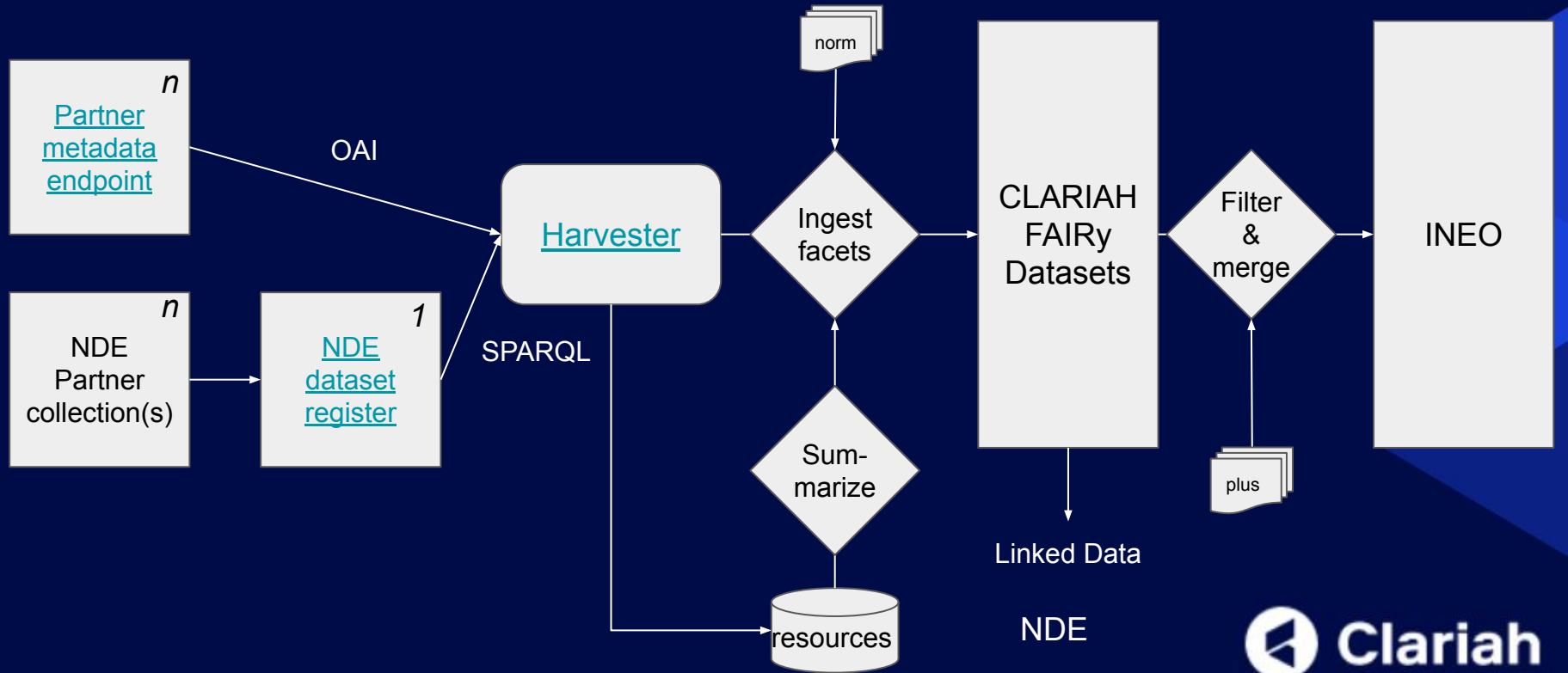


# FAIR Datasets, Vocabularies & Data Stories

Shared Development Roadmap  
Q2-demo 2022

# FAIR Datasets



# Facets for FAIR Datasets

- Current descriptive set  $\cong$  INEO set + FAIR facets
- FAIR Implementation Profiles
  - Which FAIR facets do we need?
    - F: PID types/links
    - F: Metadata formats
    - F: search engines for (meta)data
    - A: communication protocol for (meta)data
    - A: AAI for (meta)data
    - A: longevity for metadata
    - I: knowledge representation for (meta)data
    - I: vocabularies for (meta)data
    - I: schemas for (meta)data
    - R: license for (meta)data
    - R: provenance for (meta)data
  - Leading to F-A-I-R “stars”, with “enough stars” the dataset can potentially go to INEO
- F-UJI
  - Assess the FAIRness of our datasets?
  - Technical hints on how to make our FAIRness explicit
  - How can the registry help the partners repository?

# Why do we use CMDI as a pivot format?

- CMDI is an XML-based format where you can combine reusable components into resource type specific profiles
- It's in use by CLARIN and 7 of our partners provide their records as CMDI
- CMDI can easily morph to mimic other metadata formats, e.g. [the DCat of NDE](#)
- There is software for a central catalogue, i.e. [the VLO](#), including an ingest and normalization pipeline
- Facets can be adapted into the FAIR dataset facets we need
- Concept Links provide a bridge to the Linked Data world
- Lots of knowledge and tools in our community
- CMDI is not a goal/endpoint/result just a (convenient) mean to deal with the heterogeneity in our community

# What's new?

- Harvester
  - Speak SPARQL to the NDE dataset registry to get the records and turn them into CMDI
- Resource Summarizer
  - Find used Schemas and Vocabularies
  - Also for closed data? e.g. run internally & provide reports for facets (no sensitive data)
  - On the roadmap for LD, but to be extended for more resource types
- Registry endpoints
  - OAI for the full records
  - LD for the facets
- INEO sideloader
  - Merge current metadata with static descriptive info, e.g. screenshots
  - Editor for the descriptive info

# FAIR Vocabularies

- To be FAIR Datasets need to make explicit their
  - Knowledge representation
  - Vocabularies
  - Schemas
  - We use a looser LD view on vocabularies, which includes all of these
    - However, still with a LD focus, which we should try to loose now or in SSHOC.NL
- CLARIAH helps to
  - Find vocabularies recommended by the community
    - Positive or negative reviews
    - Crawled from a repository/registry, e.g.
    - Popularity with known datasets, as assessed by the summarizer
  - Retain access to vocabulary versions by caching
  - Access vocabularies in a common way by a common API
  - Crosswalk between vocabularies by linksets
  - Alternative hierarchical overlays for vocabularies
- We did trials with [a local copy of Bartoc](#) filled by crawls from [Awesome Humanities](#) and [YALC](#)
  - Now working on a better suitable data model for our own registry

# Data Stories

1. WP4: [The 'Spanish' Flu Epidemic in The Netherlands 1918-19](#)
  - a. Hosted by DRUID
2. WP5: [15 years of the popular Dutch chat show 'DWDD' - in data](#)
  - a. Hosted by the MediaSuite

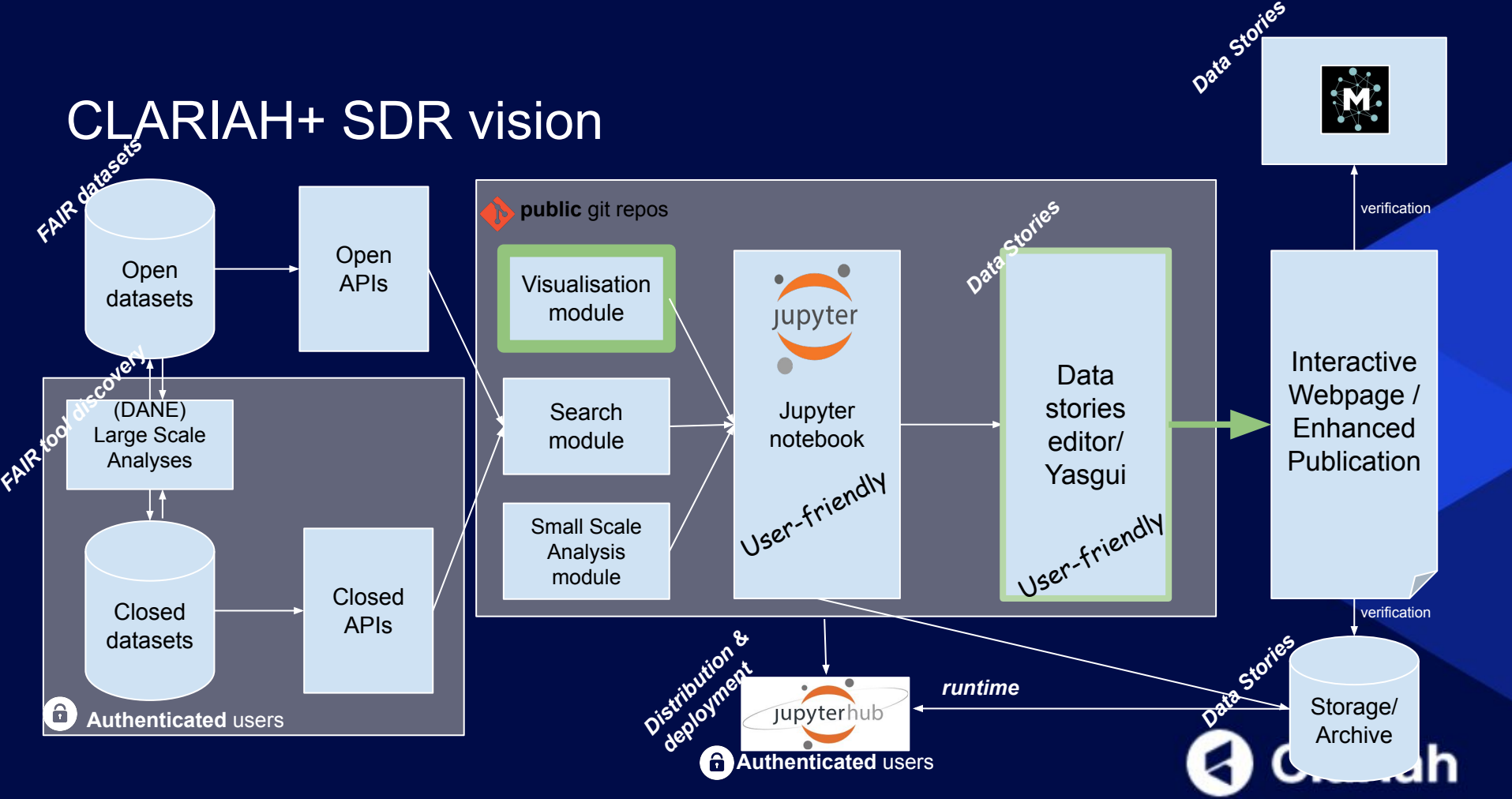
Merged their workflows into one

Wrote a [spec/schema](#) for the Data Stories, incl. [sample instances](#)

Started on the requirements for/design of the editor

Started on open source visualisation plugins, e.g. [geo](#) and [superset](#)

# CLARIAH+ SDR vision







## My Data Story

Metadata
  Notes & commentaries
  Provenance
  Edit view
  Preview

TEXT

## Cultural objectivism and capitalist narrative

The subject is contextualised into a Marxist capitalism that includes language as a totality. In a sense, the futility, and eventually the defining characteristic, of capitalist narrative depicted in Stone's \*Natural Born Killers\* emerges again in \*Heaven and Earth\*.

Derrida promotes the use of cultural objectivism to challenge archaic perceptions of class. Thus, in \*JFK\*, Stone examines Marxist capitalism; in \*Natural Born Killers\*, although, he reiterates cultural objectivism.

### Cultural objectivism and capitalist narrative

The subject is contextualised into a Marxist capitalism that includes language as a totality. In a sense, the futility, and eventually the defining characteristic, of capitalist narrative depicted in Stone's *Natural Born Killers* emerges again in *Heaven and Earth*.

Derrida promotes the use of cultural objectivism to challenge archaic perceptions of class. Thus, in *JFK*, Stone examines Marxist capitalism; in *Natural Born Killers*, although, he reiterates cultural objectivism.

Add block below
  Move this block
  Collapse block

**TEXT**

Markdown

HTML

WYSISYG

Wiki

**MEDIA**

Image

Sound file

Movie

**QUERY**

SQL (Postgresql)

SQL (MySQL)

SPARQL

GraphQL

Metadata
  Notes & commentaries
  Provenance

QUERY

Query

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 SELECT * WHERE {
4   ?sub ?pred ?obj .
5 } LIMIT 10
  
```

