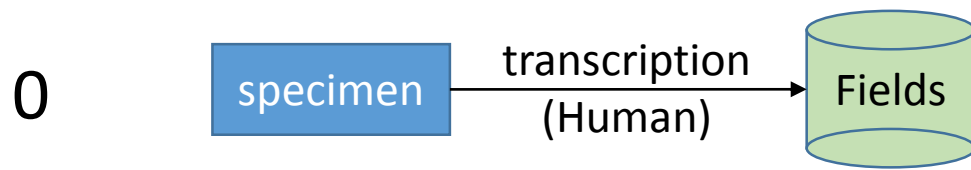


Hybrid workflows for knowledge extraction

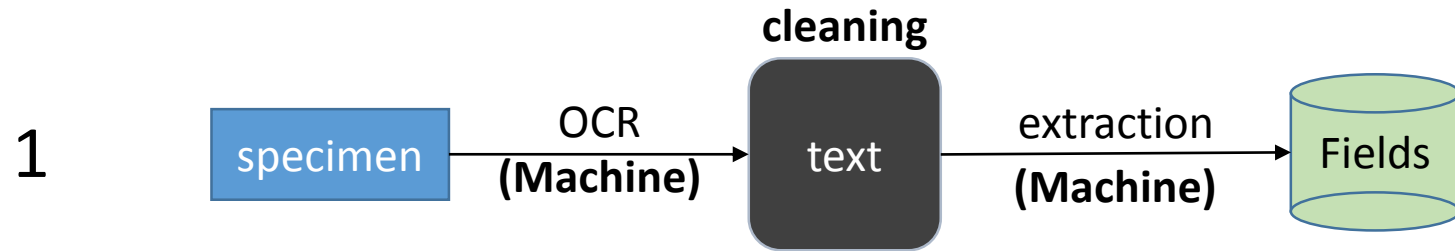
Hybrid (human- and machine-intelligent) workflows for knowledge extraction

Cases of study



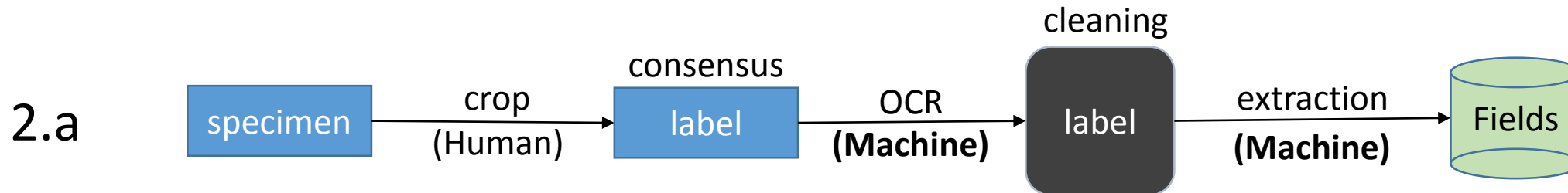
Only Crowdsourcing:

Users transcribe the fields from the image.
We have this data, but **we do not have the exact execution time**: 3 min. average

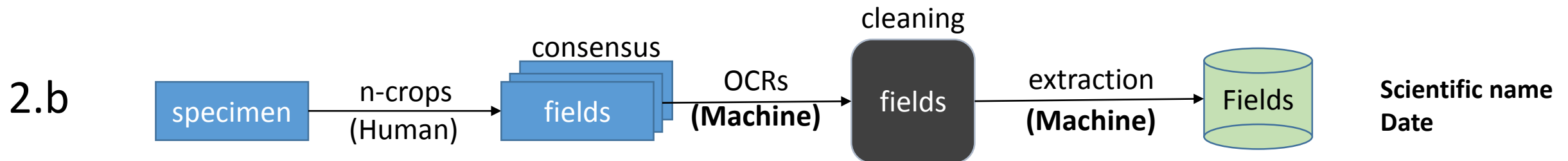


Only machine-intelligent processes:

Machine converts the image to text and extract the fields from it. **Not possible at this moment.**
Expected result: slower?, worse quality (and more expensive?) than 0.



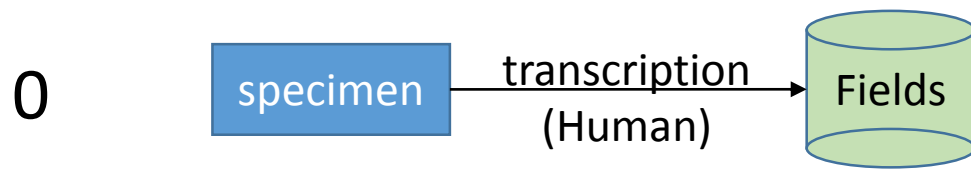
Users crop the label, which is converted to text (OCR), and ML algorithms get the fields from the text.
Expected result: Worse than 0. Better quality but slower than 1.



Scientific name
Date

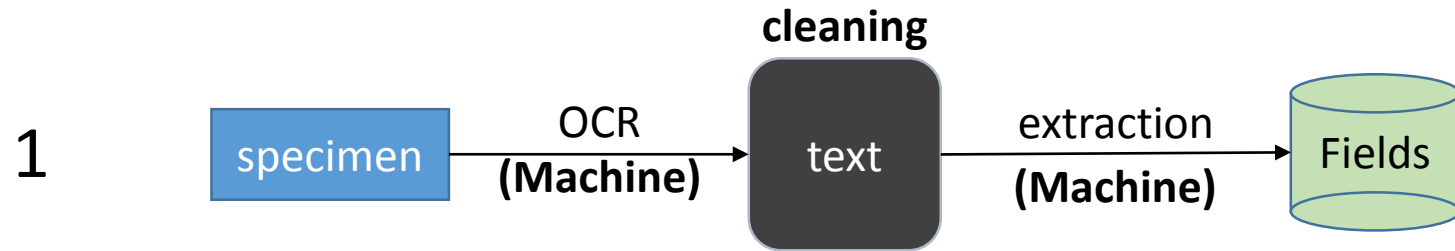
Users crop the label several times to get an image for each field, which are converted to text (OCR), and ML algorithms get the fields' values from text.
Expected result: Worse than 0. Better quality but much slower than 1. Better quality but slower than 2.a

Cases of study



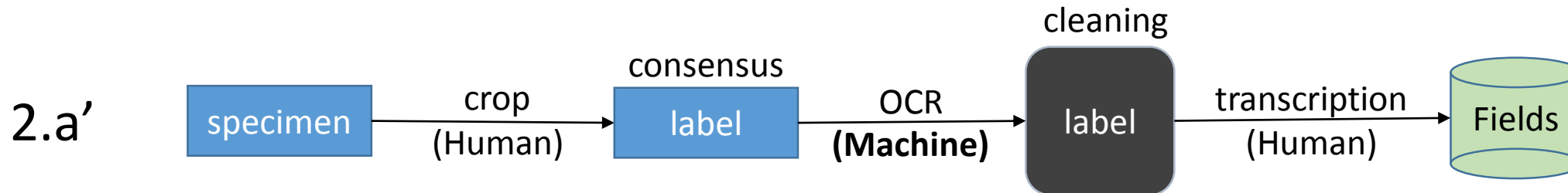
Only Crowdsourcing:

Users transcribe the fields from the image.
We have this data, but **we do not have the exact execution time**: 3 min. average



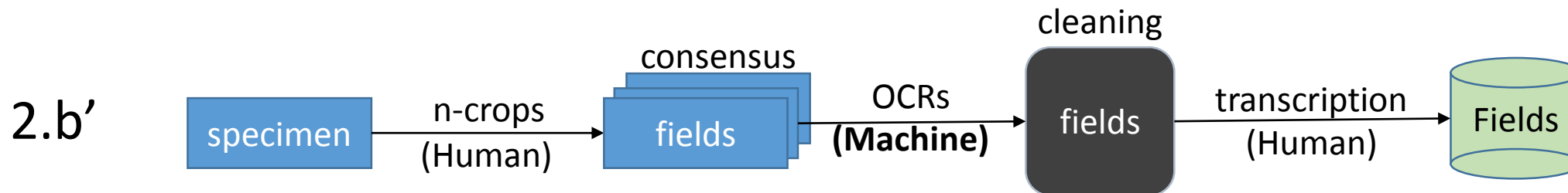
Only machine-intelligent processes:

Machine converts the image to text and extract the fields from it. **Not possible at this moment.**
Expected result: slower?, worse quality (and more expensive?) than 0.



Users crop the label, which is converted to text (OCR), and users transcribe the fields from the text.

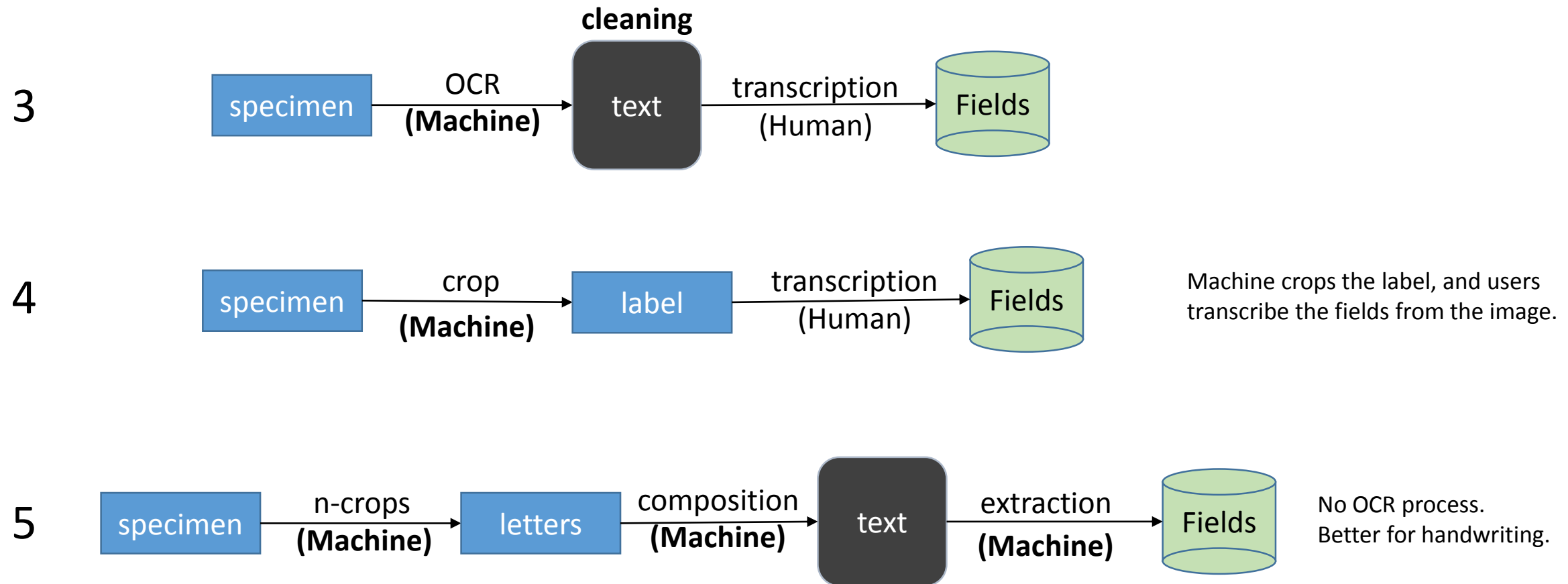
Expected result: Worse than 0. Much better quality but slower than 1.



Users crop the label several times to get an images for each field, which are converted to text (OCR), and users transcribe these fields from the text.

Expected result: Worse than 0. Much better quality but much slower than 1. Better quality but slower than 2.a

Other cases of study



Other ideas:

- i) can machine detect from 2.a or 2.b when it is doing a bad job, to then send it to crowdsourcing only the pieces it is failing to get?
- ii) can machine learn the relative positions of different fields to improve NLP?