

## Room for other projects

[NEOS](#) focuses on environmental risk factors in France, in air and water, with a strong ontology and evidence based science focus.

It is **useful to be united** to be more efficient with our limited number and personal time, and therefore not to create various projects. **But** some may feel more useful differently. Therefore, and also in order not to forget essential bricks of the cancer and environment puzzle, as NEOS was being described we tried to list the parts of the puzzle it does not address, so that others can complement the work:

- **Data analysis on how to use risk factors across time and space**, to predict risk factors. This is part of the NEOS approach but can be handled with a separate project.
  - For example  $Y=f(X)$  where Y is cancer risk per age, year, department and body location (breast/prostate/colon etc) and X is available per day over many years and sub-city locations, is there a link between both?
  - Ex: a randomforest or XGBoost analysis (a regression that uses bagging, given the high number of explanatory variables) could then lead to analyse (using variable importance, typically) what elements of X matter in predicting Y<sup>1</sup> The topic is discussed here: <https://app.ioql.io/post/2045#comment-1294>
  - Of note, institutions like IARC have already spent years analyzing relationships, so searching the scientific litterature for an article that indicates how to specifically use X probably makes more sense. Still, there are cases when the measures of X or Y are not done like in the litterature - then data makes sense. Also when X is lacking, then regressing the meaning of not finding this information makes sense.
  - **Who would strongly prefer contributing here?** (for maximal collective achievement, useful to read [NEOS](#) first)
- **Access to healthcare**
  - If you live in a medical desert, you will very probably be diagnosed later in case of cancer and we also have more difficulties getting adequate treatments and surveillance. Of note, this is not in the IARC types of factors
  - **How to measure this ?**
- **Fdep/social status at local level** *Risk that there is dispersion*

---

<sup>1</sup> X: the most granular data on air pollutants I know in France is <https://www.data.gouv.fr/fr/datasets/donnees-temps-reel-de-mesure-des-concentrations-de-polluants-atmospheriques-reglementes-1/> - however one needs to make the link between things like "Metz-Centre" with a department

Y: the most granular cancer risk statistics I know in France is <http://cepidc-data.inserm.fr/inserm/html/index2.htm> - it is by department and big-city level (as well as age, gender, and CIM-10 mortality cause) - however it is needed to code a scraper to extract much data. Also it does not contain the number of people living there. Alternatively here you find cancer mortality rates by age at department level: <https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers>

- In France, the Inserm unit “CépiDC” have created a geographical indicator called [Fdep](#) that is designed to represent risk specifically represents social status and published that it is correlated with mortality
- Of note, the IARC monographs indicate 5 types of cancer risk factors. NEOs focuses on the three first ones. The 4th one is occupation (so we think of social status), the 5th one behavior like smoking. Even though behavior likely is the most important component, we have to admit it is less amenable with the goal to give sense for individual data: if there are 3% heavy smokers in your local area but you are not one of them (most likely situation), this does not affect (or little) your cancer risk.
- *Risk that there is dispersion.* Asking for such individual data might be easier and better than associating such a location-based indicator
- **Other countries.** *Risk that there is dispersion of human resources.*