histories can also be an aid for interviewing participants for surveys concerning professional trajectories, by making it possible to have the professional trajectory available during the interview (by interview or on paper questionnaire).

**Important note:** the coding uses French nomenclatures: PCS 2003 for the occupations ([www.insee.fr/fr/information/2400059](http://www.insee.fr/fr/information/2400059)) and NAF 2008 ([www.insee.fr/fr/information/2406147](http://www.insee.fr/fr/information/2406147)) for the sectors of economic activity. The PCS nomenclature is hierarchical and has 4 digits; the NAF nomenclature is also hierarchical with 5 digits. Transcoding in ISCO-ISIC nomenclatures is currently in progress.

## 2 HOW ARE THE JOB HISTORIES CODED?

The coding of job histories is the subject of a collaboration between Constances and IRSET - Inserm UMR 1085 - ESTER team in Angers (Pr. Yves Roquelaure) which is in charge of coding.

The coding consists in assigning, for each participant of Constances, an occupation code and a sector of activity code for each of the jobs lasting at least 6 months; periods of work interruption are also coded.

The coding is done from the job history questionnaire completed at inclusion, and must be updated in the event of a change of job during follow-up. The procedures implemented are as follows:

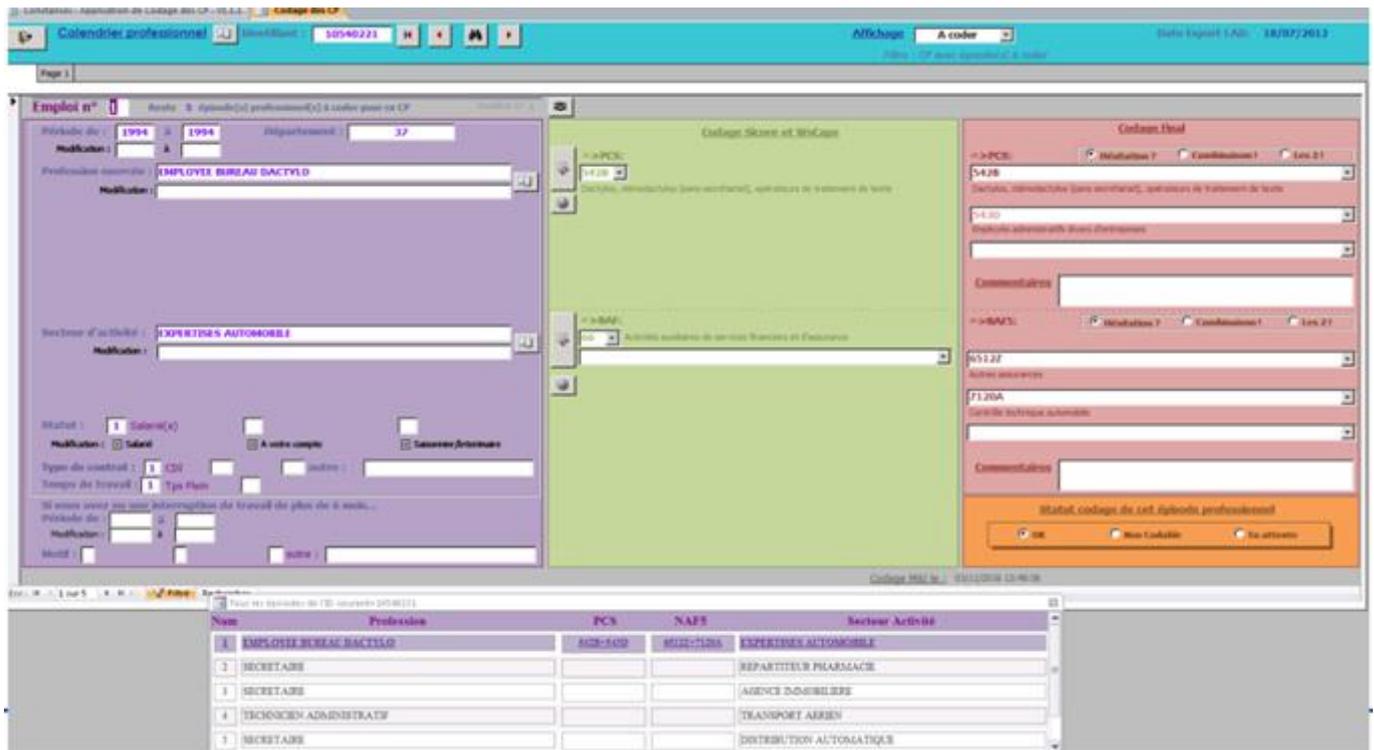
- The job history questionnaires are digitized, as well as the scan of the original paper questionnaires; the file is sent to ESTER-IRSET in Angers, to the team in charge of the coding.
- Rules for checking, cleaning and imputing incomplete or erroneous data have been defined upstream and are applied by ESTER-IRSET. The data entered in the encoding application is directly integrated into a file of the Constances infrastructure.
- Various treatments are carried out in order to obtain a deliverable to provide to users according to their needs:
  - Automatic correction of some incorrect dates
  - Imputation of certain missing dates taking into account the sequence of professional episodes
  - Chronological ordering of the professional episodes
  - Systematic recoding of NAF and PCS codes according to certain associations of sector of activity and profession wordings
  - Cleaning of certain PCS or NAF codes for which there is no ambiguity for the choice between a short or long code: they are systematically coded with the long code (for example PCS student code 8400 instead of 84 ## )
  - Construction of additional variables

### Coding rules

Coding each episode of a job history is complex, and it can be difficult to code an episode correctly when the information provided by participants in their questionnaire is inaccurate or incomplete. This work is done manually by coders, who often have to make decisions requiring significant expertise.

Coders need to analyze the entire job history first, as all of the information provided is critical to coding. This is done through an application created by Constances which provides access to the scanned questionnaires. It allows to fully see all the episodes of the job history. It consists of 3 frames of different colors:

- The purple frame gathers all the data written on the questionnaire which has been digitized.
- The green frame, which is linked to SICORE (automatic coding software developed by INSEE), is an aid for coding, but must be checked before entering the code.
- The pink frame is the entry frame.



Before entering and coding manually, several steps are essential to have a quality coding.

### Step 1: Understanding the job history

The first question when starting to code is: "How did the person answer the questionnaire?" "Indeed, when analyzing the job history, the answers must be analyzed given especially the sector of activity. There are generally three ways to answer the questionnaire:

- Very general responses (e.g. industry, construction, transport, etc.). It is therefore impossible to precisely code the sectors of activity and focus is more on the occupation code which often provide more information.
- More precise answers that provide information on the sector of activity, but not enough to help coding precisely (e.g. accounting, IT, commerce, etc.). The sector of activity must be taken into account and linked to the profession to code as well as possible.
- Very specific answers that inform either the name of the company, or the products that the company manufactures or sells, or even in which sector of activity the person works precisely. The coding will then be the most accurate.

### 2nd step: Analyzing the information

All information provided is essential for better coding and have to be taken into account.

- Is the individual an employee or self-employed? depending on the status, the codes will not be the same (e.g.: salaried IT manager, the code will be 388 #, whereas for a self-employed person the code will be 312e).
- What place does he/she work in? some have an impact on the coding (e.g.: a firefighter who works in Paris or Marseilles will be a military firefighter therefore with an NAF code 8422Z, while firefighters from other regions will be firefighters with a NAF 8425Z code).
- In what year was the profession exercised? It is possible to determine whether the sector is public or private depending on the year and apply it automatically to the code

of the profession (ex: France Telecom in 1960 was in the public sector, while from 2010 it will be in the private sector).

### **3rd step: Coding job histories**

To code the episodes of job histories, one can use either paper nomenclatures published by INSEE, or the CAPS search engine developed by Santé Publique France, which makes it possible to search all the information available in the PCS and NAF nomenclatures by word key. The sector of activity is always coded first taking into account the profession.

- Paper nomenclatures are very useful when starting to code to see all the codes of the same family giving a panoramic view of all the codes in each category and their specificities; this helps to memorize them faster.
- The CAPS computer tool is a keyword search engine. It gives a faster result, but less precise. Indeed, it only provides the results by keyword, so it is often necessary to check several codes before finding the most relevant.

When starting coding, for best coding, it is recommended to use both sources. One can start by doing a search on CAPS, being careful to choose the code correctly by checking it in the paper nomenclatures which give more details on each code (see examples of coding procedure for job histories in [appendix 2](#)).

### **4th step: Checking the codes for obtaining consistency**

When coding a job history, it is important to have consistency between each episode.

- Career development have to be taken into account. Example: if a person has been an executive at some point in his/her career, he/she must remain in that level for subsequent episodes.
- The full professional career must also be checked to see if each episode makes sense. Often a person who stays in the same field will have a career development which will have to be represented in the codes. Example: the profession codes will generally follow one another: 5 ###, 4 ###, 3 ###.
- To code as accurately as possible, hierarchy and / or qualification are taken into account. Example: skilled worker or not, employee or supervisor, manager or not manager...

To code as correctly as possible, all these steps have to be performed. When all of the information is observed and a doubt remains, looking at the scanned questionnaire may be useful. The important thing is to take into account all the data provided and put them together to obtain the best coding.

## **3 DELIVERABLES**

The results of the coding process are presented in different databases.

- The initial database of corrected job histories containing the raw and the cleaned variables, variables for flagging modification of dates or indication of consistency or inconsistency of the professional histories.
- A database constructed in chronological order of a professional history per participant. This database can be organized in two ways, depending on the user's request:
  - One line corresponds to a complete professional episode with the start and end dates. A history can include one or more professional episodes.
  - Employment and interruption periods appear by calendar year, year by year. One line corresponds to a year. There are no more period start and end dates. There are as many lines as the period of employment or interruption lasts in years.

- The database provided to the user contains all the constructed variables useful for judging the quality of the sequence of calendar dates; it is accompanied by detailed documentation of all the controls and treatments carried out, of the variables created and how to use of the job history database.

As of December 2020, the completed job histories including 198,255 participants, comprising 637,796 career episodes had been coded. As inclusions are not completely completed, coding should continue, as well as job changes during follow-up.

## ANNEX 1: JOB HISTORY QUESTIONNAIRE



### Comment remplir ce questionnaire

- Décrivez tous les emplois de **plus de 6 mois** que vous avez eus, **en commençant par votre premier emploi**.
- Si vous avez occupé plusieurs emplois très différents dans la même entreprise, décrivez-les séparément.
- Si un emploi a été suivi d'une période sans emploi (chômage, raisons de santé...), décrivez-la dans le cadre réservé à cet effet en précisant la raison.
- Notez la période de début et de fin d'emploi. Si vous ne vous souvenez plus de l'année exacte, indiquez une année approximative. Si votre dernier emploi est encore en cours, ne remplissez pas l'année de fin.
- Pour la profession exercée, évitez de noter des informations trop vagues, comme «fonctionnaire» ou «mécanicien» mais précisez exactement votre emploi (professeur ou mécanicien poids lourd).
- Ce calendrier est prévu pour 7 emplois, si vous en avez exercé un plus grand nombre, continuez sur une feuille libre sur le même modèle.

### Exemple :

Période de : 1 9 9 8 à 2 0 0 1 Département : 3 3

Profession exercée : P E I N T R E E N C A R R O S S E R I E

Production ou secteur d'activité : R É P A R A T I O N A U T O M O B I L E

Statut :  Salarié(e)  A votre compte  Saisonnier / intérimaire

Type de contrat :  CDI  CDD  Autres

Temps de travail :  Temps plein  Temps partiel

Si vous avez eu une interruption de travail de **plus de 6 mois** entre cet épisode professionnel et le suivant, précisez la durée et le motif :

Période de : 2 0 0 1 à 2 0 0 2

Motif :  Santé  Chômage  Autres

## CALENDRIER PROFESSIONNEL

### Premier emploi

Période de : A A A A à A A A A Département :

Profession exercée :

Production ou secteur d'activité :

Statut :  Salarié(e)  A votre compte  Saisonnier / intérimaire

Type de contrat :  CDI  CDD  Autres

Temps de travail :  Temps plein  Temps partiel

Si vous avez eu une interruption de travail de **plus de 6 mois** entre cet épisode professionnel et le suivant, précisez la durée et le motif :

Période de : A A A A à A A A A

Motif :  Santé  Chômage  Autres

## ANNEX 2: EXAMPLES OF JOB HISTORIES

(only French nomenclatures are currently used)

### Example of job history with 1 professional episode

START DATE OF PERIOD	END DATE OF PERIOD	STATUS	ACTIVITY SECTOR	OCCUPATION	NAF 2008 CODE	PCS 2003 CODE	COMMENTS NAF	COMMENTS PCS
1995	2008	Salarié	Tertiaire	Employé de bureau	#####	543d	The term "tertiary" is far too vague to encode the NAF. This label cannot be coded.	Search result by keyword on CAPS with precision QND (qualification not defined).

### Example of job history with 3 professional episodes

START DATE OF PERIOD	END DATE OF PERIOD	STATUS	ACTIVITY SECTOR	OCCUPATION	NAF 2008 CODE	PCS 2003 CODE	COMMENTS NAF	COMMENTS PCS
1975	1978	Salarié	Métallurgie	Tourneur fraiseur	24## #	623f ou 623g	Not having more precision of the sector of activity, one can only code the first 2 digits of the NAF.	Looking at his entire professional career, one sees that in the next episode he is a production manager. He is certainly a skilled worker. On the other hand, one does not know if he is working on small or large series, so we cannot choose between the 2 codes.
1978	1991	Salarié	Sidérurgie	Responsable production	2410 Z	474c	Search result by keyword on CAPS	Since one don't know if he's a manager or not, he is coded as a supervisor.
1991	2003	À son compte	Chaudronnerie	Chef d'entreprise	3320 A	212b	Search result by keyword on CAPS	As he is self-employed, he has an independent status code, and knowing his activity and in view of his professional calendar, he is put into a boilermaker artisan.

**Example of job history with 5 professional episodes**

START DATE OF PERIOD	END DATE OF PERIOD	STATUS	ACTIVITY SECTOR	OCCUPATION	NAF 2008 CODE	PCS 2003 CODE	COMMENTS NAF	COMMENTS PCS
2008	2011	Salarié	Éducation à l'image	Chargé de mission	85###	543d	We know that the sector of activity is education but do not know at what level	By observing the following episode, we can deduce that he is responsible for communication in training. We can conclude that he is certainly an administrative employee in a grade 5 position and that he is training in the following episode
2011	2013	Salarié	Mécénat culturel	Chargé de communication	8412z	464a	It is part of the Ministry of Culture	Knowing that he is still in "apprenticeship" and that in the previous episode he was already on assignment, he is put in class 4 because he continues his training
2014	2014	Salarié	Cinéma	Production assistant	591##	465b	Cinema is a large sector of activity, but since he is an assistant in production, we have a little more information to code, so we can code the first 3 digits of the NAF	Search result by keyword on CAPS
2015	2018	Salarié	Syndicat Professionnel (cinéma)	Responsable des affaires économiques	9420z	461f	Search result by keyword on CAPS	He is administrative and financial manager. This code is the one that is closest to its function
2019	----	À son compte	Syndicat Professionnel (cinéma)	Délégué général	9420z	335a	Professional Union (cinema)	Even if he is self-employed, he exercises a trade union mandate, so we code him in this category