

# CEPLAS project registry

---

## Idea

Collect and share high-level metadata (bibliographical metadata, project idea, nothing sensitive) of all projects going on in [CEPLAS](#) with and for CEPLAS members (only).

## Basis

- Developments in [DataPLANT](#)
  - Platform: [DataPLANT GitLab](#)
    - Restricted user group: CEPLAS members
  - Data structure: [ARC](#) (i.e. structured git repositories)
  - Metadata schema: [ISA](#) (serialized as ISA-tab)
    - selected ontologies to categorize contributor roles, study/experiment types and outputs

## Approach

### Data collection

1. Let CEPLAS volunteers describe their project in the ISA-tab `isa.investigation.xlsx` Excel workbook.
  - More GDPR safe alternative: build dummies based on published manuscripts. Could also be a nice use-case for a retrospective project data base.
2. Store data in ARCs on DataPLANT GitLab

### Data wrangling

3. Crawl ARCs via GitLab API
  - Get info from `isa.investigation.xlsx` only
4. Feed information into database (simple container or table)
  - `bash` / `python` / `R`

### Data dissemination

5. Build a locally running (double-safe) app or interactive doc (e.g. shiny doc in R) based on database that allows searching for values in specific or any fields of the isa investigation.

## Possible hurdles

1. Voluntary input required
2. Writing metadata in isa investigation is not really fun and we actually wanted to avoid asking people to do so.

## Alternative:

Automated publication crawler -> isa.investigation filler