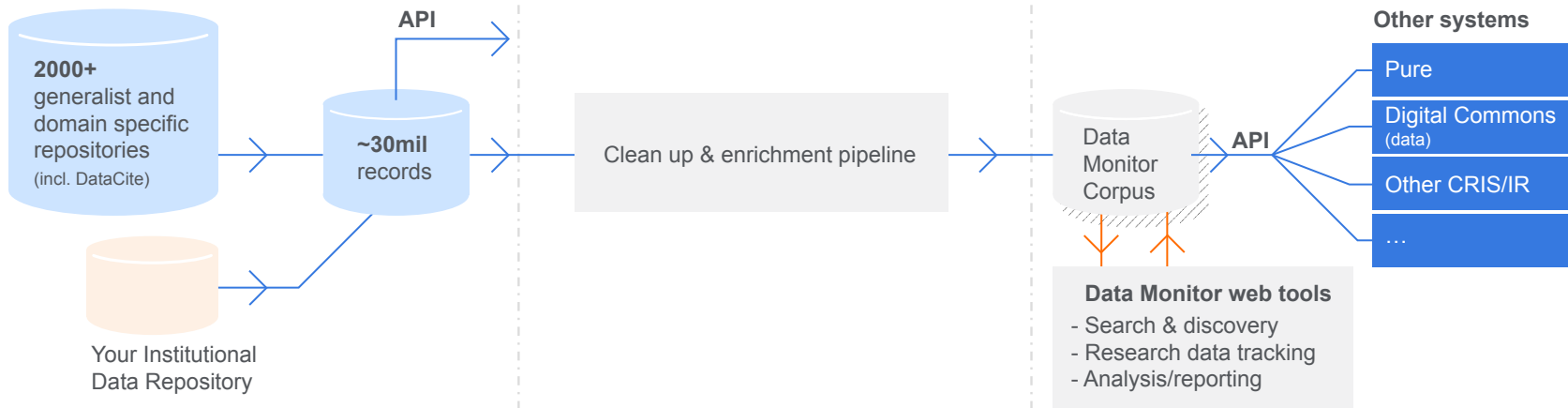


# How Data Monitor works



## Step 1

- **Harvest** metadata from 2000+ public repositories
- **Normalize** the metadata, following the OpenAIRE metadata schema
- **Accessible** via a free API

## Steps 2

- **Remove** non-reliable/non-data sources, duplicates, datasets with dead links, non-data/SPAM records
- **Enrich** metadata with publications/author/institutional linking and more by triangulating other sources (DataCite, Scholix, Scopus) + text mining and NLP

## Step 3

Data Monitor Corpus is **accessible** via the Data Monitor API (for integration) and Web Tools