TEXTCROWD
Collaborative semantic enrichment of text-based datasets

Franco Niccolucci & Achille Felicetti

EOSCpilot
www.eoscpilot.eu

Cloud-based Service
Text Documents Analysis
Cultural Heritage & Humanities
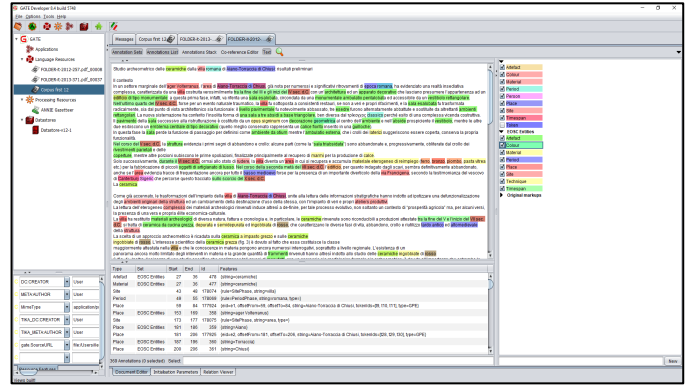Natural Language Processing
Semantic Enrichment

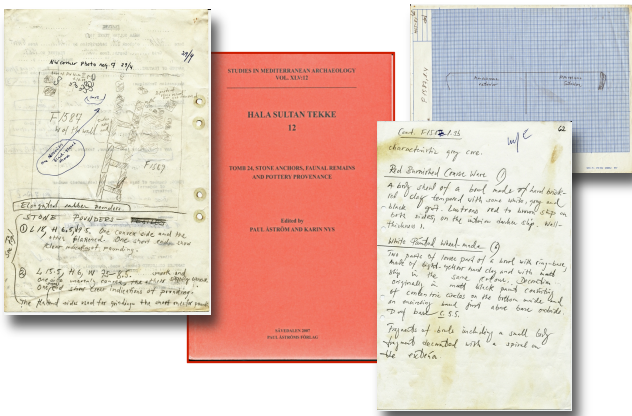VAST-LAB

PIN POLO UNIVERSITARIO CITTÀ DI PRATO
SERVIZI DIDATTICI E SCIENTIFICI PER L'UNIVERSITÀ DI FIRENZE

TEXTCROWD is an advanced cloud based tool for processing textual archaeological reports. The tool has been boosted and made capable of browsing big online knowledge repositories, educating itself on demand and used for producing semantic metadata ready to be integrated with information coming from different domains, to establish an advanced machine learning scenario

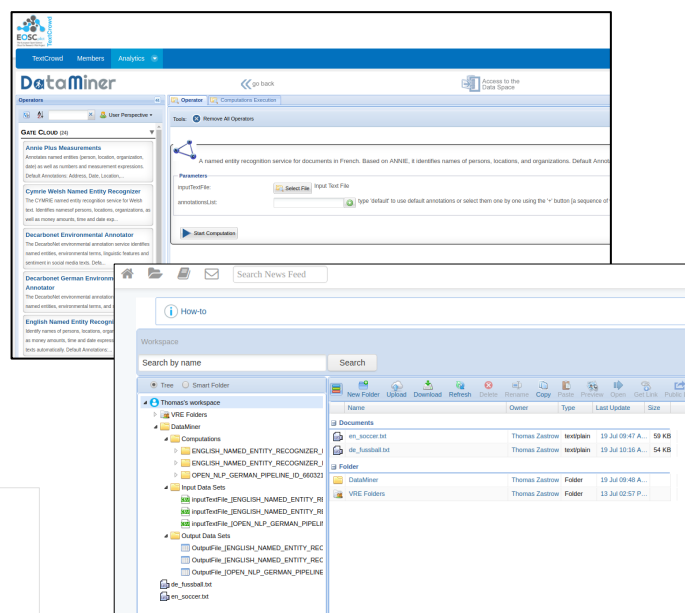## OUR PILOT

CONTEXT

TEXTUAL DOCUMENTATION from Cultural Heritage and Humanities, including:

- Reports
- Archaeology: excavations, surveys
- Conservation: diagnosis, restoration
- Grey literature
- Literary/historical sources
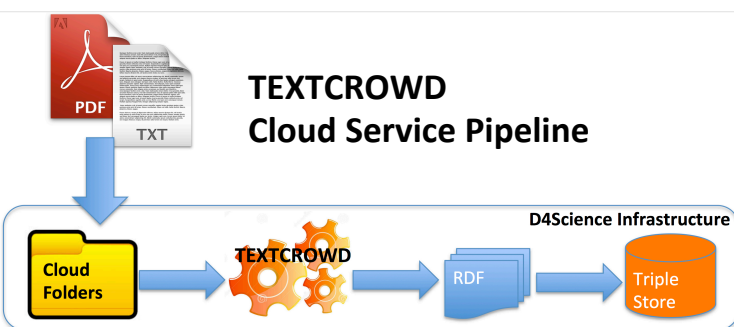- Research articles
- Monographs

EVERYTHING HAPPENS IN THE CLOUD
The TEXTCROWD service, empowered by the D4Science infrastructure, allows users to upload and store textual documents in a personal cloud folder, perform NLP and NER operations, trigger the semantic enrichment process and get CIDOC CRM information in RDF. Results can be uploaded in a triple store or in another semantic enabled system and reused within the same context or in another VRE scenario on the same cloud.

CLOUD BASED SERVICE

https://textcrowd.d4science.org

D4SCIENCE INFRASTRUCTURE

TEXTCROWD Cloud Service Pipeline

Cloud Folders → TEXTCROWD → RDF → Triple Store
D4Science Infrastructure

© 2017 EOSCPilot.eu – EOSCPilot.eu has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no 739563. The content of this dissemination material does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.