intelcomp

Artificial Intelligence Living Lab (SEDIA) Workshop
The STI Viewer Tool

September 27, 2023

I. A reminder: The IntelComp platform

- A platform to improve and automatise the tasks of managers of R&D programs at ministries (e.g. SEDIA) and funding agencies, by exploiting millions of documents with text analytics.

- Digital records of all STI activities & Automatic text analysis -> Novel insights from the data.

- The platform has been **successfully tested** by Public Administrations in three use cases (through Living Labs):

  – **R&D in Artificial Intelligence (SEDIA - Spain)**

  – Cancer Research (France)

  – R&D in Agrifood and Energy (Greece)
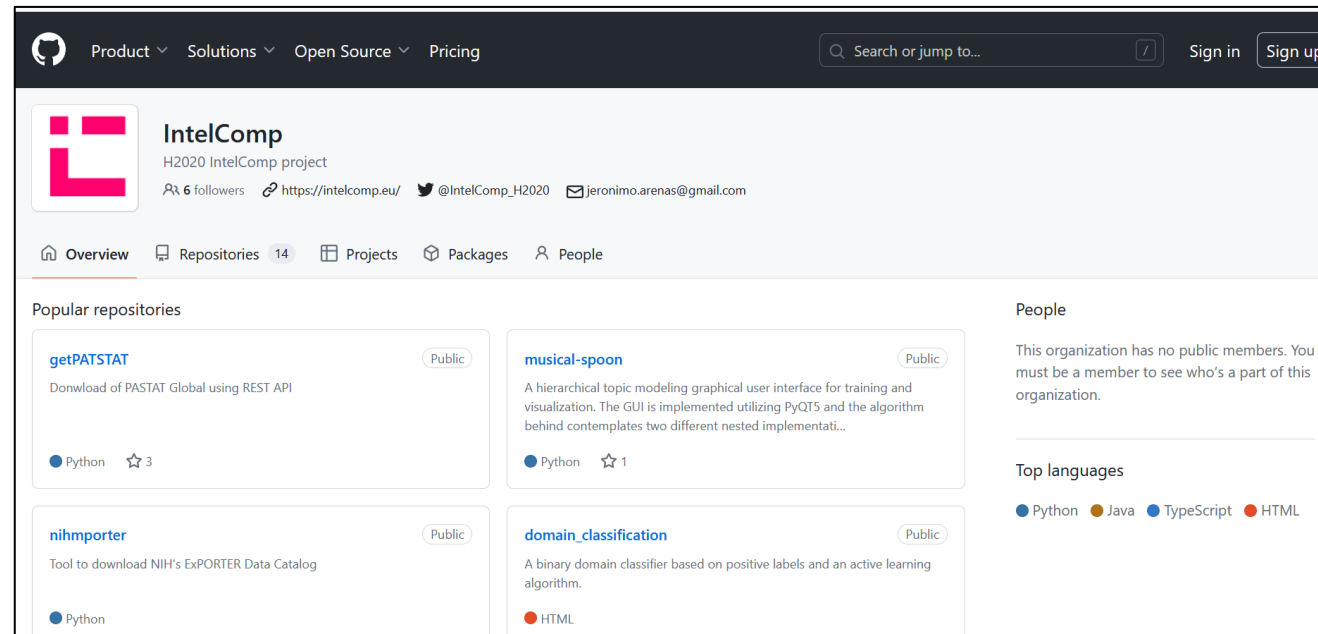
II. Our services and tools in production

- The platform already provides **solutions to ingest data, pre-process the texts and translate** millions of documents.

- We've also developed a collection of services that add value to the text data:

    1. A **domain classifier** to identify the documents in a domain.

    2. **Automatic classifiers** to assign documents to well-known taxonomies.

    3. **Topic modeling services** to extract the main themes in a collection of documents.

    4. Several software components for the computation, processing and analysis of **graph collections** - sets of nodes (e.g. projects, publications) connected by edges (that represent some kind of closeness between the nodes).

III. Our services and tools in production (cont.)

- Many of our services (NLP Preprocessing, Automatic Translation, Domain Classifier or Topic Modeling) benefit from **Large Language Models** running over GPUs.

- These text analysis services are integrated into the **Interactive Model Trainer**. This is the heart of the platform, where we train models incorporating information from experts in a domain. These models nurture the end-user tools of the platform.

- One of these tools (already in production) is the **STI Viewer**. With the models of the Interactive Model Trainer we calculate a lot of indicators that we show in the STI Viewer (TODAY, indicators in R&D in AI).

IV. Our services and tools in production (cont.)

- Most of our services are already available in **github.com/IntelCompH2020 + scientific publications + public deliverables**

V. Next steps

1. **Complete the development and test** the other three end-user tools:

   – The **STI Policy Participation Portal**, empowered by the STI Viewer and survey tools, that allows policy stakeholders (citizen associations, academia, industry) to provide feedback on STI policy.

   – The **Evaluation Workbench** tool. It assists call managers in the evaluation of grant applications.

   – The **Graph Visualizer** where we can analyze large collections of graphs - documents and their metadata.

VI. Next steps (cont.)

2. **Finish the integration and deployment of all services and tools** at the **Barcelona SuperComputing Center**. We'll have a common, flexible, incremental and highly scalable (more than 200 million documents) architecture for all possible use cases in STI policy.
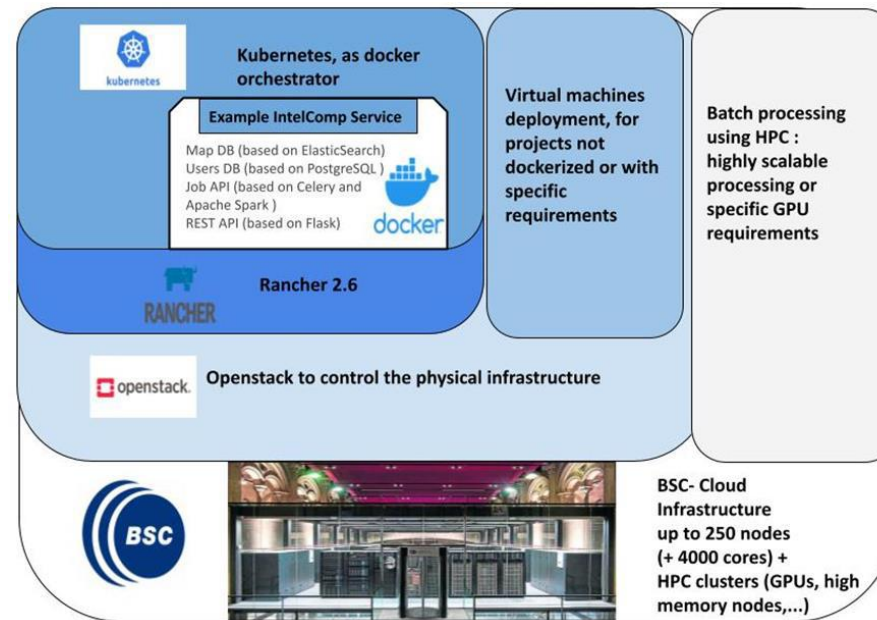
3. **Define business models** to exploit the platform.

The STI Viewer Tool
Main Functionalities

Joseba Sanmartín (Spanish Foundation for Science and Technology – FECYT)

September 27, 2023

AI Living Lab (SEDIA) Workshop: The STI Viewer Tool

1. What is the STI Viewer tool?

- The STI Viewer is an interactive tool **to visualize and explore a wide range of indicators on research and innovation activities**.

- It features **Business Intelligence dashboards** showing a set of indicators that can be "sliced" across multiple dimensions (for instance, scientific or technology domain or geographical area).

- It presents complex data in an accessible and **user-friendly manner**.

2. STI Viewer. Users and functionalities

- <u>Users</u>: **Policy analysts at ministries (e.g. SEDIA) and funding agencies** managing programs or calls that support R&D activities.

- The tool can be used for several <u>tasks</u>:

  - **To analyze the context and the problem that justifies a program**. With the STI Viewer we can identify and understand trends in R&D activities (e.g. to recognize the most promising research topics, the characteristics of firms using AI, etc.).

  - **To assist in the monitoring of the results of programs or calls**. With the STI Viewer we can use textual data to describe the outputs of the R&D projects.

3. Definition of the indicators in the STI Viewer

- The indicators stem from the **theoretical framework of the project** that combines the functions of the innovation system with the policy cycle. By crossing these dimensions, we formulated **questions** of interest to STI policymakers.

	Phase 1. Agenda setting	Phase 2. Policy formulation	Phase 3. Policy adoption	Phase 4. Policy implementation and monitoring	Phase 5. Evaluation
Function 1. Entrepreneurial activity					
Function 2. Knowledge creation					
Function 3. Knowledge diffusion through networks					
Function 4. Guidance (creating legitimacy for stakeholders, visibility and clarity)					
Function 5. Market formation (create markets through regulation of incentives)					
Function 6. Human and financial Resources mobilisation					
Function 7. Creation of legitimacy for society/counteract resistance to change					

- We focused on the questions in the **agenda setting** and **evaluation** stages of the policy cycle, and conceived a long list of **indicators** for responding to those questions.

intelcomp

4. Definition of the indicators in the STI Viewer (cont.)

Does it require text analytics?

Could we source sufficient data?

Is it technically feasible?

Is it in scope as per the proposal?

Is it in scope as per the domains (R&D in AI) and later Living Labs (SEDIA)?

5. The STI Viewer workflow