PROJET NEOS – PROJECT SUMMARY

SLIDE 1

NEOS stands for **NEoplasm Open-Source Environmental Risk Factors Standardization**. It is the project we have submitted to Epidemium Season 3.

SLIDE 2

We decided to apply to Epidemium's Challenge number 2, which is about « Accelerating research on cancer risk factors by structuring open data sources, and completing the OSIRIS clinical and -omics databases, in order to standardize variables related to the environment ».

SLIDE 3

The context of this application is multi-fold:
- First: the proposal last year by the OSIRIS consortium of a minimal set of clinical and genomic items using international standards and terminologies enabling a strong interoperability. OSIRIS common data model is modular and extensible to other types of data
- Second: the weight of environmental factors in the incidence and impact of cancers, 20% of them globally being linked to such factors according to the WHO
- Third: the heterogeneity of the environmental data sources, that makes it difficult to interpret the association between environmental exposures and health outcomes

SLIDE 4:

To answer the challenge, we gathered a team of four skilled, experienced and motivated professionals around the objective of easing analyses of environmental cancer risk factor data by structuring and harmonizing open-source epidemiological data sets, in a FAIR approach.
We organized our work into three workpackages: ontology, data & metadata, data sources.
The target outcome we defined was a standardized cancer epidemiology dataset framework and examples.

SLIDE 5:

We spent significant time defining the precise scope of our work.
We got to the conclusion that we had to select easily measurable risk factors. This is why we focused on air & water agents, in France.
In order to leverage experts' work and to use an International reference, we decided to use the monographs of IARC, The International Agency for Research on Cancer.
And we limited our choice to IARC Group 1 carcinogens, substances proven to have carcinogenic potential for humans.
At the end of the day, we worked on cancer environmental risk factors measurable in air and water media, belonging to IARC Group 1, corresponding to open data sources generated in France, for an epidemiological purpose.

SLIDE 6:

From a list of the 37 IARC Group I air & water biological, chemical and physical agents with open-source data that we identified, we selected two of the principal carcinogens in France:
- An air pollutant: **PM 2.5** (fine particle matter), associated with lung cancer risk
- A water pollutant: **arsenic**, associated with lung, urinary bladder and skin cancer risk

SLIDE 7:

We explored different data sources:
- IARC monographs
- Multiple open-source databases on PM2.5 and arsenic
- France's Ministry of Environment beta web portal on health environmental risk factors
- Numerous bibliographical references
- Queries to OSIRIS and GEOCANCER specialists, Marc Fournier, Mehdi Benchoufi

And we used the following tools:
- Collaborative tools
  - GitHub
  - JOGL
  - Slack
  - Drive
  - Inclusion of the NEOS project in the OpenData4Health site
- Google, Pubmed
- Google Image
- The OSIRIS datasets

SLIDE 8:

The outcome of our work is the NEOS framework, summarized here through its main fields, and available in an OSIRIS-like excel format, with the corresponding definitions and sources.

SLIDE 9:

The conclusions of our work are the following:
- Open-source environmental data are very heterogeneous: harmonization of references is highly needed for interoperability
- Two types of data are crucial for the NEOS Framework: place of residence/occupation, total duration of exposure
- Definition of variables must be in context and precise to avoid bias
- Data collection and analyses at the patient level require a precise address and geocoding

We identified possible limitations, particularly for rural areas: the place where agent measurements are obtained and their geographical coverage may prevent extrapolations to the rest of the territory.

SLIDE 10:

Regarding perspectives of this work:
- In the coming months, it will be expanded by our team and additional contributors to other IARC Group I environmental cancer risk factors with open sources, using the NEOS Framework
- Final use can be two-fold
  - Etiology of cancers in diagnosed patients
  - Study of populations cancer risk for prevention
- Satellite and GIS data sources could help filling the data gaps in the future

SLIDE 11:

Last but not least, we would like to thank the organizers for giving us the great opportunity to contribute to this work so much needed for the patients with cancer and the populations at risk.