



a toolkit for
a learning health system



Camila
Arias



Olivier
Birot



Kim Tam
Huynh



Adrien
Coulet



Antoine
Neuraz



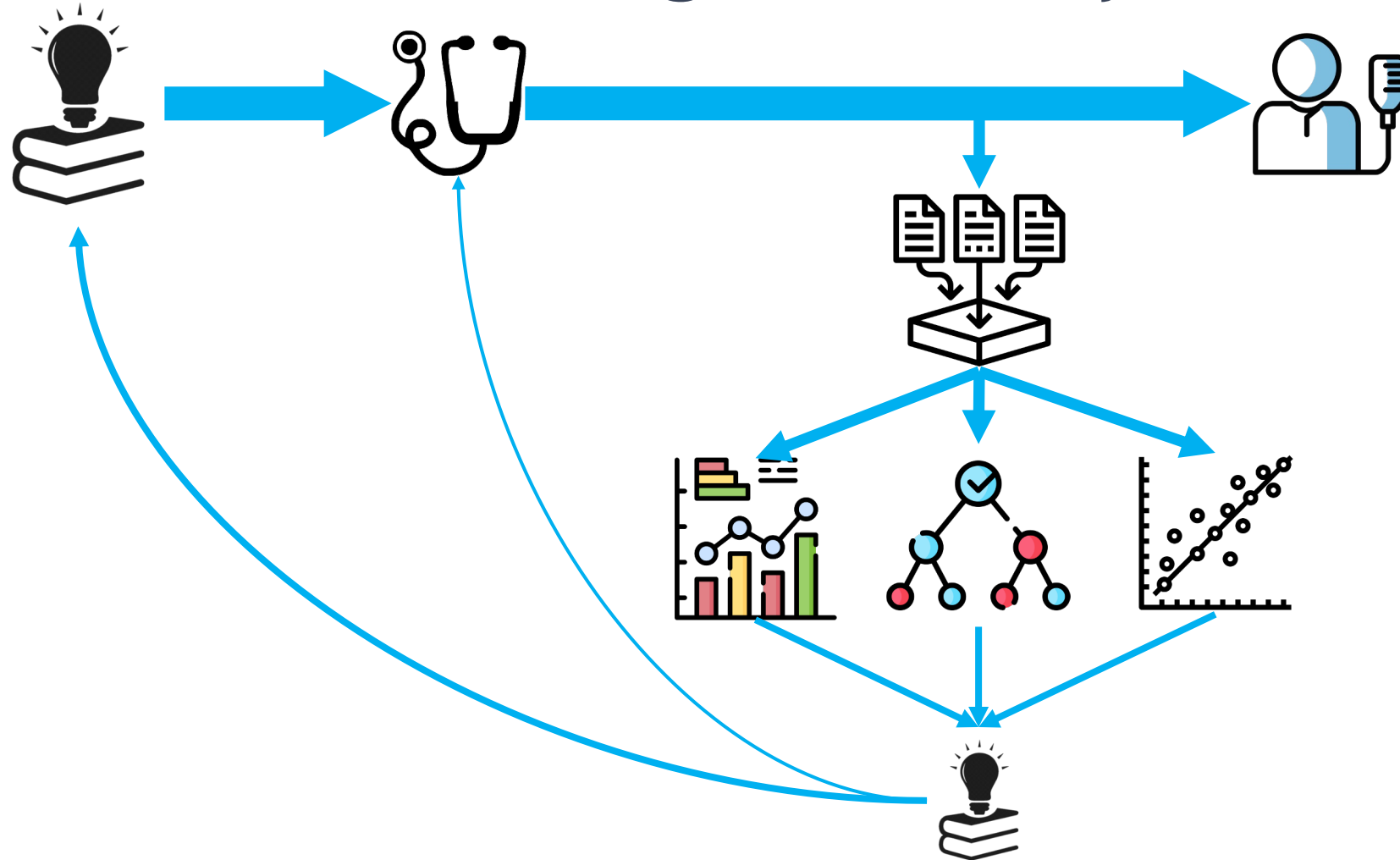
Ivan
Lerner



Bastien
Rance



Context: a learning health system



Extracting « clean » data from EHRs

oracilline 500mui x 2 par jour



**Dictionary
+ Regex**

`([0-9]+[.,]?[0-9]{0-3} ?(mui|mg|g|...)) ?x ?(par jour|/j|/ j)`

Extracting « clean » data from EHRs

oracilline 500mui x 2 par jour



Class

la pancytopenie s'est compliquée après la chimiothérapie d'un sepsis à escherichia coli résistant à la tazocilline (tazocilline* depuis le 6 septembre) traité par fortom à partir du 15 septembre

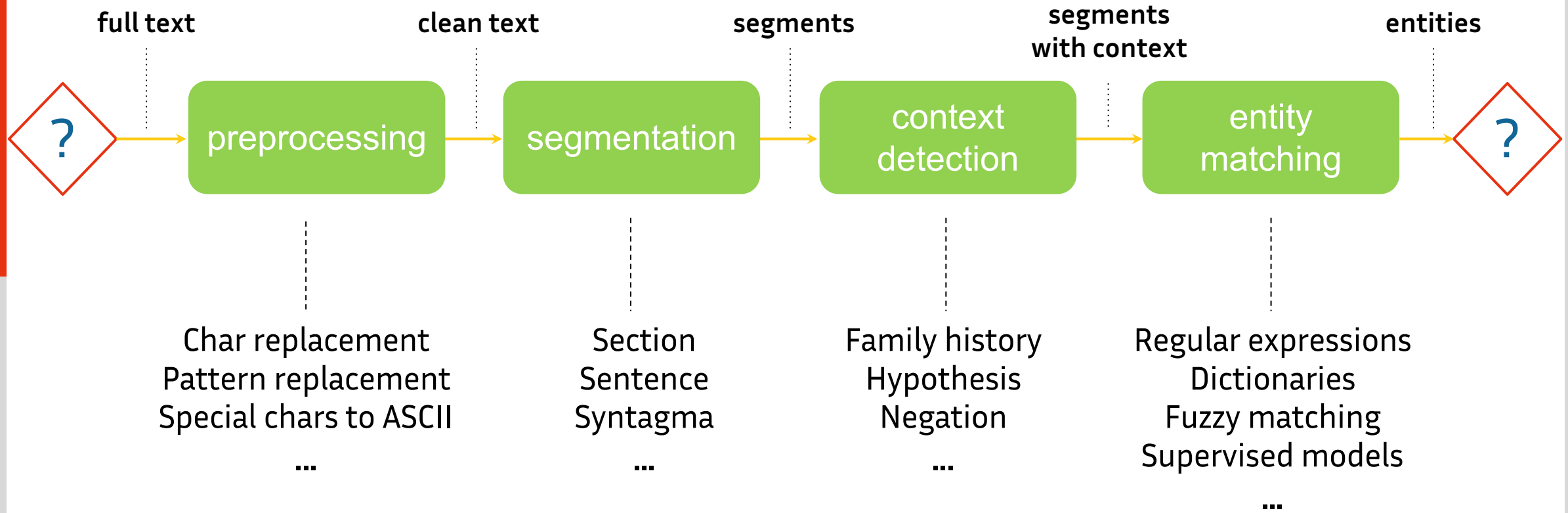
*False
positive*

*Past?
Stopped?*

*False
negative*



Example of a NER pipeline



medkit

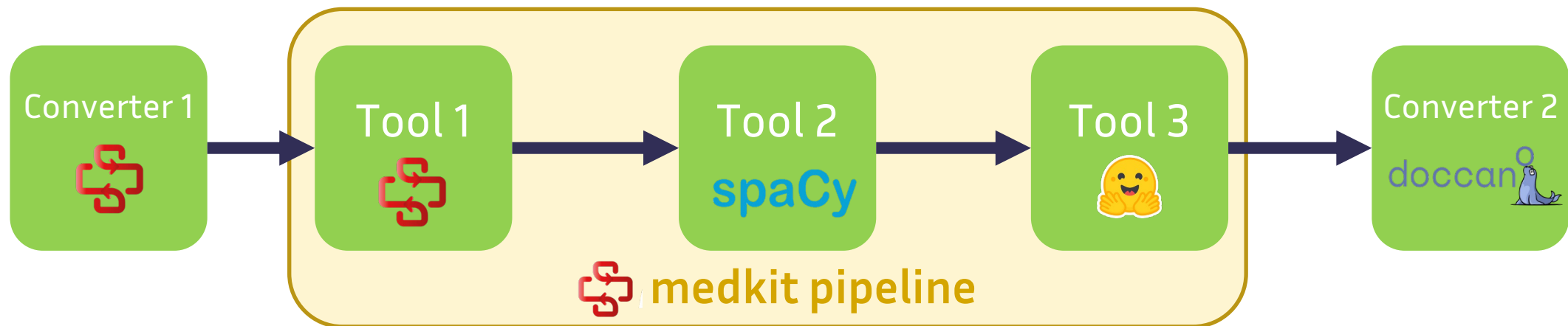
Encapsulate and chain tools
to facilitate reuse and comparison

On which data?

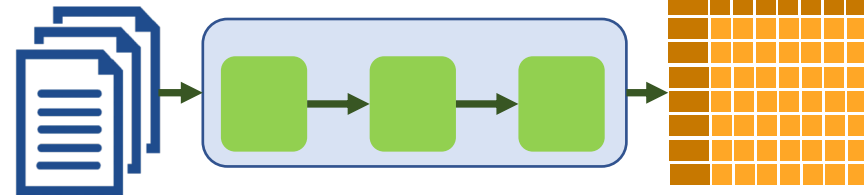
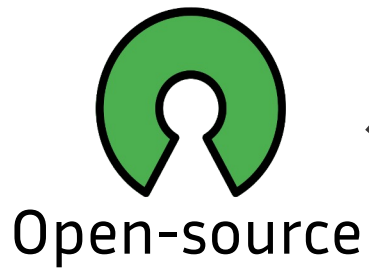
Text and audio for now, more to come

What tools?

- Preprocessing
- Extraction
- Classification
- Evaluation ...



Reuse, reuse, reuse



A rectangular box containing a screenshot of a web application interface, a yellow 'blowing a kiss' emoji, the 'doccan' logo with a blue seal character, and the 'spaCy' logo in blue.

Integrations



- **Code:** <https://github.com/TeamHeka/medkit>
- **Doc:** <https://medkit.readthedocs.io/>
- **1st medkit workshop:** 9-10 Oct. at PSC
- **if interested in another workshop session:** medkit-maintainers@inria.fr

