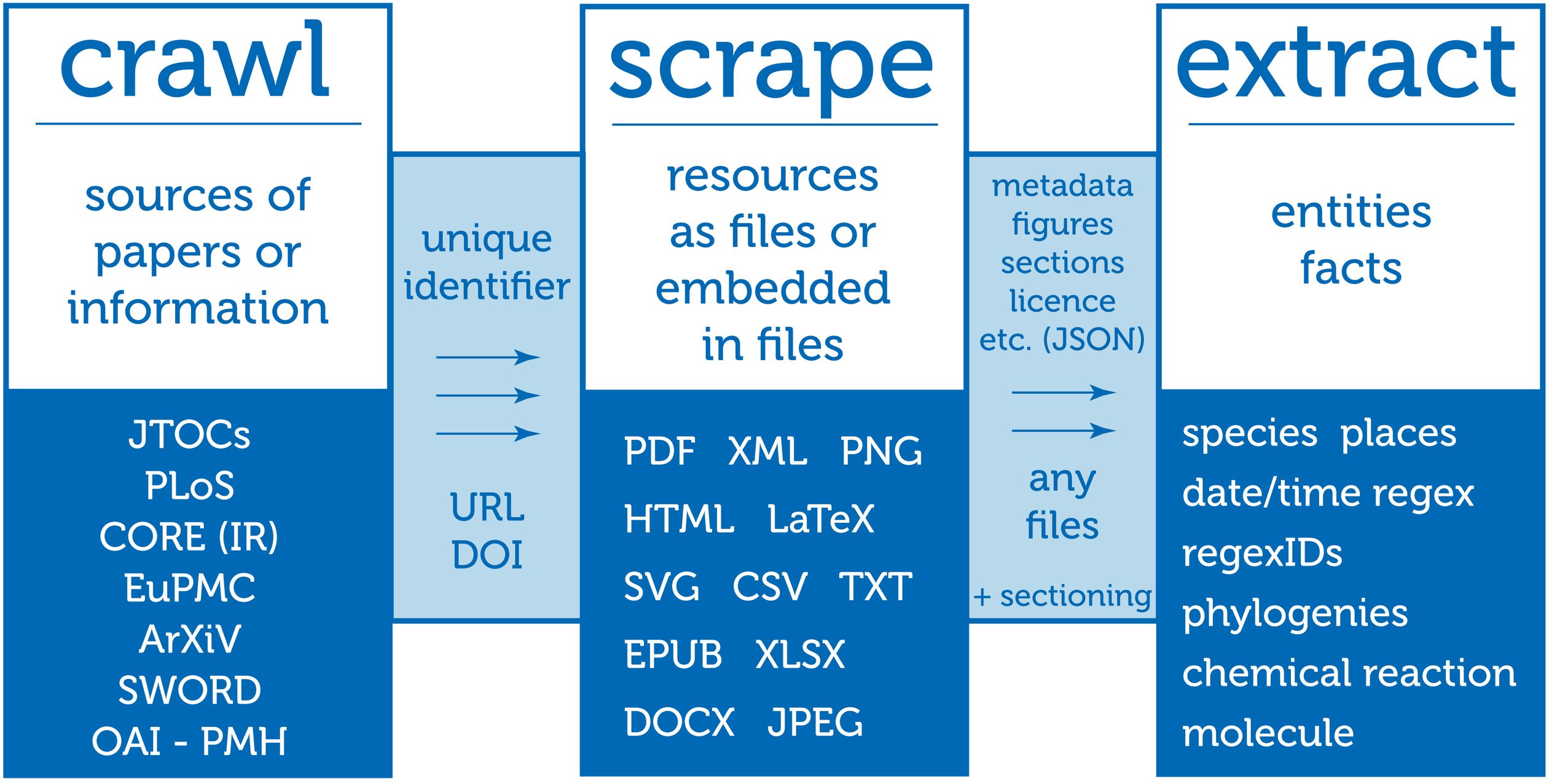
crawl
sources of papers or information
JTOCs
PLoS
CORE (IR)
EuPMC
ArXiV
SWORD
OAI - PMH
unique identifier
URL
DOI
scrape
resources as files or embedded in files
PDF XML PNG
HTML LaTeX
SVG CSV TXT
EPUB XLSX
DOCX JPEG
metadata
figures
sections
licence
etc. (JSON)
any files
+ sectioning
extract
entities
facts
species places
date/time regex
regexIDs
phylogenies
chemical reaction
molecule