INTELCOMP PROJECT

A COMPETITIVE INTELLIGENCE CLOUD/HPC PLATFORM FOR AI-BASED STI POLICY MAKING

(GRANT AGREEMENT NUMBER 101004870)

Report on Health Living Lab
D6.4

Deliverable information	
Deliverable number and name	**D6.4 Report on Health Living Lab**
Due date	December 31, 2023
Delivery date	May 16, 2024
Work Package	WP6
Lead Partner for deliverable	Hcéres
Author	Tessa Enock Levi (Hcéres)
Reviewers	Lydia Papadaki (ARC) Paresa Markianidou (Technopolis) Dietmar Lamper (ZSI)
Approved by	Jerónimo Arenas, Technical Coordinator (UC3M) Joseba Sanmartín (FECYT)
Dissemination level	Public
Version	3.3

Table 1. Document revision history

Issue Date	Version	Comments
29 Sept. 2023	v0.1	Initial document structure created
13 Oct.2023	v0.2	Introduction and methodology added
30 Nov. 2023	v0.3	Added main
20 Dec. 2023	v1.5	Ready for review
22 Dec. 2023	v2.1	Review feedback incorporated
24 Dec. 2023	v2.2	Final version approved by the Technical Manager
16 May 2024	v3.3	Under Acronyms, a List of Stakeholders in the Health Living Lab and their acronyms has been added. Subsection 4.1 has been renamed and updated to summarize with whom, how and when the Health Living Lab operated. Figure 7 has been added to show the final detailed timeline of the Living Lab activities. Subsection 4.2 has been renamed and updated to further document the activities in response to the stakeholders' requests and to showcase the innovative results of the Living Lab. Figures 13, 14 and 15 have been added to show screenshots of different STI Viewer views as a tangible outcome of the Living Lab process. Subsection 4.3 has been updated to demonstrate evidence of stakeholder involvement and details of the Living Lab external events: agendas, lists of participants and links to events' webpages with presentations. Figure 21 has been added to show screenshots of different STI Viewer views as a tangible outcome of the Living Lab process. An annex has been added to present some materials shared with stakeholders at the kick-off event of the Living Lab.

DISCLAIMER

This document contains descriptions of the **IntelComp** project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium coordinator for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The content of this publication is the sole responsibility of **IntelComp** consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors.

(http://europa.eu.int/)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004870.

CONTENTS

FIGURES

TABLES

ACRONYMS

AI — Artificial Intelligence

CC — Climate Change

LL — Living Lab(s)

PA — Public Administrations

SME — Small and Medium Enterprise

STI — Science, Technology, and Innovation

List of IntelComp Partners and their acronyms

Acronym	Full name
ARC	ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS
BSC	BARCELONA SUPERCOMPUTING CENTER
CITE	COMMUNICATION & INFORMATION TECHNOLOGIES EXPERTS ANONYMOS ETAIREIA SYMVOULEFTIKON KAI ANAPTYXIAKON YPIRESION
FECYT	FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA
Hcéres	HAUT CONSEIL DE L'EVALUATION DE LA RECHERCHE ET DE L'ENSEIGNEMENT SUPERIEUR
HFRI	ELLINIKO IDRYMA EREVNAS KAI KAINOTOMIAS
NTTD	NTT DATA
OPENAIRE	OPENAIRE MAKE
SEDIA	SECRETARIA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL - MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
TGB	TECHNOPOLIS CONSULTING GROUP BELGIUM
TILDE	TILDE SIA
UC3M	UNIVERSIDAD CARLOS III DE MADRID
ZSI	ZENTRUM FÜR SOZIALE INNOVATION GMBH / CENTRE FOR SOCIAL INNOVATION

List of Stakeholders in the Health Living Lab and their acronyms

Acronym	Full name
AFM	AFM TÉLÉTHON (patients' organisation fighting neuromuscular diseases, rare genetic disorders that kill muscle after muscle; FR)
ANR	Agence National de la Recherche (FR)
EFS	Etablissement Français du sang (institution involved on research on blood cancer and strongly involved in reflection about programs evaluations; FR)
FRM	Foundation pour la Recherche Médicale (FR)
Fondation ARC	Fondation pour la recherche sur le cancer (Cancer Research Foundation; FR)
INCa	Institut National du Cancer (FR)
ITMO	Multi-Organism Institute, Aviesan Cancer (National Alliance for Life Sciences and Health (FR)

EXECUTIVE SUMMARY

The objective of the Horizon 2020 Innovation Action IntelComp project is to build a platform to analyse large volumes of textual data using Artificial Intelligence services. IntelComp adopts a Living Labs methodology. It involves public administrations and various stakeholders (i) to co-design new tools and services; and (ii) to validate the resulting platform through the co-creation of Science, Technology and Innovation (STI) policies in three different domains: artificial intelligence, climate change/energy and health/cancer.

This document constitutes the final report of IntelComp *Health Living Lab* (LL).

The LL aimed at contributing to the development of a suite of AI models and tools for analysing STI and validating STI policies. It operated through the proper exploitation of the IntelComp results and the creation of a data space containing both raw and processed data. These goals guided the LL planning and implementation, in addition to the main objectives stated above.

To meet those objectives and goals, the IntelComp LL followed a common methodological approach that has been tailored to the needs and context of the Health LL. This includes concrete goals, policy questions and data considerations, a stakeholder engagement strategy, an alignment with the technical development, and a roadmap to capture the implementation path towards the set goals.

The Health Living Lab focused on cancer. It was welcomed by cancer research funders as a contribution to the difficult analysis of the impact of the projects they fund. The tools proposed in IntelComp, and the very broad scope of data mobilised, have given rise to considerable expectations. An extended large group representing all main French funders on cancer has been involved in the Living Lab.

Implementing the LL in order to build tools meeting these expectations constituted a rich experiment:

- Data mobilised did not answer to all expectations, being limited to the traditional dimensions of other tools (projects, publications, patents); data relating to socio-economic impact were no addressed.
- Direct interaction between technical teams and end-users on tools prototypes was difficult with our target group of policymakers. The Hcéres interface proved necessary to adapt the tools and integrate data before presenting the tools to policymakers.
- Interaction between Hcéres team and the technical teams was necessary and highly instructive. IntelComp project was essentially driven by tools development. But as the project progressed, Hcéres managed to work closely with the technical teams in order to integrate upstream stakeholders' needs.
- During the final year of the project collaborations have been very constructive and brought useful conclusions. The analysis of available data on project and publications showed the difficulties in terms of linking the different types of data. The experiment of semantic matching tools for projects and publications opened a way to resolve this issue. Thematic analyses were a good opportunity to share innovative results with policymakers on cancer research.
- The stakeholders involved in the Health LL have been particularly interested in sharing Hcéres experience of this work with new AI tools. They were interested in understanding the problems encountered and the solutions proposed as these were relevant to their own context.

1. INTRODUCTION

The IntelComp project is a Horizon 2020 Innovation Action to build a platform to analyse large volumes of textual data using artificial intelligence services. IntelComp adopts a Living Labs methodology. It involves as primary stakeholder group *public administrations* and *policymakers*, as well as other relevant stakeholders groups (such as *civil society organisations*, *academia*, or *industry organisations*), to (i) co-design tools and services and (ii) validate the resulting platform through the co-creation of STI policies in three domains as specific use cases: artificial intelligence, climate change/energy and health/cancer.

This document captures the results of the IntelComp *Health Living Lab* and constitutes deliverable D6.4. The Living Lab (LL) was implemented from Q2/2022 to Q4/2023, based on a joint approach outlined by D6.1 whose purpose was to ensure that the envisioned LL objectives would be achieved.

The deliverable presents the main results and activities of the Health LL. It starts with this introduction to provide the background and plan at the outset of the LL activities. Following the methodology that each IntelComp LL adapted to its own purposes, the main part of the report comprises the key results in terms of LL activities, as well as implications on the domain of the LL and the technical development of the IntelComp tools. The final part of the deliverable concludes.

2. LIVING LAB GOALS

2.1. Project Goals

IntelComp has been devised to build a platform that can analyse large volumes of textual data using AI services. It adopts an LL methodology and involves external stakeholders[1] to co-create the envisioned tools and services, and to validate the resulting platform through the co-creation of Science, Technology and Innovation (STI) policies in three different domains: artificial intelligence, climate change, and cancer.

Apart from these overarching goals, further goals include the following:

- IntelComp platform shall be deployed in high performance computing environment;
- A suite of AI Models and tools for STI analysis shall be developed;
- The exploitation of the results shall be achieved through adequate use of communication and dissemination processes;
- A data space of raw and processed STI sources shall be created;
- The project strives to understand the challenges of STI policy-making;
- The project aims at analysing and validating STI policy models.

2.2. Goals of the Health Living Lab

The health care system is a dynamic sector moved by multifaceted and intense medical breakthroughs. It constitutes a major concern for public policies. Hospitals deliver care and are involved in medical research, thus contributing to the transformation of care. By translation of research findings into improvement in medical care, medical innovation plays a significant role and contributes to better health, greater life expectancy and improvements in quality of life. Health research policies have also an impact in terms of new business opportunities and increased attractiveness to the next generation for careers in research and the health sector.

The Health LL focused specifically on cancer research. Cancer is the second leading cause of death globally and has been responsible for an estimated 9.6 million deaths in 2018. Moreover, many live with cancer for long periods and it is important to consider the morbidity caused by cancer. The economic impact of cancer is significant and is increasing. Only 1 in 5 low- and middle-income countries have the necessary data to drive cancer policy.

Cancer is also a topical for the Health LL because:

- It's a broad issue, ranging from basic research to clinical research, with many recent innovations in treatments and diagnosis techniques;
- It's a specific axis of Horizon 2020 and Horizon Europe with a dedicated transversal mission;
- It involves various actors: public health authorities and public research actors, but also pharmaceutical industries, non-pharmaceutical industries, and patient associations;
- It relates to several public health issues: tobacco, alcohol, food, pollution.

[1] i.e. Public Administrations (PAs) and stakeholders from civil society organisations, academia, and industry/business organisations.

Numerous studies have been carried to describe and characterise cancer research by analysing scientific production through publications, patents and clinical trials. The Cancer LL focused on two objectives: on the one hand, help policymakers to link funded projects with scientific production and on the other hand, enable policymakers to characterise the medical and societal impact of cancer research.

IntelComp goal was to provide answers to these challenges by integrating relevant data on funded projects, scientific publications, patents, impact on medical practices as well as economic and social impacts.

The mobilisation of IntelComp's AI techniques is promising to analyse the congruence between scientific themes and the strategic orientations of research policies or expectations of civil society.

The goal was to monitor a wealth of indicators on four major pillars:

- Outputs: publications produced by program/funder, publications cited in patents, patents produced.
- Medical impact: publications cited in clinical guidelines, innovations in terms of diagnostic kits, treatments, drugs, new therapies, new companies/start-ups created, newly CE-marked medical devices or technologies.
- Economic impact by tracking innovation performance of companies (enterprises with evidence of innovation activities, number of newly Conformité européenne (devices and medical technologies bearing the CE label).
- Societal impact indicators in three dimensions: societal awareness/relevance of research, congruence of research funding with societal priorities and impact on public health

2.3. Software Development Goals

The goals of the software development overlap partly with those of the LL. For instance, the collaboration with the project's stakeholders and aligning their interests with the interests and capabilities of IntelComp.

The IntelComp software development considered many services and four main tools that are of particular relevance, as the LL participants will have the opportunity to use them. The first one, the *Interactive Model Trainer* is an expert tool to (a) train new topic models, (b) edit and curate topic models, (c) train new classification models, (d) generate sub-corpora, and (e) evaluate models. The other three *STI Viewer, STI Policy Participation Portal and Evaluation Workbench* are mainly geared towards fulfilling the needs of the primary stakeholder of the LL. Table 1 characterises the main features of these tools.

Table 1: IntelComp tools – their purpose for the primary Living Lab stakeholders

	STI Viewer	**STI Policy Participation Portal**	**Evaluation Workbench**
Targeted Organisation	Public administration (Ministry), funding agency	Ministry, funding agency, academic, business and citizen organisations	Funding Agency, Evaluation Agency (if independent of the Funding Agency)
Targeted users	Policy & STI analyst	Policy officer, STI managers/agents for organisations, citizens	Call Manager
Main functionality	Analyse, compare and visualise a comprehensive set of STI related KPIs	Provide a synthetic list of measurements for participatory STI policy making	Assist in the ex-ante evaluation of STI proposals for funding
Stage of the policy-making cycle	Agenda setting, monitoring and ex post evaluation	Agenda setting, monitoring and ex post evaluation	Implementation
Tool predecessor	Data4Impact	*(simplified)* STI Viewer	Corpus Viewer

The development timeline of these tools is provided in the *Platform Development Plan*. The LL planning accommodated that timeline as much as possible by aligning its stakeholder engagement activities with the development phases laid out in that plan.

3. METHODOLOGY

Living Labs, as a concept, have long existed[2] but have more recently become popular in all kinds of research and innovation projects, including in public administration research (cf. Decker, Contreras, and Meijer, 2020). In Europe, the concept has been further developed and adapted to the needs and setup of publicly funded projects (cf. Beaudoin at al., 2022; Compagnucci, Spirgarelli, Coelho, and Duarte, 2020).

3.1. What is a Living Lab?

Numerous definitions of LLs exist (Compagnucci, Spirgarelli, Coelho, and Duarte, 2020; pp. 3). Most definitions include key characteristics, namely the relation to real-life environments, the focus on stakeholders, on collaborative activities such as validation, experimentation, or testing. Sometimes, these are part of a co-creation approach. Another important characteristic is that LLs are facilitated, not managed, i.e. the team behind a LL has no authority over the lab's participants (cf. Westerlund and Leminen, 2011). Sustainability is yet another characteristic that is often crucial (cf. Leminen et al., 2016).

As a work definition, IntelComp's understanding of LL largely matches the definition offered by Schaffers and Turkama (2012): *A living lab provides a setting for collaborative innovation by offering a collaborative platform for research, development, and experimentation with product and service innovations in real-life contexts, based on specific methodologies and tools, and implemented through concrete innovation projects and community-building activities*.

3.2. Overall IntelComp Living Lab Approach

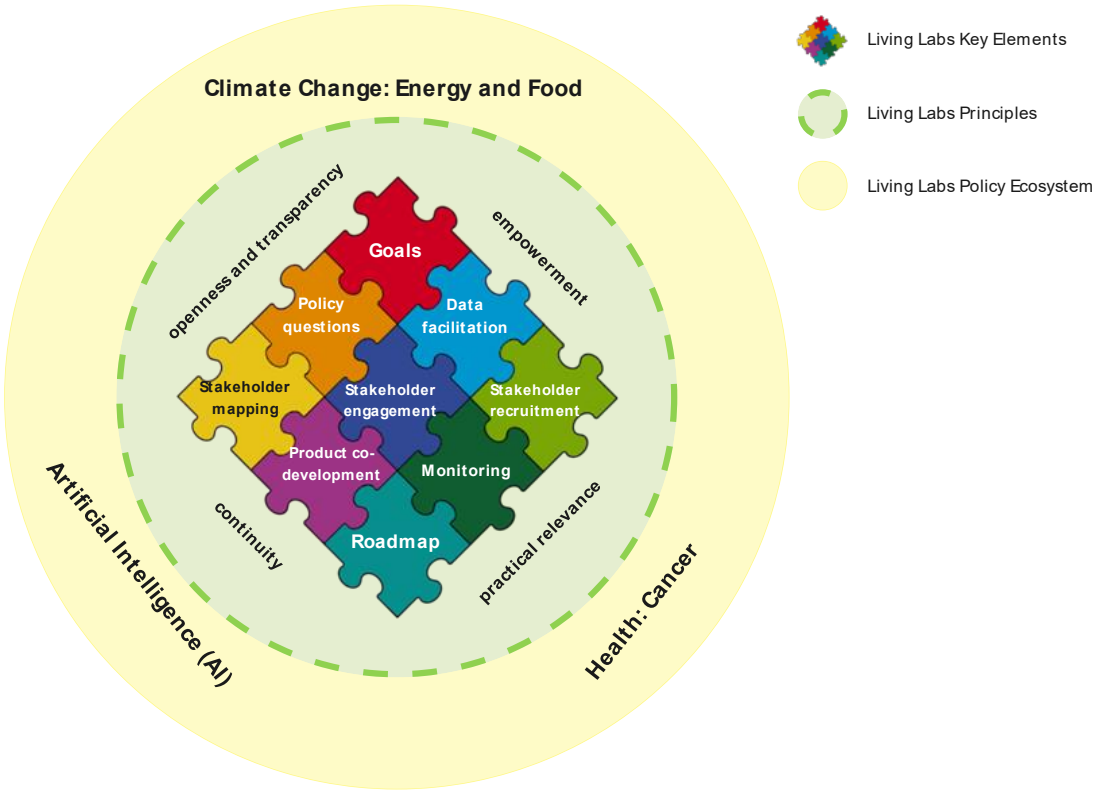
IntelComp largely followed the general approach of a LL but tweaked it such that it fit the project's setting. This includes its **policy ecosystem** which, as the figure below shows, comprises **AI, Cancer, and Climate Change**; moreover, it followed the following four **guiding principles**[3]:

- **Openness and transparency** – open to participation of many stakeholders; open to perspectives, needs, expertise, etc.; transparency regarding goals (no hidden agenda) and expected outcomes, decisions, limitations, and expectations;
- **Empowerment** – empowering LL participants by taking their inputs and contributions seriously, by enabling them to engage in the LL activities, and by helping them find answers to their (policy) questions;
- **Continuity** – continuous (mutual) learning; continuous fostering of relations between participants; and
- **Practical relevance** – relevance of activities, outputs, and results for LL participants in their real-life setting; relevance of results and outcomes for IntelComp.

[2] On the origin of the concept, cf. Eriksson, Niitamo, Kulkki, et al. (2005); Dutilleul, Birrer, and Mensink (2010); or Hossain, Leminen, and Westerlund (2019)

[3] Scholarly literature sometimes labels these differently and may include more such principles, but these are the ones that are most essential for the LLs foreseen by IntelComp.

Figure 1: Key elements and principles of living labs



3.3. Key elements of the IntelComp Living Lab

While the policy ecosystem provided the context and the principles to guide the IntelComp LL, it is the key elements that represented the building blocks of the LL. These key elements comprised the **goals**, specific **policy questions** and **data sources**, the **stakeholder** dimension (mapping, recruitment, engagement), the **co-development** of tools, the **implementation roadmap**, and the **monitoring** of the LL implementation (cf. figure above).

Although each LL tailored those key elements to their own needs, the common methodology outlines them as follows:

	In addition to the overall project goals, each LL set its own **individual goals** that it tried to realise during its lifetime. Hence, the planning and implementation of its key elements needed to be tailored to each LL.

	Each LL started with its set of domain-specific **policy questions**. During the LL preparation and planning phase, the LL teams provided key inputs to WP1 - *Evidence-based Policy Modelling* – which collected them and triangulated them with the policy framework (Deliverable 1.1). The final selection of the set of policy questions was done in WP1 by the technical teams and their assessment of the feasibility of measurements described D1.2. Those questions informed the scope of the work of the technical teams, from data sources to AI services to the user interfaces of the IntelComp tools. During the LL implementation, the initial set of policy questions was expanded and refined, depending on the needs and interests of the engaged stakeholders.

	Depending on and derived from the policy questions and indicators (also being developed by WP1), were the **data** that ought to be used, processed, and presented via the user tools. However, there was a different aspect to it, in that users may be given the means to upload their own data and possibly have them enriched and processed, to eventually use them in the given user tool. In the end, this turned out as not feasible, due to the technical complexity, which is why it was done solely on the basis of selected use cases.

The stakeholder dimension comprised three essential key elements of the LL[4]:

a) the mapping of potential stakeholders;
b) their recruitment as LL participants;
c) the ongoing stakeholder engagement to generate the envisioned goals and keep the stakeholders intellectually and emotionally linked to the LL.

a) the goal of the mapping was to identify a large enough group of stakeholders. This ensured that the ongoing participation in activities was adequate, that the results were robust, and that the burden caused by the ongoing engagement were made lighter by spreading efforts across different individuals. In practical terms, the mapping essentially prepared/collected data so that the answers to the following sentence could be determined for each potential stakeholder: *We want to recruit **whom**, **why**, **when**, **how**, and (if we don't have direct access) **by whom***.

b) the stakeholder recruitment was a concerted effort to activate suitable individuals – identified via the above-mentioned mapping – who committed to becoming involved in the LL activities, ideally regularly and throughout the lab's lifetime. While it was ultimately up to each participant to determine their own degree of involvement, the LL made a serious effort to keep their participants engaged, which leads to the next point;

c) the ongoing stakeholder engagement is the core activity of the LL (Mastelic, Sahakian, and Bonazzi, 2015), that assumes both a longer-term perspective to ensure that the LL as a whole continues to work towards its goals and a short-term perspective in that it focuses on the implementation of individual lab activities, such as workshops or trainings. It is important to note that it is easy to lose sight of the overall goals because the attention often lies on the next activities to be implemented, which is why the *LL implementation monitoring* was an integral part of the LL activities (more on this below).

[4] Note that the list ordered chronologically, which reflects the work of the creation of the initial, preliminary list of stakeholders; the illustration keeps the stakeholder engagement at the centre and is flanked by the two other activities, because it is most central to the LL activities, in terms of required efforts. Also, stakeholders will be recruited on a continuous basis, i.e. the chronological order plays a negligible role.

	Co-creation was in IntelComp's DNA and therefore one of the key elements of the LL: the co-development of the project's tools. One out of the four envisioned tools was expected to be fully co-created, the other three were being built on existing products but the basic idea was the same: the LL participants and potential users of the tools would have a big say in the development of those tools, i.e. the LL facilitators would listen to their needs, take their input seriously, transparently communicate decisions by the project partners that affect them, and in general live by the four guiding principles presented above. In practical terms, the co-creation process was closely tied to the technical development of IntelComp's tools and services, which is why the timeline laid out in the Platform Development Plan was an integral part of – and visually present in – the roadmap of each LL.

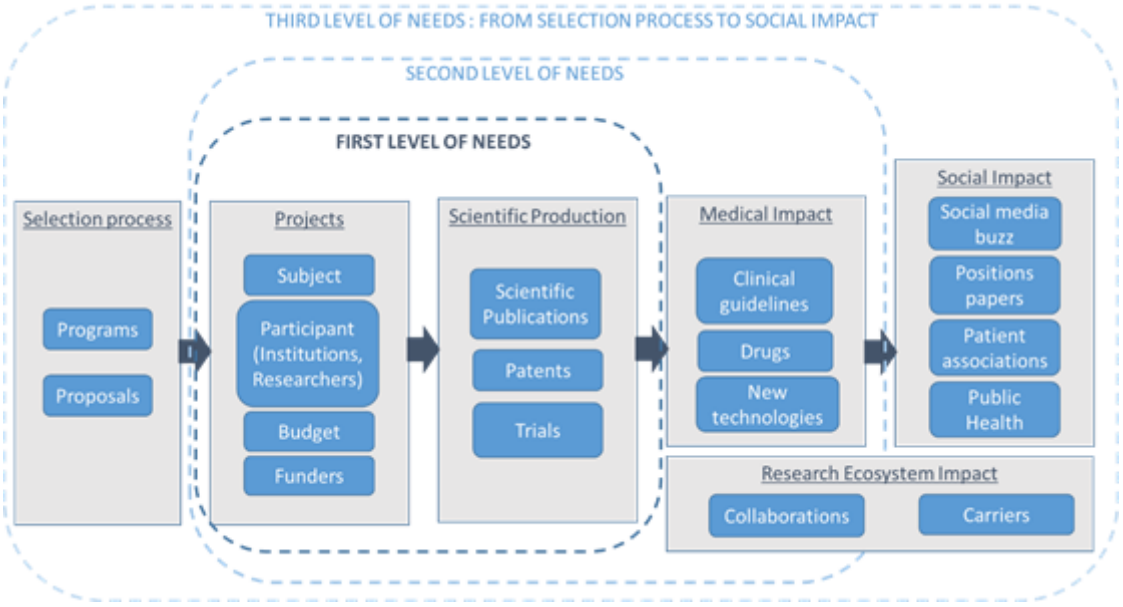
	LL Roadmap is a visual representation of the major events planned for the labs' implementation. In parallel, it shows how those are connected and, in fact, aligned with the development process. Each roadmap was tailored to its LL in terms of the number, timing, and scope of the events, as well as their target audience. The LL Roadmap served as a guideline for the implementation process, as well as with communicating that process to third parties.

3.4. Tailored Approach for Health Living Lab

3.4.1. Health Living Lab: Policy questions

The area of greatest interest to the potential stakeholders consulted by Hcéres was the analysis of the impact of funded research projects (or programmes or groups of projects) and the characterisation of 'impact pathways'. This involved monitoring and characterising the steps leading from project selection to different research results and then to their socio-economic impacts.

Figure 2: Health Living Lab – levels of needs



This approach made it possible to federate the expectations of various research stakeholders (funding agency, evaluation agency).

A first level of needs was to be able to characterise in a broad way the scientific production ("output") of funded projects in terms of:

- Scientific publications (as a common indicator of scientific production *stricto sensu*)
- Patents (as an indicator of technological production)
- Clinical trials (as an indicator of medical activity)

A second level of needs was to be able to identify and characterise the medical impact ("outcomes") of funded projects in terms of:

- Good practices (citations in clinical guidelines)
- New treatments (pharmaceutical industry)
- New diagnostic screening techniques (industrialists / start-ups)

The third and last level of needs was to be able to identify and characterise the social impact ("outcomes") of funded projects:

- Media impact (via the media & social networks)
- Topics of funded projects most often included in position papers
- Topics of funded projects corresponding to the expectations of patient organisations
- Positioning of projects in relation to public health data (incidence, mortality, quality of life of patients etc.)

Another dimension of impact raised by stakeholders was the impact of funded projects on the research ecosystem:

- Impact on the structuring of scientific communities

- Impact on the careers of young researchers funded

Related policy questions (cf. Table 2) explore this framework to address the impact pathways.

Table 2: Health Living Lab – policy questions corresponding to the analysis of impact pathways

Policy question	Sub-questions / Measurements
What is the production of scientific knowledge of funded projects	- How many scientific publications related to funded projects and details about discipline, type of research, type of cancer… - How many scientific publications related to funded projects compare to same kind of project (same discipline and same budget) - How many citations of publications related to funded projects compare to other publication (same discipline)
What is the technical production of funded projects	- How many patents related to funded projects (reference to the project or patent with a citation of a publication related to the project) with details by technologies - How many citations of publication related to the project in patent compare to other publications (same discipline)
What are the clinical trials related to the funded projects	- How many clinical trials funded by the project - How many clinical trials related to publications related to the project - Proportion of funded project related to clinical trials compare to other projects
What is the impact of funded projects on medical practices	How many medical guidelines citing the scientific publications related to the funded projects?
	How many new treatments related to the project
	How many new diagnostic screening techniques (firms / start-ups)
What is the importance of dissemination towards different audiences	- Dissemination toward scientific audience (OA, project event…) - Dissemination toward larger audience (out of academics) - "Connectivity" to society
Which societal challenges have been addressed by the selected project?	- Which societal challenges have been addressed by selected project (comparing topic in project's abstract and main to societal challenge) - Which societal challenges have been addressed by scientific production linked to selected projects? (comparing topic in publication's abstract and main societal challenge)
Which policy objectives have been addressed by the selected project?	- Which policy objectives have been addressed by selected project (comparing topic in project's abstract and main policy objectives) - Which policy objectives have been addressed by scientific production linked to selected projects? (comparing topic in publication's abstract and mains policy objectives)
What is the impact of project on researchers involved in the project	- Number of researchers involved in the project - Number of jobs created by the project - Careers of researchers involved in the project - Researcher communities created by the project

The analysis of the impact pathways was interesting to triangulate with the type of actors involved at each stage: public research actors, but also pharmaceutical companies (new treatments), engineering companies (diagnostic techniques), and patient associations.

Several thematic approaches related to cancer were interesting to study:

- Analysis by type of research (basic / clinical / translational)
- Identification of research cooperation (international cooperation, inter-institutions, public-private partnerships, etc.)
- Analysis by research discipline (e.g. epidemiology; social sciences)
- Focus on certain types of cancer (cancer location in particular) and prognosis (good / bad) or incidence (high/low)
- Possible focus on issues related to cancer: tobacco, alcohol, food, pollution, etc.
- Identification of new treatments and breakthrough technologies (genetics, biotherapies, predictive medicine, e-health)
- Characterise projects/work related to the different stages of patient care 1) prevention; 2) early detection; 3) diagnosis and treatment; and 4) quality of life for cancer patients and survivors.

3.4.2. Health Living Lab: Data facilitation

It was expected that the Health Living Lab could meet the challenge of these policy questions and help final users to characterise impact pathways of their funded projects both by the broad spectrum of data mobilised (cf. Table 3) and by the capacity of the tools to connect these data (in particular publications) with the funded projects.

Table 3: Health Living Lab – Main data sources identified to answer to policy-questions

Object		Data
Policy	Policy documents	Europe's Beating Cancer Plan
		Stratégie nationale de lutte contre les cancers 2021-2030
	Calls	Funders listed below
	Proposals	Funders listed below
Funded Projects	Datasets of founded projects, entailing titles and abstracts, budgets, partners, and detailing publications related to these projects	European Community
		NIH
		ANR
		INCa
		FRM (Fondation pour la Recherche Médiale)
		ITMO Cancer
		EFS (Établissement Français du sang)
		Fondation ARC
		Ligne nationale contre le cancer
Scientific productions	Publications	Semantic Scholar
		PubMed (including MESH)
	Patents	PATSTAT
	Clinical trials	clinicaltrials.gov

Medical impact	Clinical guidelines	Extract PubMed
	Drugs	Drugbank
	New Technologies	Industrial Web site
		Technology news derived from European Media monitoring & Meltwater ?
Social and economic impact	Data on public health	Global Burden of Disease
	Jobs in Europe for researchers and entrepreneurs	Euraxess
	web site of associations, to extract positions papers	FONDATION ARC POUR LA RECHERCHE SUR LE CANCER (https://www.fondation-arc.org) LIGUE CONTRE LE CANCER (https://www.ligue-cancer.net) ROSE UP ASSOCIATION (https://www.rose-up.fr) AFSOS (https://www.afsos.org) Lung Cancer Europe (https://www.lungcancereurope.eu https://www.lungcancereurope.eu/2021/12/16/lung-cancer-europe-luce-position-paper-2015) European Society for Paediatric Oncology (SIOP Europe or SIOPE, https://siope.eu https://siope.eu/news-and-resources/position-papers) EUROPEAN BREAST CANCER COUNCIL (The platform for breast cancer specialists and patient advocates, https://ebccouncil.com, https://ebccouncil.com/position-papers) Alcimed (https://www.alcimed.com/en/position-papers/alcimed-position-paper-quality-of-life)

The crosscutting data analysis requires to be able to connect these data together and to link them to projects funded. Linking needs are shown in the Table 4.

Table 4: Health Living Lab – Connection between databases

Priority Level	Type of Data	Link to projects
1	Scientific publications	Acknowledgements Funded authors Funding Institutions
	Patent	Citations in the Patent of a publication related to the funded project Funded inventor
	Clinical trials	NCT number
	Clinical guidelines	Citations of publications related to the funded project
2	Drugs	Patent Companies involved in the project
	New technologies and diagnostics	Patent? Companies involved in the project
	Social media buzz	Mention of a funded researcher or funded project in medias Common topic
3	Data on health	Common topic

	Positions papers	Mention of a funded researcher or funded project in positions papers Common topic
	patient associations' positions	Common topic

The challenge for impact path analyses was the ability to link projects to scientific outputs and then to the impacts of these outputs.

3.4.3. Health Living Lab: Stakeholders

The methodology has foreseen a broad consultation of users in the framework of three working groups: one with policymakers, another with contributors from the research system (researchers, industrialists) and a third with civil society. The consultation envisaged via design thinking methods, widely used in services and industry, provided for groups of several participants with common needs.

In 2021, Hcéres met several actors of the cancer domain (researchers, analysts, decision makers, etc.) to have a first round of a gathering of needs and of analysing how these actors could be involved in a living lab.

In view of the first exchanges, it seemed difficult to implement a design thinking approach with a large group of users representing all kind of policymakers:

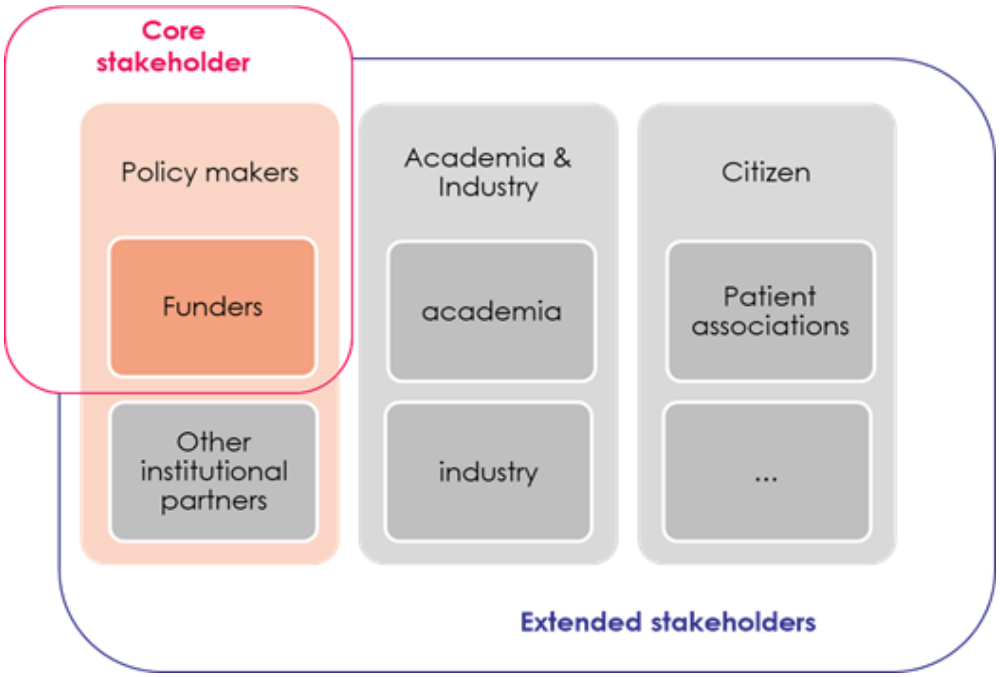
- It is difficult to approach and involve policymakers in a long and exploratory process
- The policymakers may have specific needs and it is not certain that a mutualised approach is relevant (at the risk of remaining only conceptual)
- The final product of the Living Lab is difficult to visualise until it has been applied to concrete needs
- IntelComp tools are very technical and will require a significant investment to understand and explore their potential

The funding agency appeared to be a particularly relevant user of IntelComp. Funders could ask questions upstream of funding (which field to be funded, which team), as well as downstream (what the scientific results of funded projects are, what is the societal impact of funded projects). All funders could share a common type of use case.

In that context, we proposed to work with two levels of engagement (cf. Figure 3):

- a group of "CORE" stakeholders with research funding stakeholders for whom use cases related to Impact pathways of funded projects will be investigated in the Living Lab. Each use case may be considering a specific study.
- a group of "EXTENDED" stakeholders with other policymakers, academia, industry and citizen representatives who will be associated with the main results and feedback of the IntelComp platform functionalities.

Figure 3: Health Living Lab – *core* and *extended* living lab stakeholders



The Health Living Lab first involved the core group of stakeholders (C), then the extended group (E) was confronted with the result of the core group (cf. Table 5).

Table 5: Health Living Lab – First stakeholder mapping

Name	Policymakers		Academia and Industry		Citizens	Why are they to be / were they recruited
	Funders	Other institutional stakeholders	Academia	Industry	patient associations	
INCa	C					main national funder on cancer research
ANR	C					major public funding agency in France
FRM	C					funding agency with willing to measure their impact
ITMO Cancer	C					large views on cancer research projects
Fondation ARC	C				E	foundation (both funder and patient association)
EFS	C					involved in reflexion about programs evaluations
Ligue	C				E	foundation (simultaneously funder and patient association)
IGR			E			research and medical institution specialised on cancer
Pasteur			E			research and medical institution involved on cancer
HCSP		E				main institution involved on health program
IRESP		E				main research institution about public health
LEEM		E				representative of pharmaceutical industry
UniCancer			E			in charge of promotion of clinical and translational research on cancer
BMS				E		biopharmaceutical company deeply involved in cancer research
Takeda				E		pharmaceutical company involved on cancer research
DGOS						Department of Ministry of health
EU Mission Cancer		E				dedicated transversal mission on Cancer at European level

3.4.4. Alignment with technical development

The living lab approach was grounded on agile co-creation involving some partners bringing their technical competency, the users, and other partners acting as an interface between them, so coordination is crucial.

In the case of the Health Living Lab, implementation teams Hcéres was structurally disjoint from technical works, which are assigned to different partners (and moreover, partners from different Members-states).

Hcéres has taken various actions to address this lack of connection with the technical teams:

I. Proactive communication on First Round Needs

During the first phase of consultation with external stakeholders of the Health Living Lab, Hcéres formalised several notes in order to ensure proper coordination with the technical teams in the definition of the Living Lab and the feasibility of the expectations of its stakeholders:

- The first note “Health Living Lab Needs” aimed at formalising the main use cases of the IntelComp platform in the Health Living Lab in order to discuss with the project's technical teams the feasibility of the services envisaged and to specify which tools will be available in practice when the Health Living Lab is launched in 2022.
- The second note “Statement of data need” aimed at formalising the data requirements in IntelComp platform in order to challenge the use cases submitted by Health Living lab’s stakeholders.

In parallel, Hcéres got involved in the WP1 meetings, which was not scheduled in initial Description of action, in order to bring up the policy questions that external stakeholders wish to address through the Health Living Lab and the data needed to provide a credible answer to these policy questions.

The feasibility of stakeholders’ expectations as well as the data engaged to address them needed to be clarified, in order to be able to continue collaborating with the external partners involved in the Living Lab.

II. Alignment with technical planning

The Health Living Lab events had been planned (cf. Roadmap section below) to be aligned with technical development of the project: dates for events and workshops had been scheduled to run in parallel with the delivery dates of different IntelComp tools for better synergy.

This good articulation between the planning of the Health Living Lab and the planning of the development of the tools was essential to be able to hold successful co-creation sessions in the framework of the Health Living Lab.

III. Implementation of operational collaborations

Hcéres committed itself in enhanced information exchange between technics and implementation by suggesting early access for its teams to some IntelComp tools and data. The content of this collaborative work is described in part 4.

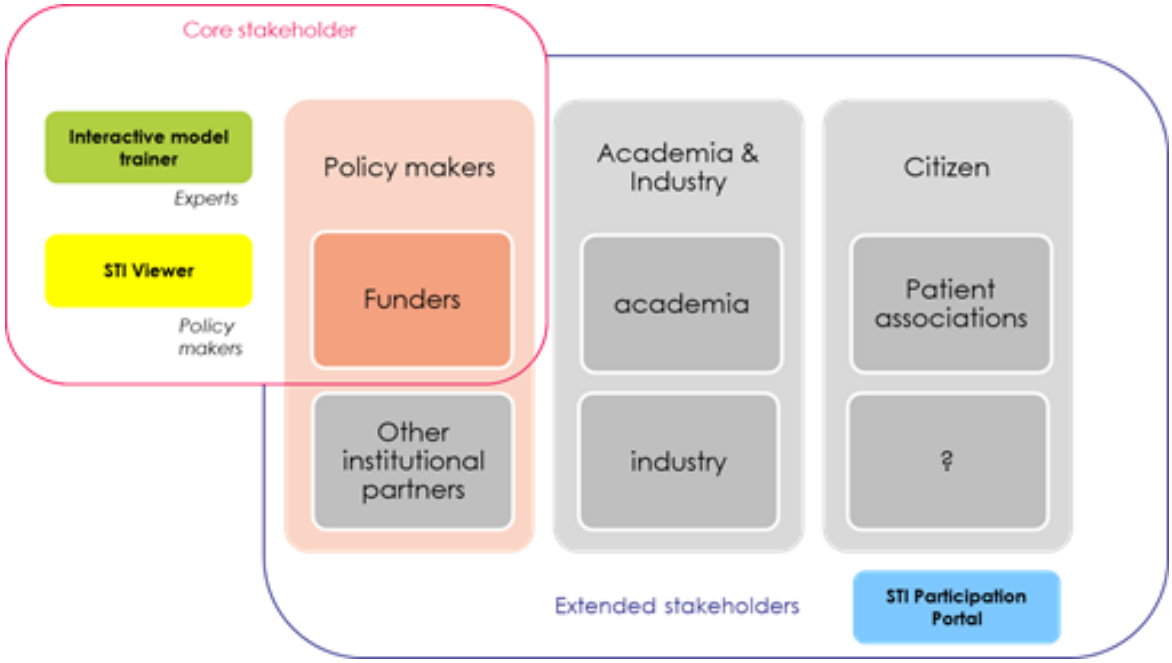
3.4.5. Health Living Lab: Target Roadmap

The Health Living Lab objective was to first address the needs of the core group of stakeholders and exploit the potential of:

- The Interactive Model Trainer (IMT) tool with a group of cancer expert & technical advanced users from core stakeholders to determine the relevant thematic approaches to characterise cancer research.
- The STI Viewer tool with a group of data analyst & decision makers from core stakeholders to study the impact pathways of programs or group of projects funded by the core stakeholders.

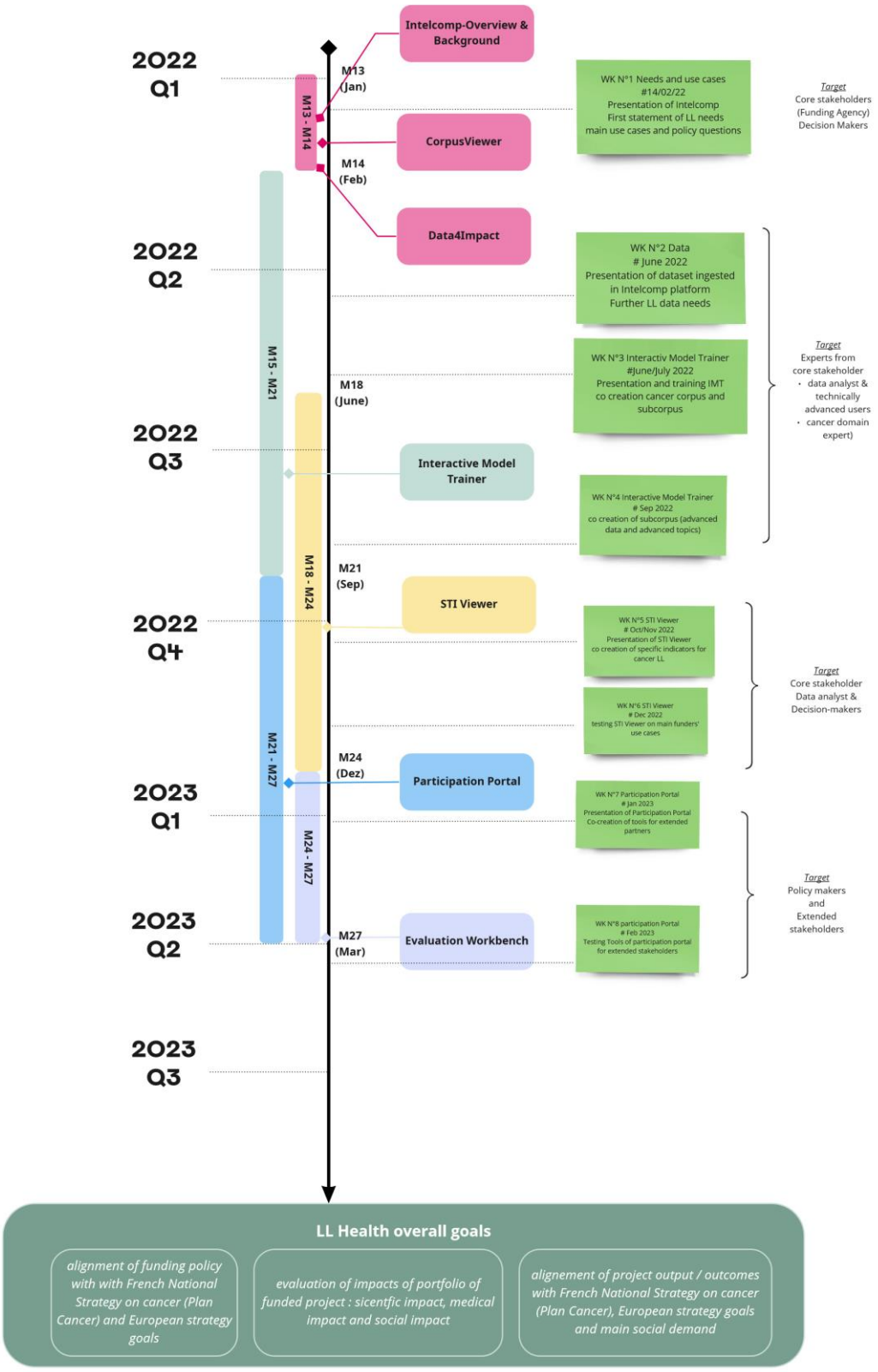
Secondly, the Health Living Lab planned to address the needs of the extended group of stakeholders focused on co-creation tools from STI Participation Portal.

Figure 4: Health Living Lab – IntelComp services co creation with *core* and *extended* living lab stakeholders



The Health Living Lab roadmap have been planned to be aligned with technical development of the project:

Figure 5: Health Living Lab – Target Roadmap



4. LIVING LAB RESULTS

4.1. Stakeholders' involvement and updated roadmap

Hcéres met with several actors of the medical research domain to have a first round of a gathering of needs and to analyse how these actors could be involved in the living lab. After this broad consultation of potential users (researchers, analysts, decision makers, etc.), it appeared essential to have efficient involvement and co-construction process, as well as to work with a group of limited participants with common needs.

Stakeholders chose to focus the Health Living Lab on the impact of funded projects and a group of actors interested in this goal was involved in the co-creation process. This group was set up with the support of INCa (Institut National du Cancer), which is the main funder of cancer research in France and a key player in the field in charge of coordination of all research on cancer domain in France. The group of stakeholders involved in the living lab has brought together the main policy makers and funders of cancer research in France and policy stakeholders involved in cancer research projects such as patient associations or research institutions.

The external participant (stakeholders) involved in the Health Living Lab were:

- **INCa** (Institut national du cancer) is the major national funder on cancer research also in charge of coordination of cancer research in France.
- **ANR** (Agence National de la Recherche) is the project-based funding agency for research in France and a leading actor in France on impact evaluation of projects.
- **FRM** (Foundation pour la Recherche Médicale) is a funding agency on medical research with strong willing to measure their impact.
- **EFS** (Etablissement Français du sang) is an institution involved on research on blood cancer and strongly involved in reflections about programs evaluations.
- **Fondation ARC pour le Cancer** is a non-profit association working both as a patient association and funder of research projects on cancer.
- **AFM-Telethon** is a patients' organisation fighting neuromuscular diseases, rare genetic disorders that kill muscle after muscle**.** AFM-Telethon is also a funder in the research in neuromuscular and rare genetic diseases.
- **ITMO Cancer** Institution in charge of coordination and policy making of cancer research.

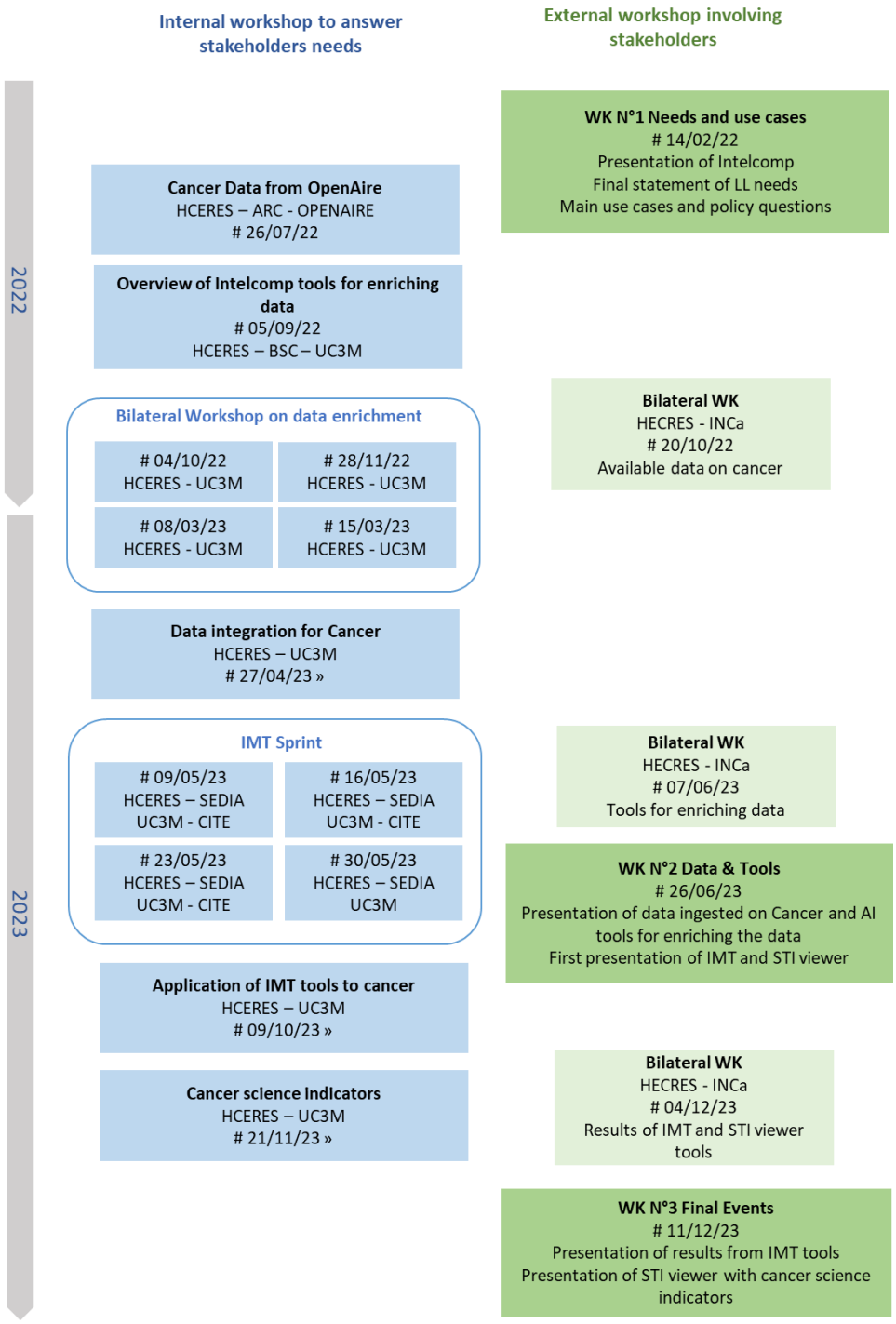
Figure 6: External stakeholders involved in the Health Living Lab

To adjust the tools to the stakeholders' needs in a co-creative process, the Health Living Lab developed two main types of activities.

- The first type of activities were internal activities to prepare meetings with external stakeholders and adjust data and tools to their needs. These activities did not involve external stakeholders, but mobilised members of the consortium.
- The second type of activities were events involving the previously listed external stakeholders. The first event established the framework for participation in the living lab, with ever-increasing engagement by stakeholders in the following events of the Health Living Lab. Most of the participants were responsible for the evaluation of research in their institutions. The main goals of these events were to show the functionalities of the IntelComp tools and get feedback on functionalities and on the main results of using such tools.

The roadmap had to be adjusted to this living lab organisation and to the technical issues and delivery dates of the different IntelComp tools. The final list of meetings and events of this LL is given in Fig. 7.

Figure 7: Health Living Lab – Adjusted Roadmap



4.2. Outcomes of the Living Lab process

4.2.1. Data exploration

First, as mentioned in subsection 3.4.2 (data facilitation), the broad spectrum of integrated data was a major goal to be able to study the impact pathway of funded projects. It was essential to give external stakeholders engaged in the Health Living Lab accurate information about data available in the IntelComp platform and about the ingestion process of new datasets.

To clarify the data available on funded projects and publications linked to these projects, Hcéres explored data that IntelComp could include in its database about the main funder's projects (INCa) via OpenAIRE. These data built on the basis of INCa data on projects were rather poor concerning publications. This is due to the lack of data on *funding acknowledgement* in publication funded by INCa, as well as the lack of reporting accessible on INCa projects. The constitution of a subset of publications related to project was not feasible using only these two methods.

Figure 8: Data in OpenAIRE about INCa projects and related publications

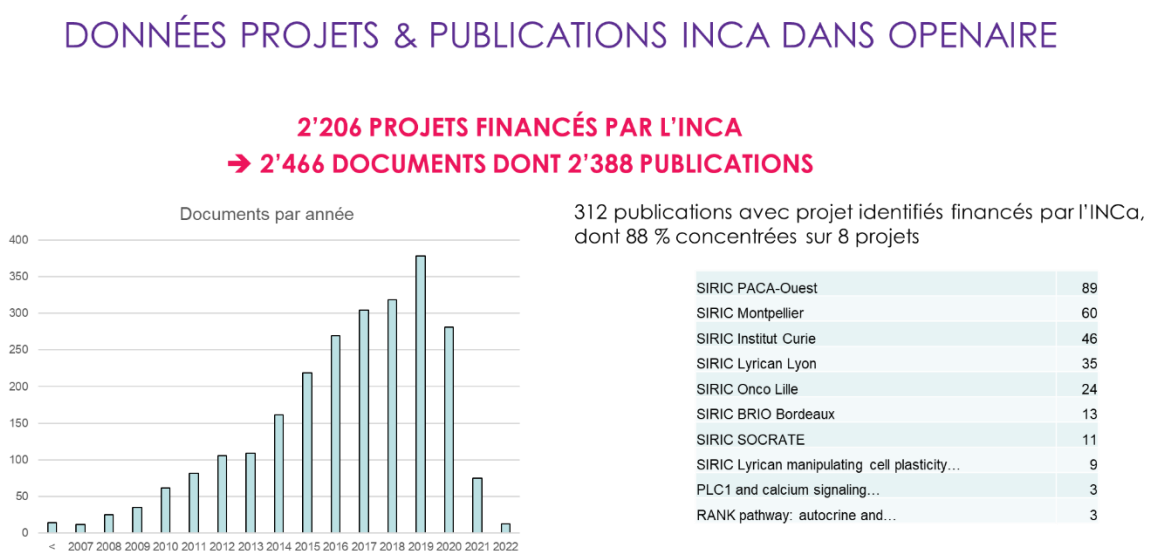
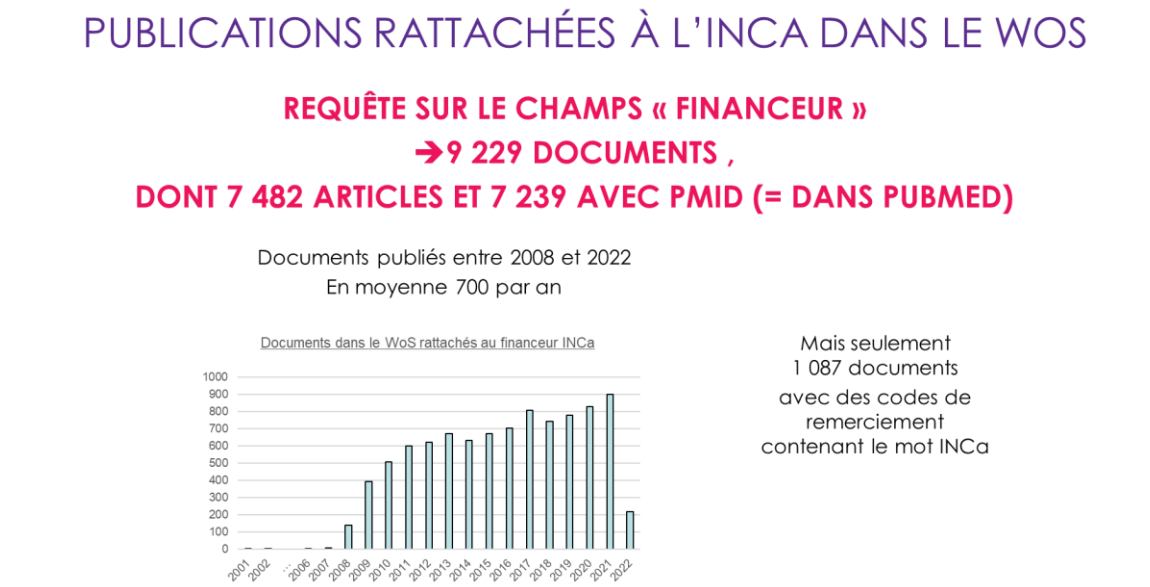


Figure 9: Exploration of data available about publications related to INCa project on Web of Science



Following these results, other techniques to enrich the data on publications related to projects were needed. Hcéres worked with UC3M to apply these techniques to the Health Living Lab, as it is explained in the next subsubsection.

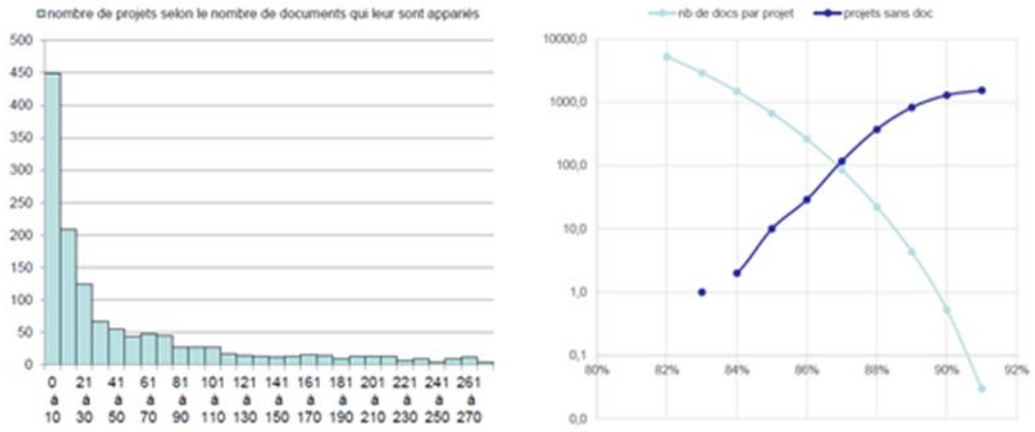
4.2.2. Application of IntelComp Graph Service to INCa dataset

Given the lack of data on publications linked to projects funded by the main cancer research funding agencies in France, Hcéres and UC3M worked together on the application of the graph analysis service to this problem and had various meetings:

- 05/09/22: Overview of IntelComp tools for enriching data (Hcéres – BSC – UC3M)
- 04/10/22: Exploring the potential uses of graph analysis to identify publications related to project (Hcéres – UC3M- ZSI)
- 28/11/22: providing available data and organizing work (Hcéres – UC3M)
- 08/03/23: analysis of preliminary results (Hcéres –UC3M)
- 15/03/23: analysis of additional results (Hcéres – UC3M)

This collaboration was highly instructive and showed the potential capabilities of semantic analysis to identify the scientific output of projects. The experiment was carried out mostly on an exploratory basis. On the one hand, tests have shown that interesting results could be obtained with semantic similarity between projects' abstract and publications' abstract for certain calls for projects. On the other hand, however, semantic analysis does not work to identify the scientific output of projects whose abstracts do not include any scientific content. Indeed, some projects are intended to fund research infrastructures or the setting up of collaborations, and do not mention the scientific content of the research to be carried out. In this case, semantic analysis is not the appropriate tool for identifying the scientific production linked to these projects.

Figure 10: Exploration of semantic similarity between INCa Projects and publications on Cancer



Hcéres presented these results to the group of stakeholders (cf. 4.3.3). Following these results and the available time, Hcéres proposed to the stakeholder group to go on experimentation of the IMT tool on European projects for which there is data on publications coming from reporting and acknowledgement.

4.2.3. Use of the IMT tool to answer to the Cancer Research Funders' needs

IMT internal sprints were rolled out during the months of May and June 2023 (on a weekly basis in May, and every 10 days approximately in June), under the aegis of the University Carlos III technical team. Both the AI LL (SEDIA) and the Cancer LL (Hcères) were involved in a fruitful cocreation process developed via Teams, which aimed the following objectives:

- Create cancer elated subsets for the following corpora: CORDIS, OpenAIRE publications, and PATSTAT, and create topic models for these subsets;
- Validate the services of the current IMT. Modify how functionality is presented in terms of usability ("cosmetic changes");
- Identify new functionalities that the technical team could work on after Summer 2023.

This sprint activities were supported by various meetings also involving the AI Living Lab:

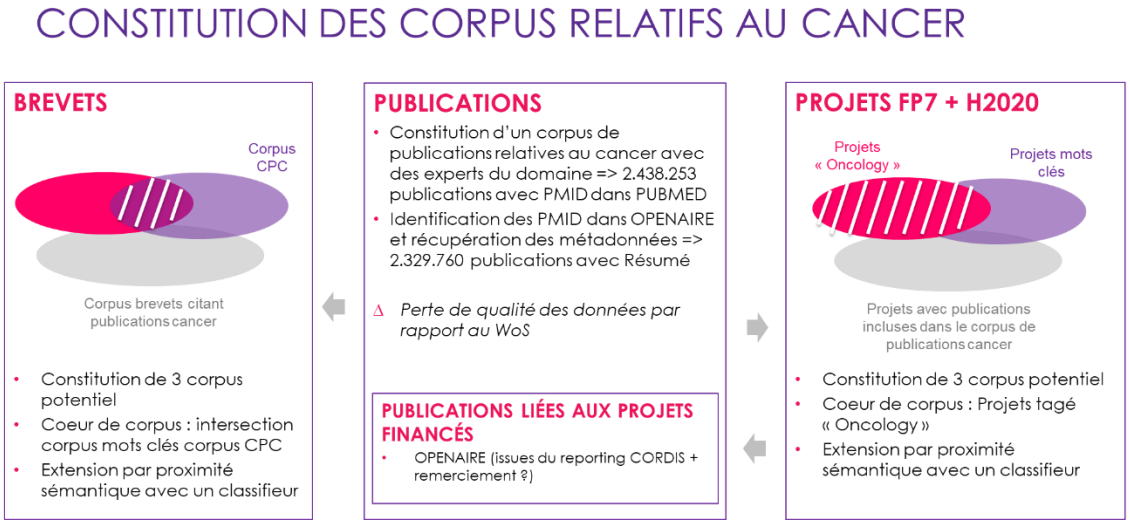
- 09/05/23: ingestion of data, preparation of deployment of main services (UC3M – CITE – SEDIA – Hcéres)
- 16/05/23: deployment of topic modelling components (UC3M – CITE – SEDIA – Hcéres)
- 23/05/23: deployment of domain classification component (UC3M – CITE – SEDIA – Hcéres)
- 30/05/23: creation of topic models as needed (UC3M– SEDIA – Hcéres)

After this the sprint session Hcéres and UC3M have continued to work together to prepare the Health Living Lab workshop on results of IMT tool applied to cancer. These complementary activities were supported by two bilateral meetings:

- 09/10/23: Application of IMT tool to cancer (Hcéres – UC3M)
- 21/11/23: Cancer science indicators (Hcéres – UC3M)

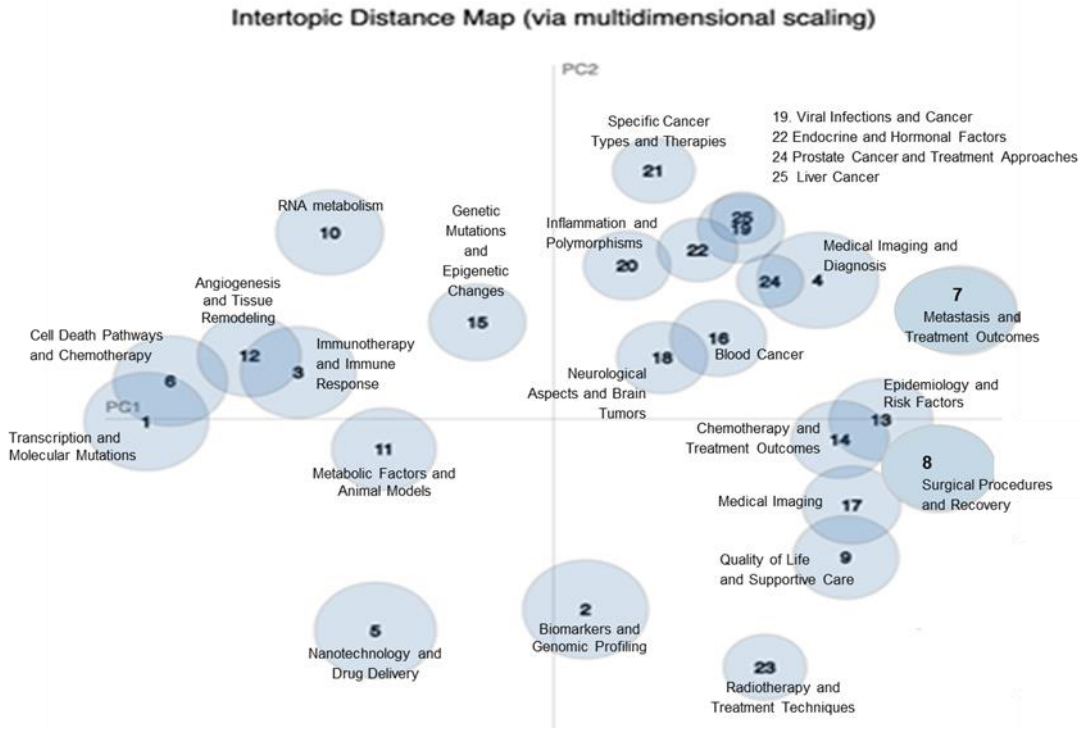
Hcéres and UC3M have continued to collaborate. First to Integrate cancer publications databases and build specific patent and European funded projects databases related to cancer.

Figure 11: Construction of Sub Corpus on Cancer



Second, UC3M created a topic model for the publications corpus and use ChatGPT to label the different topics. This work was carried out using the functionalities available in the IMT. UC3M and Hcéres had very fruitful exchanges on the methods used to build the corpus and built the topic model.

Figure 12: Topic Model on Cancer Publications



IMT tool appear to be too technical to work with for stakeholders of the cancer Living Lab. That said, the latter found the results of domain classifier to build thematic corpus and of topic modelling to explore content of corpus very innovative.

4.2.4. Integration of results in the STI Viewer tool

STI Viewer and STI Policy Participation Portal were the two main tools to address the impact pathway, which was the core interest of external stakeholders in the Health Living Lab. However, the stakeholders involved in the LL were reluctant to work on a tool intended for a large audience (the STI Policy Participation Portal), and the Health Living Lab therefore focused solely on the STI Viewer.

Topic model enriched indicators are an innovative approach specific to IntelComp tools. The indicators in the health domain calculated by UC3M had been included in the STI Viewer. Hcéres and UC3M worked together to define different views that have been integrated on world / Europe / France publications on cancer research and publications related to European funded projects on cancer to compare publications from these different scopes. Figures 13, 14 and 15 below illustrate just some examples of the indicators available in the STI Viewer and the analysis that this tool enables on cancer research.

Figure 13 shows the total number of articles on cancer available in OpenAIRE, as well as number of articles in Open Access.

Figure 13. Worldwide scientific production on cancer over time

Figure 14 shows the evolution of scientific production in cancer where at least one of the authors belongs to a European institution. The figure shows the distribution of the scientific production across the topics identified using the Interactive Model Trainer.

Figure 14. Evolution of scientific production in cancer

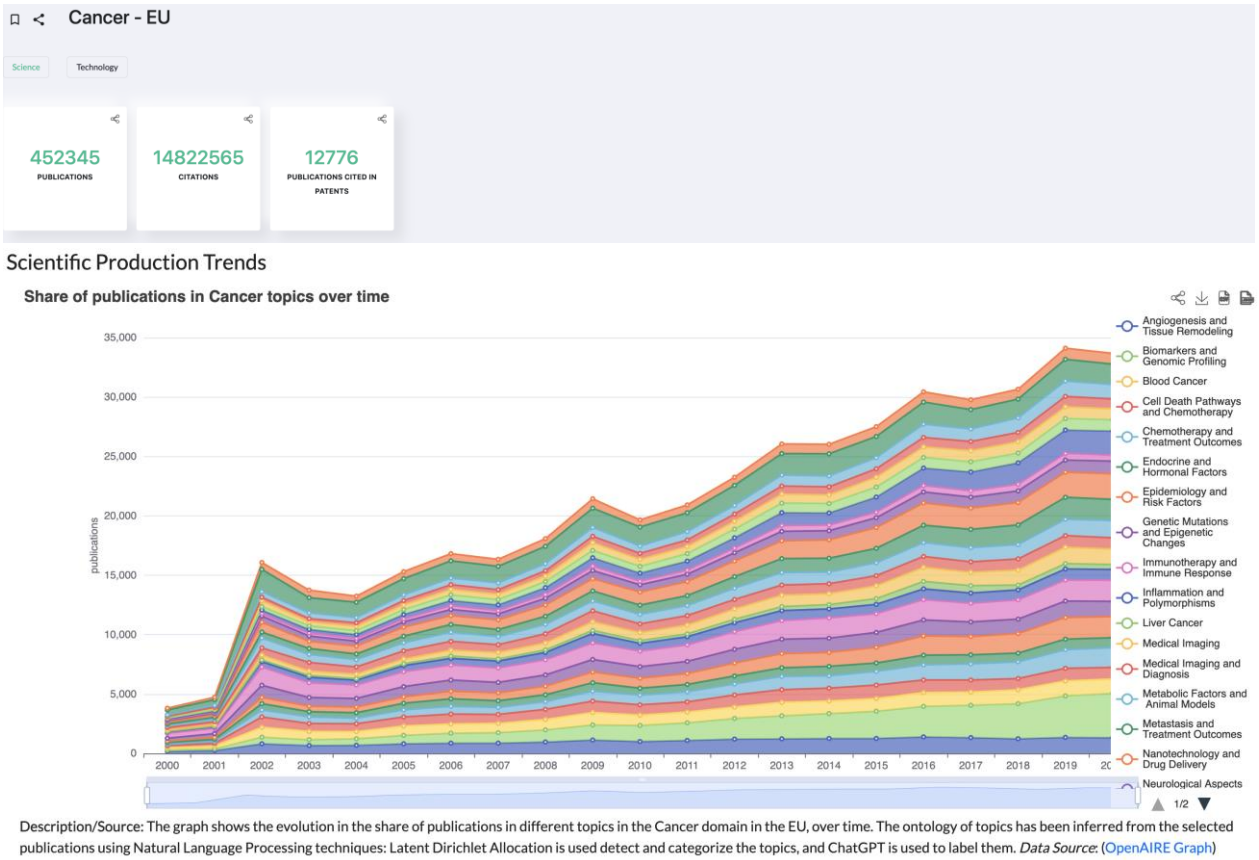
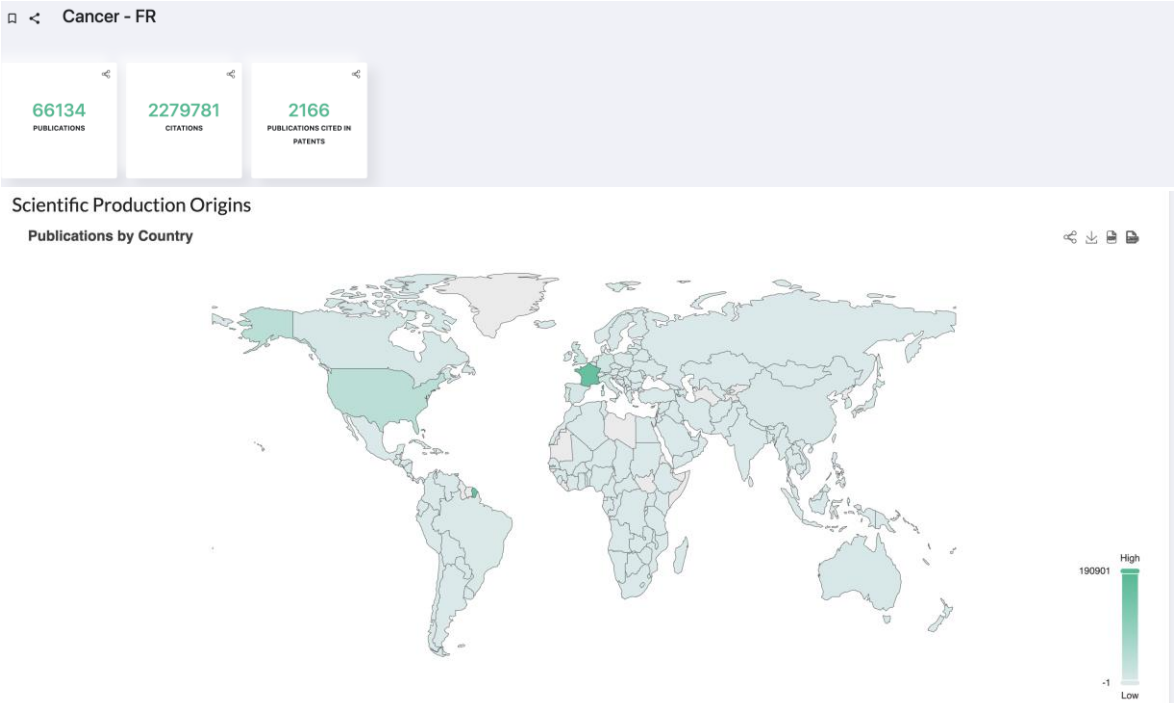


Figure 15 shows the scientific production on cancer in France, i.e., selecting only the papers where at least one of the authors belongs to a French institution. The figure shows country affiliation distribution of the papers authors.

Figure 15. Scientific production on cancer in France



Description/Source: The graph shows the number of publications in Cancer by the country of the affiliated organization of an author (with at least one author affiliated to an FR organization). In the case of multiple authors each organization and corresponding country is counted as a separate publication (e.g., one publication with three authors, where two are from Greece and one is from Spain is counted as two publications in Greece and one in Spain). *Data Source*: OpenAIRE Graph

4.3. External events

The external events with stakeholders were organised mainly in 2023 to be aligned with available tools and data.

Figure 16: External events with the stakeholders

Consultation Workshop on STI Policy Needs
June 10th 2021
Virtual workshop
Target group : Policy makers and stakeholders in Cancer area
Consultation Workshop dedicated to exploring the needs of STI policy makers and policy stakeholders in the Health-Cancer domain

Kick Off Workshop with Funders to present IntelComp Project and identify specific needs
feb 14th 2022,
Virtual workshop
Target group : French Funders of Cancer Research
Consultation Workshop dedicated to present IntelComp Project and exploring which specific needs of funders in the Health-Cancer domain
Intelcomp tools could answer

First Workshop on Intelcomp Tools
June 26th 2023
Face-to-face workshop
Target group : French Funders of Cancer Research
Applications of AI tools to identify publications related to funded projects
First presentation of IMT and STI viewer tools

Second Workshop on Intelcomp Tools
December 2023
Face-to-face & virtual workshop
Target group : French Funders of Cancer Research
Application of IMT to define domain corpus and topic modelling
Presentation of STI viewer tools applied to cancer data with a focus on projects funded by EU

First Consultation Workshop was organised by WP1.

4.3.1. Kick-off Event: Consultation meeting with French Funders

4.3.1.1. Background and objectives of the event

The objective of this meeting was to present the IntelComp project to the main French funders of research on Cancer and to offer them the opportunity to participate in the health Living Lab. This event was also the time to collect theirs needs regarding this Living Lab.

4.3.1.2. Preparation of the event

This kick-off event was prepared by two meetings with INCa which is a key player in the cancer in France. It was essential to involve INca in the project and to involve it in the implementation of the LL.

The first meeting took place at INCa on September 16, 2021, in order to present IntelComp and the way to work in the Living Lab and to discuss the content of the Living Lab and the main actors to involve. The second meeting took place remotely on January 18, 2022, in order to prepare the kick off meeting.

A first note presenting the IntelComp project and proposal to join the Health Living Lab was sent to participants before the Kick-off event. (cf. Annex).

4.3.1.3. Details about event

The first event (kick-off) involving external stakeholders took place online on February 14, 2022 from 10:00 to 12:00 AM.

The participants of this event were representing all the main French funders of research on Cancer.

External participants:

- **Carla Estaquio** (Manager Evaluation Mission - Research and Innovation Division at INCa – Institut National du Cancer)
- **Martine Garnier-Rizet** (Head of Digital Strategy and Data Directorate – at ANR - Agence nationale de la recherche)
- **Marianne Lanoë** (Data Analyst at ANR - Agence nationale de la recherche)
- **Florence Guibal** (Head of the strategical scientific domains / Analysis of the research programs at FRM - Fondation pour la Recherche Médicale
- **Valérie Lemarchandel** (Chief scientific officer at FRM Fondation pour la recherche médicale)
- **Hugo Juraver** (In charge of research program analysis at FRM Fondation pour la Recherche Médicale)
- **Sandrine Daubeuf** (Head of non-thematic projects at Fondation ARC)
- **Angélique Michaut** (Research Manager at EFS (Establishment Français du sang)
- **François Castagner** (Project Manager at Aviesan - ITMO Cancer)

Hcéres participants:

- **Frédérique Sachwald** (Director, OST - Observatoire des Sciences et Techniques)
- **Tessa Enock-Levi**, (Project Manager at OST- Observatoire des Sciences et Techniques)
- **Dominique Guellec**, (Scientfic Advisor at OST - Observatoire des Sciences et Techniques)
- **Luis Miotti**, (Scientfic Advisor at OST - Observatoire des Sciences et Techniques)
- **Aurélien Leynet** (Data Scientist at OST - Observatoire des Sciences et Techniques)
- **David Sapinho** (Senior Data Scientist at OST - Observatoire des Sciences et Techniques)

This event was hosted by Hcéres team who took charge of the entire presentation following this agenda:

1. Tour de table - introduction of participants.
2. Presentation of the IntelComp project.
3. Presentation of the living lab and how to participate.
4. Questions and answers on the project and the living lab.

Presentation used on this workshop can be found in the webpage of the event[5].

4.3.1.4. Main results of the kick-off meeting

The stakeholders expressed their interest in the IntelComp project and their desire to follow the progress of the Health Living Lab. The stakeholders agreed to form the "core group" of participants in the Health Living Lab and to be involved in different steps of the Health Living Lab.

The area of greatest interest to the stakeholders consulted by Hcéres during this event was the analysis of the impact of funded research projects (or programs or group of projects) and the characterisation of 'impact pathways'. This use case involves monitoring and characterising the different stages leading from project selection up to different research results and then to their socio-economic impacts as presented in Subsection 3.4.2.

This use case was chosen to develop IntelComp tools in the Health Living Lab with the group of funders attending this first event. This use case was described and forwarded to the technical teams.

Another use case for the IntelComp platform was mentioned later in the event, which concerned the identification of scientific breakthroughs in a field of research. This use case corresponds to the Living Lab's focus on AI. Depending on the progress of the Cancer Living Lab's work, and the time and resources available, this approach was also identified to be tested in terms of it being an opportunity regarding the development of the Health Living Lab.

The stakeholders also raised several questions about the project:

[5] https://intelcomp.eu/events/kick-event-consultation-meeting-french-funders-february-2022

Questions raised about the data used in IntelComp

The potential partners were interested in the large scope of the data that can be mobilised with IntelComp and in the possibility of connecting these data together and linking them to projects funded by their agency.

They also asked to be more precise about the data that seemed realistic to mobilise in the health living lab. They needed to know what would be available for the next workshop. The main questions raised by stakeholders about feasibility of their data needs forwarded to technical teams were the following:

- Which data will be available when the living lab starts (early 2022)?
- Which data has already been identified and will be integrated later (and when)?
- How will we proceed to integrate new data sets (when should we identify them, what constraints should we anticipate, who will be in charge of what?)
- Will it be possible to connect IntelComp with non-open data: user-owned data that cannot be open, such as paid databases (WoS) or confidential data (budget data, application data...)?
- What data seems to be much more difficult or impossible to integrate into IntelComp?

Questions raised about the tools used in IntelComp

The challenge for impact path analyses is the ability to link projects to scientific outputs and then to the impacts of these outputs. The main questions raised by stakeholders about tools available in IntelComp were the following:

- What tools are already being considered (available or in development) to make these links?
- Links by acknowledgements in scientific publications?
- Links by researcher's names or institutions involved in funded project?
- Links by publications citations in patent?
- Links by common topic (topic modelling) or common words (lexical approach) ?
- What are the links that can be implemented at the start of the living lab? / planned but under development / that seem impossible to implement

These questions about tools to link data were forwarded to technical teams.

The analysis of the impact pathway will be interesting to cross with the type of actors involved at each stage: public research actors, but also pharmaceutical companies (new treatments), engineering companies (diagnostic techniques) and patient associations.

Several thematic approaches related to cancer were also raised by stakeholders:

- Analysis by type of research (basic / clinical / translational);
- Identification of research cooperation (international cooperation, inter-institutions, public-private partnerships, etc.)
- Analysis by research discipline (e.g. epidemiology; social sciences…)
- Focus on certain types of cancer (cancer location in particular) & prognosis (good / bad) or incidence
- Possible focus on issues related to cancer: tobacco, alcohol, food, pollution...

- Identification of new treatments and breakthrough technologies (genetics, biotherapies, predictive medicine, e-health)
- Characterise projects/work related to the different stages of patient care 1) prevention; 2) early detection; 3) diagnosis and treatment; and 4) quality of life for cancer patients and survivors.

These stakeholders expectations, forwarded to the technical teams needed to be prioritized depending on their feasibility.

Questions raised about implication and workload of stakeholders to be involved

First stakeholders asked if it is necessary to have a technical background and previous knowledge of existing tools (Corpus Viewer, Data4Impact and OpenAIRE to be able to follow the living Lab. Second, they asked what the workload would be to attend to all events and work on use cases.

Hcéres team explained the developing tools are supposed to be for policy makers and it is not necessary to have a high technical background, neither to have previously experiment of existing tools to participate. Hcéres proposed to shape the Health Living Lab workshop focused for a policy makers' audience. If necessary, more technical meetings would be organised with a more targeted audience.

4.3.1.5. Next steps

The various participants showed a strong interest in the next stage of the project and requested more information on the data that can be used to characterize the impact of funded projects. This was a point to be clarified in the next workshop. Minutes of this meeting, sent to all the participants, are available in the Annex.

4.3.2. Workshop on data and services to connect projects to publications

4.3.2.1. Background and objectives of the event

The instruction of questions raised by involved stakeholders during the first event about available data and services to link this data has been quite laborious. Hcéres committed itself to make a first diagnostic on available data on research projects about cancer funded by the main stakeholder of the Living Lab (INCa) and tested with the help of UC3M the feasibility to connect projects to publications. This work was presented in Subsection 4.2.2.

The objective of this event was first to present the diagnostic of data available on publications relative to projects funded by INCa, crossing several sources: OpenAIRE, INCa databases, Publications in Web of Science. Second objective of the workshop was to present the result of the work to connect publications to projects. This event was also an opportunity to quickly present the IMT and STI Viewer tools and collect suggestions for exploratory work for a next workshop.

4.3.2.2. Preparation of the event

This event required a deep preparation work with IntelComp technical team as described on in 4.2.1 and 4.2.2.

During this period, two bilateral meetings with INCa were organised to present technical issues on data and prepare the workshop with all the external stakeholders. The first meeting took place

online on October 20, 2022, in order to present data potentially available on IntelComp about INCa-funded projects. Due to the lack of data on publications related to INCa-funded projects, it was decided to continue to investigate the potential of IntelComp tools to enhance this data before holding the next workshop. The second meeting took place online on June 7, 2023 to prepare the presentation of the second event presenting the investigations made on data and a first demonstration of the services.

4.3.2.3. Details about event

The second event involving external stakeholders was organised face to face and took place on June, 26, 2023 from 10:00 to 12:00 at the INCa offices.

Despite it was a face-to-face event, main French Funders on Cancer (INCa, FRM, ITMO, AFM) and one research institution (Institut Pasteur) could attend.

External participants

- **Carla Estaquio** (Manager Evaluation Mission - Research and Innovation Division at INCa – Institut National du Cancer)
- **Florence Guibal** (Head of the strategical scientific domains / Analysis of the research programs at FRM - Fondation pour la Recherche Médicale
- **Valérie Lemarchandel** (Chief scientific officer at FRM Fondation pour la Recherche médicale)
- **Hugo Juraver** (In charge of research program analysis at FRM Fondation pour la Recherche Médicale)
- **Sandrine Daubeuf** (Head of non-thematic projects at Fondation ARC)
- **Angélique Michaut** (Research Manager at EFS, Establishment Français du sang)
- **Isabelle Besson Fauré** (Head Scientific Evaluation Department at AFM - Association Française contre les Myopathies)
- **Berangere Virlon** (Department Manager at Institut Pasteur)
- **François Castagner** (Project Manager at Aviesan - ITMO Cancer)

Hcéres participants:

- **Tessa Enock-Levi**, (Project Manager at OST- Observatoire des Sciences et Techniques)
- **Aurélien Leynet** (Data Scientist at OST - Observatoire des Sciences et Techniques)

This event was hosted by Hcéres team who took charge of the entire presentation following this agenda:

1. Tour de table - introduction of participants.
2. Background and ambition of the project.
3. Review of data and tools developed – Questions & Answers about data available.
4. Exploring AI services to identify publications from INCa-funded projects – Questions & Answers about possibility of AI services.

5. Other possible explorations: presentation of IMT and STI viewer and first feedback of stakeholders.

The presentation used for this workshop is available in the event webpage[6].

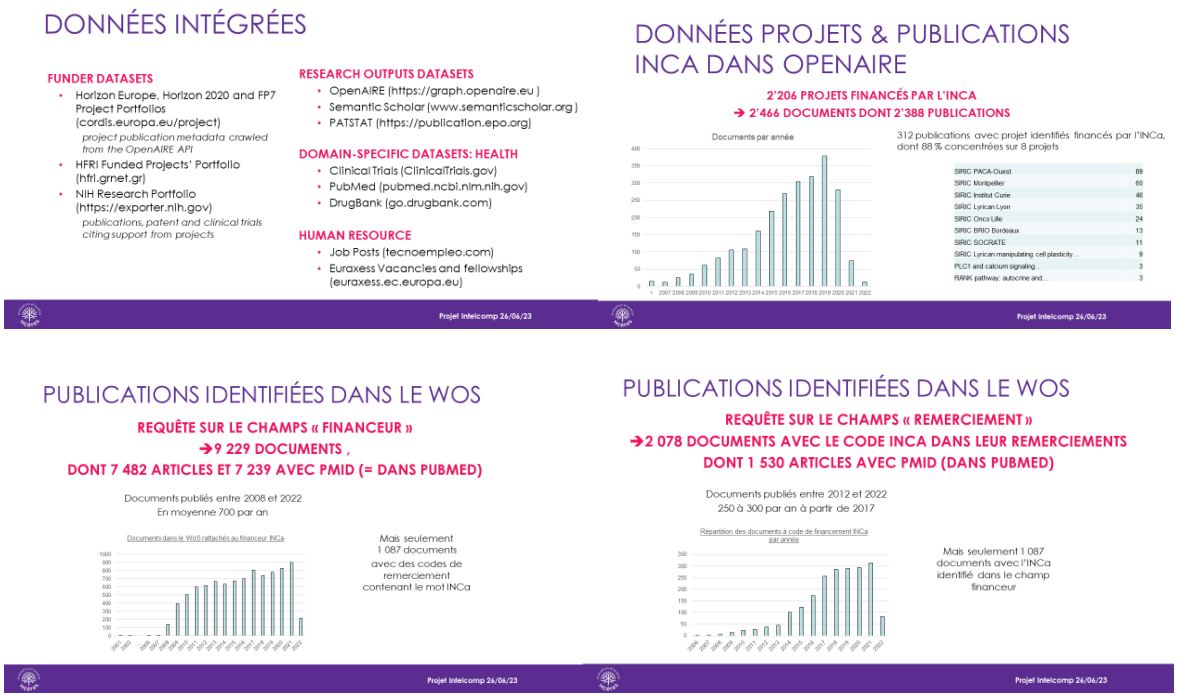
4.3.2.4. Main results

Due to the face to face organisation, the audience was smaller, but the exchanges were very rich, even though the subjects were highly technical.

The group of funders attending the event demonstrated a strong interest in the diagnostic of available data on publications related to project from different sources. They were also very interested in the practical implementation of AI services to connect publications and project data. Their enthusiasm extended to the valuable feedback provided by the INCa / Hcéres team, shedding light on the capabilities and limitations of AI in this application.

The experimentation of AI services to connect projects to publications related to these projects showed the necessity to adapt AI services to the type of projects funded and to involve human in the loop to have significant results. The general impression was that this process could not be fully automatized.

Figure 17: Results presented on available data on projects and related publications



[6] https://intelcomp.eu/events/health-living-lab-imt-workshop

Figure 18: Results presented about connections between projects and publications

Score	Projet	Document
92,0%	Role of radiotherapy with modified fractionation in locally advanced...	Altered fractionated radiotherapy in the management of head and neck...
91,8%	Randomized Open Phase III Trial Testing Efficacy of Gemtuzumab Ozoga...	Gemtuzumab ozogamicin for treatment of acute myeloid leukemia
91,6%	Blastic plasmacytoid dendritic cell neoplasm (BPDCN): in vitro and i...	CD28/4-1BB CD123 CAR T cells in blastic plasmacytoid dendritic cell ...
91,6%	NORAD01 (Non inferiority study of preoperative chemotherapy without ...	NORAD01-GRECCAR16 multicenter phase III non-inferiority randomized t...
91,5%	CBF-2018. A Dose-finding phase II Study of Gemtuzumab Ozogamicin in ...	Gemtuzumab ozogamicin for treatment of acute myeloid leukemia
91,5%	Clinical Benefit of Genetic Biomarkers for Guiding Treatment Decisio...	Individualized anticancer therapies: which regulatory guidelines?
91,5%	Exploitation of the anti-tumor properties of alpha-galactosylceramid...	Targeted Delivery of alpha-Galactosylceramide to CD8 alpha(+) Dendri...
91,4%	Highthroughput technologies to drive metastatic breast cancer patien...	Array CGH and PIK3CA/AKT1 mutations to drive patients to specific ta...
91,4%	Randomized phase III study of a treatment driven by early PET respon...	Efficacy of chemotherapy or chemo-anti-PD-1 combination after failed...
91,4%	MRI-FIRST 01:Improvement in the detection of aggressive prostate can...	Independent Evaluation of the Respective Predictive Values for High-...

Stakeholders also gave first feedback about IMT and STI Viewer after presentation of the tools with available data.

IMT tool appears to be too technical for the policy makers' audience. In any case, the stakeholders emphasised the interest in a tool that can precisely define corpus domains and identify subtopics. They showed a strong interest in results of IMT tool applied to real cancer data.

After the demonstration of STI Viewer with available data at that time of the project, the participating stakeholders were sceptical about the added value of this tool, compared to other tools on the market (like insight) that seem to offer the same kind of data but are more flexible. To ensure effective utilisation, they stressed the importance of practising with real cancer data to provide relevant feedback.

However, the funders also voiced concerns about the potential risks associated with the uncontrolled dissemination of unreliable data, a growing issue amid the increasing controversies surrounding health research.

4.3.2.5. Next steps

The various participants showed a strong interest in the thematic approach made possible by IMT tool. Even if using the tool doesn't seem to be their area of expertise they asked to see what kind of results and analysis can be made on real cancer data.

Even if participant were sceptical about version of the presented STI Viewer, they showed a real interest to put this tool to the test with real cancer data, to see to what degree it could be used to analyse impact of funded projects.

The next workshop was planned to focus on results and training of these tools with real cancer data.

4.3.3. Workshop on thematic exploration

4.3.3.1. Background and objectives of the event

The tools developed in IntelComp assume the existence of already connected data. Considering the lack of time until the end of the project, the living lab team proposed to explore IMT and STI Viewer tools on data available: publications about cancer and publications related to projects funded by European Community.

The objectives of the last event involving external stakeholders was to present the application of IMT tool to define domain corpus and topic modelling and to present STI Viewer tool applied to cancer data with a focus on projects funded by EU.

4.3.3.2. Preparation of the event

This event asked a preparation work both with INCa and with IntelComp technical team.

Preparation work with IntelComp technical team to create specific corpus on cancer data, apply topic modelling and integrate real cancer data in STI viewer have been described in Subsections 4.2.3 and 4.2.4.

Hcéres had a bilateral meeting with INCa which took place on line on December 4, 2023 in order to present main results, and to choose the ones to present at the final event and prepare the event.

4.3.3.3. Details about the event

The last event involving external stakeholders was organised both face to face at the INCa offices and online, to mobilise all the stakeholders involved in the living lab. It took place on December 11, 2023, from 14:00 to 16:00.

The workshop was attended by all the main French Funders on Cancer (INCa, ANR, FRM, AFM and Fondation ARC) and major research institutions (Institut Pasteur, INSERM, ITMO)

External participants:

- **Carla Estaquio** (Manager Evaluation Mission - Research and Innovation Division at INCa – Institut National du Cancer)
- **Marie-Alexandra Neouze** (Head of the Studies and Impact Unit – at ANR - Agence Nationale de la Recherche)
- **Florence Guibal** (Head of the strategical scientific domains / Analysis of the research programs at FRM - Fondation pour la Recherche Médicale
- **Nicola Eugénie** (In charge of research program analysis at FRM Fondation pour la Recherche Médicale)
- **Sandrine Daubeuf** (Head of non-thematic projects at Fondation ARC)
- **Florence Noel Boulain** (Scientific Project Manager, Foundation ARC)
- **Isabelle Besson Fauré** (Head Scientific Evaluation Department at AFM - Association Française contre les Myopathies)
- **Berangere Virlon** (Department Manager at Institut Pasteur)
- **Angélique Michaut** (Medical Research Coordination and Communication Manager at Institut Pasteur)
- **Muriel Altabef** (Project Manager at Aviesan - ITMO Cancer)
- **Gérard Bréard** (Director of the public health institute at INSERM)

Hcéres participants:

- **Tessa Enock-Levi**, (Project Manager at OST- Observatoire des Sciences et Techniques)
- **Dominique Guellec**, (Scientific Advisor at OST - Observatoire des Sciences et Techniques)
- **Marianne Lanoë** (Data Scientist at OST - Observatoire des Sciences et Techniques)
- **David Sapinho** (Senior Data Scientist at OST - Observatoire des Sciences et Techniques)

This event was hosted by Hcéres, who took charge of the entire presentation following this agenda:

1. Tour de table - introduction of participants.
2. Background and progress of IntelComp project.
3. Creation of specific corpus on cancer with domain classifier (from IMT) - Questions and answers on the tool.
4. Thematic exploration of corpus with topic modelling (from IMT) - Questions and answers on the tool.
5. Using STI Viewer to visualize impact of European founded project - Questions and answers on the tool.

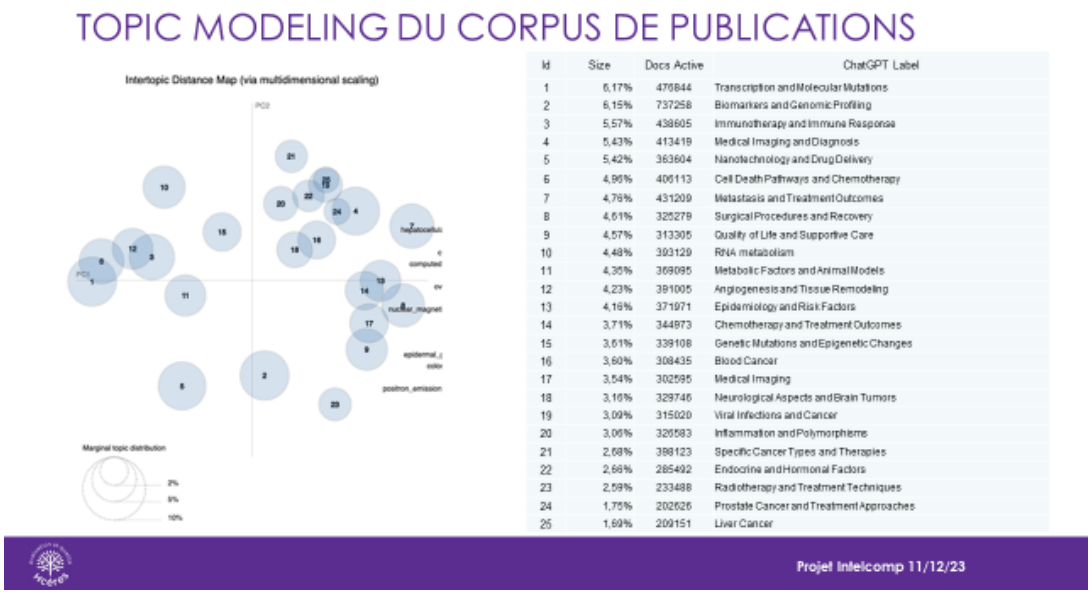
The presentation used for this workshop is available in the event webpage[7].

4.3.3.4. Relevant results

The creation of cancer related corpus with Domain Classifier from IMT didn't raise any questions in the audience; it was probably too technical, but also seemed professional to them.

The presentation of the topic model developed with UC3M was of great interest to participants. This technique allows them to explore emerging topic in cancer research, without preconceived ideas.

Figure 19: Presentation of topic modelling in IMT



[7] https://intelcomp.eu/events/cancer-living-lab-final-event

However, the results of topic labelling by ChatGPT raised questions and showed the need to adjust the process through expert feedback, which was not possible given the project's deadline.

Figure 20: Presentation of topic extraction and labelling with Chat GPT

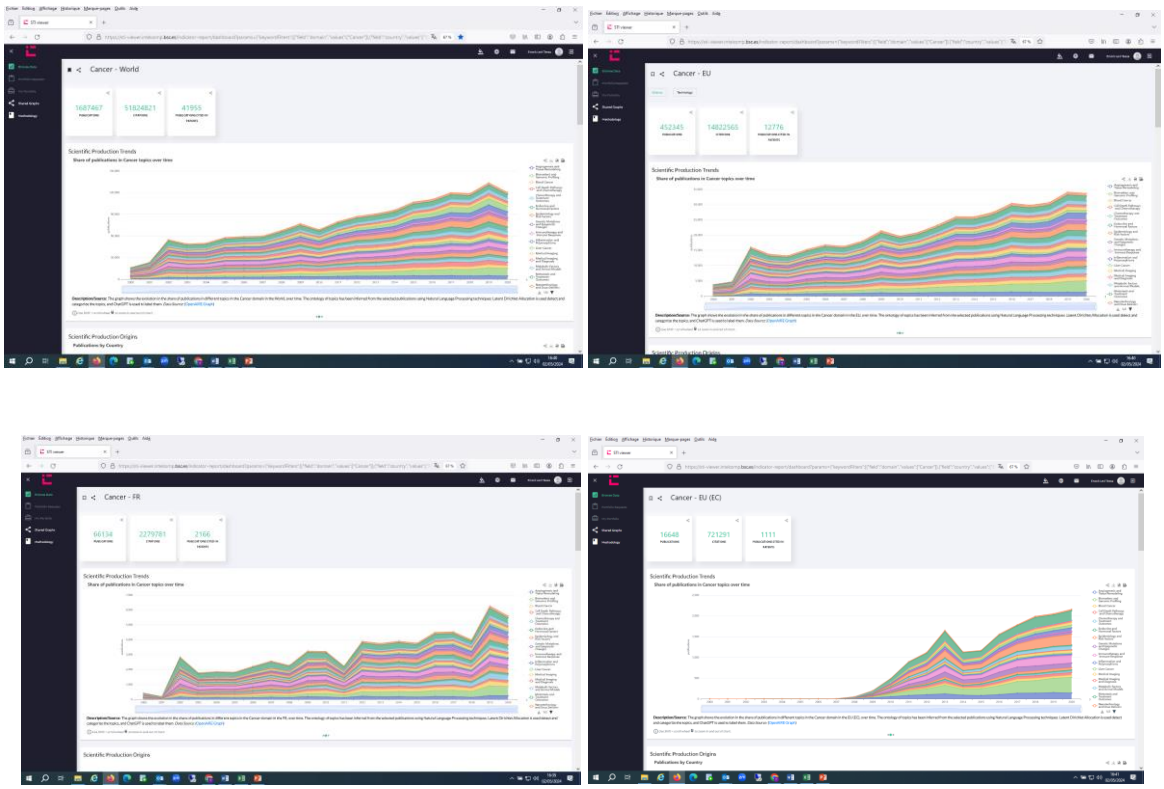
Id	Size	Docs Active	ChatGPT Label	Word description
1	6,17%	476844	Transcription and Molecular Mutations	protein, binding, transcription, DNA, kinase, promoter, mutation, phosphorylation, mouse, cellular, myc_oncogene, suppressor, ribonucleic_acid, ras_mutations, messenger_rna
2	6,15%	737258	Biomarkers and Genomic Profiling	molecule, biomarker, biological, therapeutic, genome, signature, classification, cellular, immunity, metabolic, heterogeneity, high_throughput, profiling, proteomic, immunotherapy
3	5,57%	438605	Immunotherapy and Immune Response	immunity, mouse, dendritic_cell, cytokine, lymphocyte, inflammation, macrophage, tumor_necrosis_factor, antigen, vaccine, nk_cells, lipopolysaccharide, antibody, immunotherapy, necrosis_factor
4	5,43%	413419	Medical Imaging and Diagnosis	cystic, benign, metastasis, dog, imaging, neoplasm, surgery, mass, renal, pancreatic, computed_tomography, soft_tissue, malignant, differential_diagnosis, child
5	5,42%	363604	Nanotechnology and Drug Delivery	nanoparticle, cytotoxicity, binding, peptide, synthesis, photodynamic_therapy, vivo, doxorubicin, conjugate, anticancer, DNA, molecule, neuromuscular, fluorescence, imaging
6	4,96%	406113	Cell Death Pathways and Chemotherapy	apoptosis, protein, kinase, nf_kappab, protein_kinase_b, autophagy, phosphorylation, mitochondrial, death, cytotoxicity, caspase, cisplatin_chemotherapy, mapk, bax_activation, vivo
7	4,76%	431209	Metastasis and Treatment Outcomes	metastasis, prognosis, overall_survival, lymph_node, surgery, recurrence, resection, chemotherapy, heart_rate, bladder, adjuvant, radiotherapy, preoperative, multivariate, hazard_ratio
8	4,61%	325279	Surgical Procedures and Recovery	surgery, resection, postoperative, reconstruction, laparoscopic, endoscopy, flap_surgery, preoperative, intraoperative, excision, perioperative, pain, rectal, defect, hospital_stay
9	4,57%	313305	Quality of Life and Supportive Care	quality_life, pain, survivor, child, breast, exercise, fatigue, palliative_care, physical, interview, training, life, nurse, social, psychosocial
10	4,48%	393129	RNA metabolism	micro_rna, protein, colorectal_cancer, tissue, messenger_rna, metastasis, gastric, overexpression, ribonucleic_acid, epithelial_mesenchymal_transitions, gas_chromatographic, pancreatic, nasopharyngeal_carcinoma, prognosis, apoptosis
11	4,35%	369095	Metabolic Factors and Animal Models	rat, mouse, metabolic, diet, antioxidant, inflammation, enzyme, animal, cytotoxicity, fatty_acid, DNA, oxidative, toxicity, specie, liver
12	4,23%	391005	Angiogenesis and Tissue Remodeling	vascular_endothelial_growth_factor, angiogenesis, stem, mouse, endothelial, metastasis, migration, extracellular_matrix, integrin, hypoxia, adhesion, matrix_metalloproteinase_inhibitors, tissue, vivo, fibroblast
13	4,16%	371971	Epidemiology and Risk Factors	colorectal_cancer, mortality, smoker, confidence_interval, cohort, HIV, colonoscopy, death, heart_rate, lung, relative_risk, hazard_ratio, diet, polyp, sex

For example, the topic named “Immunotherapy and Immune Response” mixed two subjects completely different: immunotherapy and immune response.

Furthermore, if this thematic exploration using topic modelling may correspond to a search for emerging topics, it does not meet all the needs of funders, who would have also liked to have a thematic analysis according to their own nomenclatures. They pointed the Common Scientific Outline (CSO) from International Cancer Research Partnership. This analysis would have been possible with domain classifier from IMT but would have needed more time.

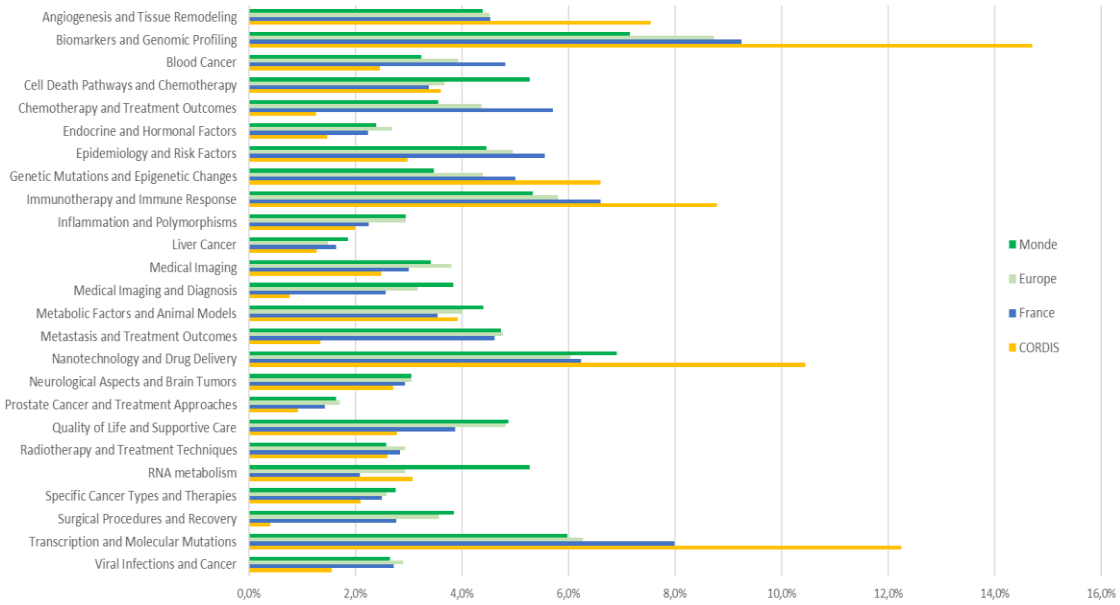
The last part of the workshop was devoted to practice of STI Viewer to analyse the research topics emerging from European community funded projects, in comparison with the research topics observed in France, Europe and the rest of the world.

Figure 21: Screenshots of different STI Viewer views to compare research topics



Exploring topics along various parameters (World, Europe, France and Project funded by European Commission) was of great interest to participants. On the one hand, they appreciated to have the same views / indicators on all parameters and the ability to define the time period and export all the indicators. On the other hand, they would have preferred visualisations of various parameters to be able to compare results, similar to what is shown in the next figure made with Excel, with an export of indicators from STI Viewer.

Figure 22: Scientific publications and publications funded by the European Union in sub-fields of cancer research (2010-2020)



This approach was of great interest to workshop participants.

Finally, the funders once again alerted the project team to the risk of presenting results based on uncontrolled data or tools, particularly in the health sector, which is subject to a great deal of controversy.

5. CONCLUSION

The Health Living Lab was met with high expectations from cancer research funders, in terms of the analysis of the impact of their funded projects. The services and tools proposed by IntelComp and the broad scope of the used data gave rise to those kind of expectations. Adding to this potential, the effort of the Health Living Lab in its stakeholder engagement, a large group representing all main French funders on cancer has been involved in the lab activities.

Implementing the Living Lab to build tools meeting these expectations constituted a rich experiment. First, the initial approach was to mobilise data from OpenAIRE on French cancer research funders' projects, incorporate them in IntelComp's datalake and, ideally, enrich them with additional data from both internal and external databases. However, the link between projects and publications proved to be insufficient due to the lack of data on funding acknowledgements. Hcéres, as Health Living Lab lead, tried to process data from French cancer research funders with AI means provided by IntelComp partners. This approach turned out to be too complex, as the technological development focused on the establishment of more fundamental services. The possibilities offered at the time were limited to European-funded projects and to traditional dimensions, such as information on projects, publications, or patents; data relating to socio-economic impact, which would have been interesting to the engaged stakeholders, were not addressed at all.

Secondly, the co-construction process was complex to set up. Direct interaction between technical teams and end-users on tools prototypes proved difficult, with the primary target group being comprised of policymakers. The first tool available (IMT) turned out to be too technical to mobilise relevant policymakers. On the other hand, the STI Viewer tool needed to have ingested more data on cancer research, to be relevant to policymakers. Having Hcéres act as an interface proved necessary and, ultimately, essential to adapting the tools to the interest of the Living Lab's key stakeholders and integrating the desired data, before presenting the improved tools to policymakers.

Finally, the interaction between the Hcéres team and the technical teams was both necessary and highly instructive, as the IntelComp project was fundamentally driven by a tools development approach. LL Health needed to determine a more flexible and tailored approach than initially foreseen. Through a series of targeted interactions, Hcéres managed to successfully establish a close collaboration with the technical teams to integrate upstream stakeholders' needs. During the final year of the project, these collaborations became constructive and produced several useful conclusions. The analysis of available data on project and publications showed the difficulties in terms of linking the different types of data. Experiments with the semantic matching service for projects and publications became increasingly successful and opened a way to resolve this critical issue. Thematic analyses were a good opportunity to showcase and share innovative results with policymakers on cancer research.

The stakeholders involved in the Health Living Lab have been particularly interested in learning about Hcéres' experience of working with new AI-based tools. They are interested in understanding the encountered problems and proposed solutions, as those are relevant to their own context.

REFERENCES

Beaudoin Christine at al. (2022)	A research agenda for evaluating living labs as an open innovation model for environmental and agricultural sustainability
Brugha, R., & Varvasovszky, Z. (2000)	Stakeholder analysis: a review. Health policy and planning, 15(3), 239-246.
Compagnucci, Lorenzo, Francesca Spirgarelli, José Coelho, and Carlos Duarte (2020)	Living Labs and user engagement for innovation and sustainability
De Vicente López, J. and Matti, C. (2016)	Visual toolbox for system innovation. A resource book for practitioners to map, analyse and facilitate sustainability transitions. Transitions Hub series. Climate-KIC, Brussels 2016.
Decker, Rianne, Juan Franco Contreras, and Albert Meijer (2020)	The Living Lab as a Methodology for Public Administration Research – a Systematic Literature Review of its Applications in the Social Sciences
Dutilleul Benoît, Frans A. J. Birrer, and Wouter Mensink (2010)	Unpacking European Living Labs Analysing Innovation's Social Dimensions
Dutilleul Benoît, Frans A. J. Birrer, and Wouter Mensink (2010)	Unpacking European Living Labs Analysing Innovation's Social Dimensions
Eden, C., & Ackerman, F. (1998)	Making Strategy: The Journey of Strategic Management. London: Sage.
Eriksson, Niitamo, Kulkki, et al. (2005)	State-of-the-art in Utilizing Living Labs Approach to User-Centric ICT Innovation-A European Approach, Lulea: Center for Distance-Spanning Technology. Lulea University of Technology Sweden, Lulea.
Gascó, Mila (2017)	Living labs – Implementing open innovation in the public sector
Greve, Katharina (2018)	Facilitating Co-creation in Living Labs
Greve, Katharina, Riccardo De Vita, Seppo Leminen, and Mika Westerlund (2021)	Living Labs -- From Niche to Mainstream Innovation Management
Hakkarainen, L., Hyysalo, S., 2013	How do we keep the living laboratory alive? learning and conflicts in living lab collaboration. Technol. Innovat. Manag. Rev. 3 (12), 16e22.
Hossain, Mokter, Seppo Leminen, and Mika Westerlund (2019)	A systemic review of living lab literature
Leminen, S., Nyström, A.G., Westerlund, M., Kortelainen, M.J., 2016	The effect of network structure on radical innovation in living labs. J. Bus. Ind. Market. 31 (6), 743-757

Leminen, Seppo, und Mika Westerlund (2012)	Towards innovation in Living Labs networks. *International Journal of Product Development* 17, Nr. 1–2 (1. January 2012): 43–59. https://doi.org/10.1504/IJPD.2012.051161.
Mastelic, Joëlle, Marlyne Sahakian, and Riccardo Bonazzi (2015)	How to keep a living lab alive
Mendelow, A. L. (1981)	Environmental Scanning -The Impact of the Stakeholder Concept.
Paskaleva, Krassimara, & Ian Cooper (2021)	Are living labs effective
Schaffers and Turkama (2012)	Living Labs for Cross-Border Systemic Innovation https://doi.org/10.22215/timreview605
Torma, Veronika (2020)	Analysing stakeholder engagement – stakeholder involvement in urban living labs and the main processes needed to establish a living laboratory

ANNEX: Material of kick off meeting

Note presenting IntelComp project to stakeholders sent before kick off meeting

OBSERVATOIRE DES SCIENCES ET TECHNIQUES

18 Janvier 2022

Tessa Enock-Levi
(OST)

PROJET INTELCOMP : LE LIVING LAB SANTE

Le projet INTELCOMP

INTELCOMP est un projet européen du programme H2020 qui se déroule de 2021 à 2023. Il élabore une plateforme de données relatives aux activités de recherche et d'innovation ainsi que des outils d'IA permettant d'analyser ces données de manière transverse. Il s'agit notamment de mieux suivre et mesurer les impacts de financements de projets de recherche. Trois Living Labs sont conçus pour tester l'intérêt de cette plateforme auprès des utilisateurs potentiels dans trois domaines, l'intelligence artificielle, le changement climatique et la santé.

L'OST est en charge de la mise en place et de l'animation du Living Lab regroupant des acteurs de la recherche en santé pour tester les usages potentiels de l'outil et orienter les développements. Ce Living Lab sera officiellement lancé au premier trimestre 2022 et les travaux auront lieu en 2022 et 2023.

Le Living Lab Santé se concentre sur le domaine du cancer. La thématique du (des) cancer(s) a été retenue pour plusieurs raisons :

- elle représente un enjeu de santé majeur et à ce titre fait l'objet d'un axe du programme cadre Horizon 2020 avec une mission transversale dédiée ;
- Les recherches sur le cancer se développent sur l'ensemble du spectre, des plus fondamentales à la clinique ;
- elle implique des acteurs variés du domaine de la santé - recherche publique, entreprises pharmaceutiques (nouveaux traitements) et techniques (outils de diagnostic), associations de patients.

Coeur du Living Lab Santé

Dans le cadre de l'étude récente réalisée par le CNCR et la FHF sur la place de la France en recherche en cancérologie[1], le lien entre production scientifique et financement n'a pas pu être analysé. La plateforme INTELCOMP semble particulièrement prometteuse pour intégrer des données sur les projets financés (via OpenAire ou autre) et faire le lien avec la production scientifique issue de ces projets et son impact (analyse des publications, essais cliniques, brevets, citations dans les clinical guidelines…)

Compte-tenu des premiers échanges que l'OST a eu avec différentes institutions et chercheurs dans le domaine de la santé et du cancer, il est proposé d'explorer les possibilités de la plateforme INTELCOMP en terme d'analyse des « chemins d'impact » des projets financés. Il s'agit de suivre et caractériser les différentes étapes qui mènent de la sélection du projet, en amont, à différents résultats de la recherche, puis à leurs impacts socio-économiques.

[1] Quelle est la place de la France en recherche en cancérologie ? Analyse de la production scientifique 2010-2019 France & Monde ; CNCR-FHF

Un premier niveau d'analyse consiste à caractériser de manière large la production scientifique (« output ») des projets financés en matière de :

- Publications scientifiques
- Brevets déposés
- Essais cliniques réalisés.

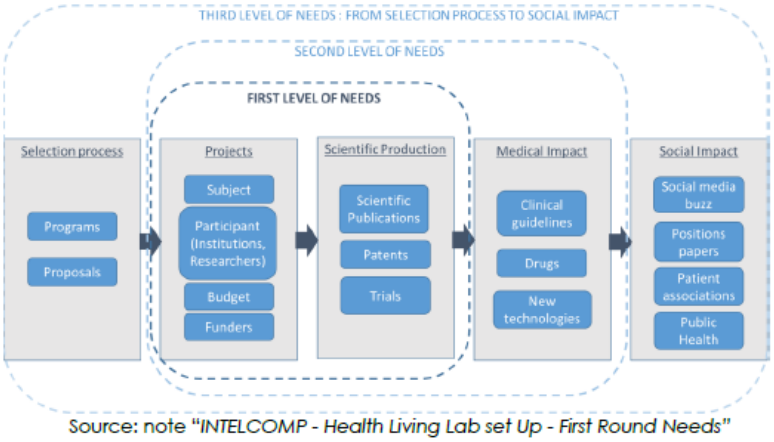
Un deuxième niveau d'analyse consiste identifier et caractériser l'impact (« outcomes ») médical des projets de recherche en matière de :

- Bonnes pratiques (Citations dans les guides de bonnes pratiques / clinical guidelines)
- Nouveaux traitements (industrie pharmaceutiques)
- Nouvelles techniques de diagnostic dépistage (industriels / startup)

Enfin, le troisième niveau d'analyse consiste à identifier et caractériser l'impact social des projets financés:

- Impact médiatique (média, réseaux sociaux)
- Thématiques des projets financés le plus souvent reprises dans les *position papers*
- Thématiques des projets financés correspondant aux les attentes des associations de patients ;
- Positionnement des projets par rapport aux données de santé publique (incidence, mortalités, qualité de vie des patients etc.).

Les différents niveaux d'analyse des chemins d'impact



Source: note "*INTELCOMP - Health Living Lab set Up - First Round Needs*"

L'analyse du chemin d'impact pourra être croisée avec le type d'acteurs impliqués à chaque étape : recherche publique, entreprises, associations de patients.

Plusieurs approches thématiques liées au cancer paraissent intéressantes à étudier avec la plateforme INTELCOMP:

- Analyse par type de recherche (fondamentale / Clinique / translationelle) ;
- Identification des coopérations en recherche (coopérations internationales, inter-institutions, partenariats public-privé etc.)

- Analyse par discipline ou thématique
- Focus sur certains types de cancer (localisation cancéreuse), pronostic (bon / mauvais) ou encore incidence
- Focus éventuel sur les problématiques adjacentes au cancer et en lien avec la santé publique : tabac, alcool, alimentation, pollution…
- Identification des nouveaux traitements et technologies de rupture (génétique, biothérapies, médecine prédictive, e-santé)
- Caractérisation des projets / travaux concernant les différentes étapes de la prise en charge du patient 1) la prévention; 2) la détection précoce; 3) le diagnostic et le traitement; et 4) la qualité de vie des patients atteints d'un cancer et des personnes ayant survécu à la maladie.

Fonctionnement du Living Lab Santé

La méthodologie du projet prévoit un dispositif de co-construction avec trois types d'utilisateurs potentiels:

- Les « décideurs des politiques de recherche »,
- Les « acteurs du système de recherche » (chercheurs, industriels…)
- Les acteurs de la « société civile » (citoyens, associations de patients…).

Le dispositif de co-construction avec les « décideurs des politiques de recherche » constitue le cœur du projet et sera lancé en premier. La consultation des autres cibles sera définie plus précisément fin 2022, en fonction des résultats et des opportunités identifiés suite aux travaux engagés avec les « décideurs des politiques de recherche».

Le déroulement du dispositif de co-construction avec les « décideurs des politiques de recherche » est en construction et sera défini précisément avec les participants pour s'ajuster à leurs besoins. Le living Lab s'enrichira des contributions et attentes des partenaires.

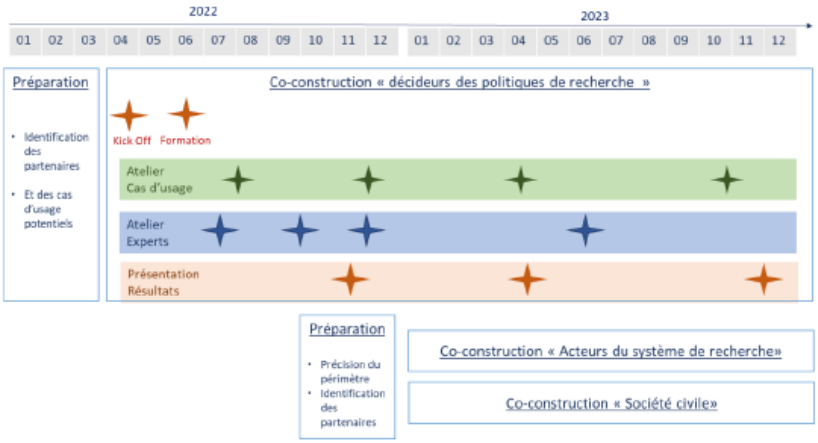
Nous pouvons envisager plusieurs manières de participer pour les partenaires :

- Proposer un cas d'usage spécifique à instruire dans le cadre du Living Lab : Dans ce cas, le partenaire participe au projet en tant qu'utilisateur final de l'analyse, comme le destinataire d'une étude classique. L'étude est réalisée par l'OST en collaboration avec les équipes techniques d'INTELCOMP. Le partenaire est impliqué dans le déroulement de l'étude en particulier en phase de cadrage puis pour la validation de certaines étapes.

- Participer à des ateliers d'experts pour ajuster la plateforme INTELCOMP à la thématique du cancer : identification de sources de données spécifiques, constructions de nomenclatures thématiques ad hoc…

- Assister aux réunions de présentation des principaux résultats et participer à l'analyse de ces résultats et au retour d'expérience de la plateforme INTELCOMP.

Planning envisagé



Minutes sent to the stakeholders after the meeting

COMPTE RENDU DE REUNION

PROJET INTELCOMP
RÉUNION OST / PARTENAIRES EXTERNES 14/02/2022

Participants

- ANR (Martine Garnier-Rizet ; Marianne Lanoë)
- FRM (Florence Guibal ; Valérie Lemarchandel ; Hugo Juraver)
- ITMO Cancer (François Castagner)
- FONDATION ARC (Sandrine Daubeuf)
- EFS (Angélique Michaut)
- INCa (Carla Estaquio)

- OST (Frédérique Sachwald, Tessa Enock-Levi, Dominique Guellec, Luis Miotti, Aurélien Leynet, David Sapinho)

Objet : réunion de présentation du projet INTELCOMP et proposition de participation au LIVING LAB Santé centré sur le cancer.

Compte-rendu rédigé par : Tessa Enock Levi / Aurélien Leynet

Documents associés : Présentation INTELCOM OST 140222

Présentation du Projet INTELCOMP

Le projet INTELCOMP est un projet européen financé dans le cadre du programme H2020 dont l'objectif est de construire une plateforme de données concernant les activités de recherche et d'innovation ainsi que des outils d'IA permettant d'analyser ces données de manière transverse. Trois Living labs doivent permettre de tester, auprès des utilisateurs potentiels, l'intérêt de cette plateforme pour suivre et orienter les politiques de recherche dans trois domaines : l'intelligence artificielle, le changement climatique et la santé.

L'OST est en charge de la mise en place et de l'animation du Living lab regroupant des acteurs de la recherche en santé pour tester les usages potentiels de l'outil et orienter les développements. Ce Living lab débutera en janvier 2022 et les travaux auront lieu en 2022 et 2023.

L'OST a proposé de centrer le Living lab Santé sur le cancer et sur l'analyse des chemins d'impact des projets financés, c'est-à-dire la séquence des étapes menant de la sélection des projets de recherche jusqu'à la mise en œuvre de leurs résultats scientifiques, techniques et sociétaux.

Résumé de l'échange entre les participants

Les partenaires expriment leur intérêt pour le projet IntelComp et leur souhait de suivre l'avancement du ***Living lab* Santé**. Le living lab en est à son lancement, c'est donc le moment de s'y associer et l'orienter pour en tirer le meilleur parti.

Les partenaires réunis lors de cette présentation pourraient constituer le « core group » des participants au Living lab santé qui serait étendu à d'autres institutions, en particulier européennes.

Ils soulignent le besoin de préciser les **données** qui y seront disponibles ; dans sa formulation actuelle, la plateforme IntelComp pourra mobiliser à la fois des données ouvertes et des données propriétaires dont l'usage sera réservé aux participants fournissant ces données.

Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur

Le périmètre des données qui seront intégrées à la plateforme reste à définir et s'appuiera sur :

1. Les données intégrées à la plateforme issues des données ouvertes utilisées dans les outils existants (Corpus Viewer, Data4impact, Open Aire).
    - On peut par exemple pointer les données dont l'intégration est en cours d'étude : données projets issues de Cordis et du NIH, données de publications de PubMed, données brevets de PATSAT, données sur les essais cliniques issues de ClinicalTrials.gov, certaines données issues d'OpenAire, données sur les médicaments (DrugBank)

2. Les données ouvertes identifiées par le Living Lab Cancer.
    - Sources de données potentielles dont l'intégration pourra être étudiée : données projets issues d'Open Aire (23 financeurs), données projets des partenaires du LL (si elles ne sont pas dans Open Aire), données complémentaires sur les essais cliniques (Cochrane library ?), données sur les clinical guidelines (identification spécifique dans PubMed), données de sante (Global Burden Disease)

3. Des données propriétaires si besoin pour compléter l'analyse.
    - En particulier l'OST pourra voir comment mobiliser les données du Web of Science (dont l'OST maintient une version en interne) si les données de publications de PubMed ne sont pas suffisantes ou pour les compléter. Le Living lab pourra aussi étudier comment mobiliser des données de financement qui ne sont pas intégrées aux bases de projets ouvertes

Les partenaires montrent leur intérêt pour plusieurs cas d'études qui pourront s'appuyer sur la plateforme Intelcomp:

- Le suivi de l'impact médical et sociétal des projets suscite beaucoup d'intérêt
    - La temporalité de l'impact pose un défi tout particulier pour la recherche fondamentale et sera à prendre en compte pour le choix de l'historique des données sur lesquelles travailler
    - L'impact sociétal peut être appréhendé de plusieurs manières, comme par exemple en mesurant :
        - La mise en œuvre de thérapeutiques nouvelles liés à des domaines de recherche
        - La reprise des connaissances scientifiques nouvelles hors du milieu scientifique (« connexion avec la société »)
        - La correspondance entre la production scientifique et les orientations politiques ou encore les attentes sociales (par exemple alignement de l'effort de recherche avec les priorités du plan Cancer)

- Une autre dimension de l'impact soulevée par les participants est l'impact des projets financés sur l'écosystème de la recherche :
    - Impact sur la structuration des communautés scientifiques
    - Impact sur les carrières des jeunes chercheurs financés

- Enfin un autre cas d'usage de la plateforme Intelcomp évoqué ultérieurement (revue de littérature) concerne l'identification des ruptures scientifiques dans un domaine de recherche. Ce cas d'usage correspond à l'orientation retenue par le Living Lab sur l'IA. En fonction de l'avancée des travaux du Living Lab Santé, du temps et des ressources disponibles, cette approche pourra être éventuellement testée.