LUCIA Understanding Lung Cancer related risk factors and their Impact

Horizon Europe Grant Agreement Number: 101096473

Deliverable Number	D6.13
Deliverable Title	Policy brief formulating recommendations based on the research and innovation strand of the 'Understanding' annual cluster meeting-M36
Due date of deliverable	31.12.2025
Actual Submission Date	23.12.2025
Responsible partner	TECH
Contributors	All partners, projects within 'Understanding' cluster
Revision (draft, 1, 2, …)	1.0
Dissemination Level	PU

Start Date of the project: January 1, 2023

Duration: 48 months

Document information\ Revision History

Description \ Status	Revision date	Authors
1st Draft	26.11.25	Lopez Adrian (MELCAYA)
2nd Draft	04.12.25	Baruch Polis (LUCIA)

Executive Board Document Sign Off

Role	Partner	Signature	Date
Project coordinator	TECH	Baruch Polis	17/12/2025
WP1 Lead	ULSTER	Jonathan Wallace	22/12/2025
WP2 Lead	VICOM	Alba Garin-Muga	22/12/2025
WP3 Lead	TECH	Baruch Polis	17/12/2025
WP4 Lead	BB	Jon Eneko Idoyaga Uribarrena	22/12/2025
WP5 Lead	UHEI	Jonathan Sleeman	22/12/2025
WP6 Lead	HOPE	Marie Nabbe	23/12/2025
WP7 Lead	TECH	Liat Tsuri	17/12/2025

Executive summary

This policy brief summarises recommendations from the third "Understanding (Risk Factors & Determinants)" cluster meeting to strengthen cancer prevention, early detection and equity access in the context of the EU Cancer Mission. It calls for support on interoperable, FAIR-by-design cancer data ecosystems that connect project-level resources with national cancer data nodes and emerging European hubs such as EUCAIM, UNCAN and CANDLE, ensuring sustainable reuse of legacy and new datasets. The report urges EU policymakers to prioritise integrated risk-stratification tools and minimally invasive biomarkers, backed by trustworthy AI, federated analytics and shared quality standards, to accelerate translation into screening, surveillance and personalised care. It further recommends that EU-level action on cancer inequalities adopt a broader perspective, including prevention, diagnostics, treatment and survivorship, while systematically embedding patient and citizen voices and clear governance frameworks for digital health and AI. Finally, the brief proposes that the European Commission and Member States should use these suggestions to guide future legislation, funding programmes and mission-oriented initiatives, ensuring that scientific advances are matched by equitable access and long-term infrastructure support.

Contents

Acronyms & abbreviations

Term	Description
AI	Artificial Intelligence
API	Application Programming Interface
CAR-T	Chimeric Antigen Receptor T-cell
CAYA	Children, Adolescents and Young Adults
CDISC	Clinical Data Interchange Standards Consortium
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CT	Computed Tomography
DCAT-AP	Data Catalogue Application Profile
DICOM	Digital Imaging and Communications in Medicine
eCRF	Electronic Case Report Form
EFPIA	European Federation of Pharmaceutical Industries and Associations
EGA	European Genome-Phenome Archive
EHDS	European Health Data Space
EHR	Electronic Health Records
EOSC4Cancer	European Open Science Cloud for Cancer
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
GP	General Practitioner
GWAS	Genome-Wide Association Study
HCC	HepatoCellular Carcinoma
IARC	International Agency for Research on Cancer
KPI	Key Performance Indicator

MGUS	Monoclonal Gammopathy of Undetermined Significance
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NAFLD	Non-Alcoholic Fatty Liver Disease
OMOP	Observational Medical Outcomes Partnership
QC	Quality Control
SMM	Smoldering Multiple Myeloma
VOC	Volatile Organic Compounds
VPN	Virtual Private Network

1 Introduction

The third "Understanding (Risk Factors and Determinants)" annual cluster meeting was organized by the MELCAYA project and held in Barcelona (Spain) on the 15th of October 2025. During this meeting, cluster projects reviewed collaborative work on data management, convergence on healthcare data standards and planned for long-term interoperability with emerging EU infrastructures such as UNCAN and EUCAIM. Scientific sessions showcased cross-project results on genetic susceptibility, exposome, omics and functional models, as well as risk stratification and early diagnosis tools, including AI-based models, smart sensors and proteomic signatures. The goal of this document will be to report on the general recommendations based on the work presented and discussions held during this meeting.

2 Recommendations based on the research and innovation (R&I) activities

2.1 Sharing and agreeing on common practices for data/material management

All projects acknowledge the need for secure, well-governed platforms. LUCIA and DISCERN already run controlled "safe haven" environments (LUCIA's Health Data Platform, IARC SIT/Innovica), where partners access de-identified data via VPN and role-based permissions, working with tools like RStudio and Jupyter and without free data export. MELCAYA and ELMUMY centralise core clinical and omics datasets, relying on strict pseudonymisation, REDCap/eCRFs and institutional infrastructures. GENIAL is more decentralised but is harmonising legacy cohorts and ontologies to enable comparability. FAIR principles are widely endorsed but unevenly implemented, with all projects planning or already making use of persistent identifiers and community repositories (e.g. EGA for genomics). Interoperability standards such as OMOP, CDISC and FHIR are being adopted in some cases, while others are still selecting frameworks. A shared challenge is integrating legacy cohorts and heterogeneous lab data, whereas standard omics and large-scale imaging are already relatively well structured for reuse.

Overall, the common recommendations proposed to advance work in this area are the following:

1. **Establish cluster-wide metadata and ontology harmonisation:** the projects should create a common minimum metadata and variable set (for clinical, exposure, omics and imaging data), aligned with existing standards (OMOP, CDISC, FHIR and domain ontologies), and share practical templates between technical teams. This directly responds to the identified gap of limited metadata harmonisation across the cluster.
2. **Converge on a coherent repository and access strategy:** building on current plans (EGA, EUCAIM, EOSC4Cancer, etc), the cluster should define disease and data type specific repository choices,

with harmonised access governance and model data use agreements and access committee procedures.

3. **Strengthen technical interoperability and tooling:** projects already developing APIs, toolkits and synthetic data (e.g. LUCIA's platform and metadata harvesting) can share reusable components with the rest of the cluster, supporting common pipelines for data ingestion, quality control, pseudonymisation and FAIR metadata generation.
4. **Formalise cross-project governance and legal alignment**: given the shared constraints of GDPR and national regulations, the data management group should further align consent language, re-use conditions and risk-mitigation measures and document best practices around controlled access environments and de-identification strategies.

The next steps proposed by the cluster working group include:

- Creation of a metadata harmonisation task force to deliver a shared metadata schema, map guidelines for legacy variables and coordinate ontology choices across cancers and platforms.
- Development of a sustainability roadmap for repositories and platforms, clarifying long-term hosting (e.g. migration into UNCAN/EOSC4Cancer nodes or institutional infrastructures), funding models and responsibilities for updates, versioning and user support beyond each project's lifetime.
- Launch joint demonstrator use cases, for example, cross-cohort analyses of environmental and lifestyle exposures or shared AI pipelines for early detection across melanoma, lung cancer, liver cancer, multiple myeloma and other cancer types. This would showcase the added value of interoperable data while stress-testing governance and technical solutions.
- Perform data-stewardship workshops and shared documentation (covering standards, repositories and legal/ethical templates) to ensure that the rich datasets generated by the cluster projects remain findable, accessible under clear conditions, interoperable and truly reusable for future cancer research.

2.2 Sharing and cross-comparison of risk factors and molecular features

All projects emphasise that many cancers are preventable by managing exposures and early lesions. GENIAL links most hepatocellular carcinoma to modifiable risks such as alcohol, obesity, viral hepatitis and other exposures. LUCIA has found strong geospatial predictors of lung cancer risk, including radon,

pollution and limited green space. MELCAYA shows how UV exposure, skin phenotype and MC1R variants drive melanoma in the young. DISCERN characterises the exposome, mapping pollution, climate and toxic agents to footprints in kidney and other cancers. ELMUMY shows that microenvironmental and molecular signals in MGUS/SMM progression may be modifiable. All projects highlighted gene-environment interactions: large GENIAL cohorts dissect germline modifiers of liver cancer risk, MELCAYA and LUCIA link inherited variants and methylation to UV and smoking signatures, DISCERN connects environmental maps to tumor mutational patterns and ELMUMY integrates multi-omics and clinical data. Together, they argue that multi-omics and advanced models are needed to move from correlation to causation and identify actionable targets.

The common recommendations proposed to advance work in this area are the following:

1. **Develop and refine integrated risk models:** the projects collectively recommend combining clinical, genetic, epigenetic and exposomic variables into robust risk scores. Examples include HCC risk models that merge alcohol/NAFLD status with germline variants (GENIAL), melanoma tools that integrate monogenic and MC1R/polygenic risk for young patients (MELCAYA), geospatial-plus-omics models for lung and kidney (LUCIA, DISCERN), and clinical/flow/urine-based risk stratifiers for MGUS/SMM (ELMUMY).
2. **Prioritize assays and pipelines that are close to routine practice:** tools that rely on standard clinical data should be prioritized, as they can be widely adopted and scaled. ELMUMY's classifiers based on common clinical variables and urine peptidomics, as well as its flow-cytometry markers of progression, are an excellent example of this pragmatic approach. Similarly, LUCIA and GENIAL emphasize leveraging existing clinical cohorts and registries rather than only experimental datasets.
3. **Strengthening cross-project synergies and shared infrastructure:** several speakers explicitly called for deeper collaboration: comparing molecular drivers such as MDM2 and stemness pathways across cancers, aligning environmental layers (pollution, radon, toxic plants, UV) and sharing experimental systems (organoids, premalignant models or CRISPR screens) and pipelines across projects.

The next steps in this area include the following points:

- Complete and exploit large cohorts and biobanks: finalise recruitment and sequencing in MELCAYA's cohort, fully analyse GENIAL's expanded GWAS and rare-variant sets, mine DISCERN's population cohorts and tumor series with additional exposomic layers and complete ELMUMY's multi-omics and trial preparations.

- Validate candidate biomarkers and targets in independent datasets and functional models, including MC1R and other melanoma risk variants, epigenetic and geospatial lung risk markers and ID1/ID3 inhibition, HCC risk variants and gene-environment interactions, ELMUMY's protein markers (MDM2 and CAR-T targets) and DISCERN's exposomic-signature links in new geographic areas.
- Translate findings into prevention and early-intervention strategies: including risk-adapted screening or intensified monitoring for high-risk MGUS/SMM, young melanoma patients, and heavily exposed liver and lung populations, study pharmacological clearance of premalignant lesions (as piloted in LUCIA) to reduce future cancer incidence and evaluate public health measures to curb key exposures such as alcohol misuse, UV overexposure, air and soil pollution, radon, obesogenic environments and toxic herbal practices.
- Maintain a cross-project analytical and data-sharing platform so that the rich multi-cancer datasets and tools produced by the cluster projects remain reusable for future mechanistic studies and for the broader Cancer Mission community beyond the lifetime of the individual projects.

2.3 Collaboration in technology, tools, knowledge and best practices for data exploitation and computational modelling

All projects are building ecosystems for data, models and AI rather than isolated analyses. ELMUMY's cloud warehouse integrates multi-layer clinical, omics, cell line and animal data with AI models. GENIAL is developing an open, explainable deep-learning pipeline using histology models to predict HCC risk from cirrhotic biopsies. LUCIA's Health Data Platform connects clinical records, imaging, omics, environment and sensors across the full AI lifecycle. MELCAYA operates a secure pan-European digital pathology hub for CAYA melanoma. DISCERN makes uses of IARC SAT to host omics and case-series data with controlled, standardised pipelines. All projects are relying on multimodal, longitudinal and pre-symptomatic data, insisting on trustworthy, reproducible and reusable AI. For instance, ELMUMY is building ML pipelines and open biomarker methods, GENIAL open-sources and benchmarks histopathology models, LUCIA is validating risk, EHR and imaging models across health systems, MELCAYA is documenting image variability to guide standards and DISCERN is evaluating R tools for central and federated analyses.

The common recommendations proposed to advance work in this area are the following:

1. **Consolidate shared AI-ready infrastructure across projects:** the proposed platforms (such as GENIAL's pathology pipeline, LUCIA's Health Data Platform or MELCAYA's Halo Link) should define common practices for user roles, logging, dataset versioning and model catalogues, in order to enable cross-project reuse and joint demonstrators.
2. **Promote standardisation and quality control for complex data types:** building on MELCAYA's detailed assessment of slide quality and scanner parameters and DISCERN's standardised omics pipelines, all projects should converge on minimal QC checklists and reporting standards for images, omics and clinical/exposure variables.
3. **Embed explainability and clinical feasibility from the beginning:** risk scores based on routine variables (LUCIA), interpretable gene-ratio signatures (ELMUMY) and visual attribution maps in histopathology (GENIAL) are good examples of this approach. Models should be co-designed with clinicians, favouring variables and workflows that can realistically be adopted in practice.
4. **Strengthen federated and privacy-preserving analytics:** given that parts of the cohort data in DISCERN and other projects cannot be centralised, federated methods that approximate pooled analyses while respecting local constraints should be further developed and shared across the cluster.

As next steps in this topic, the following points have been proposed:

- Complete data collection and integration of the remaining experimental and clinical datasets and systematically feed them into the respective platforms and warehouses.
- Advance multi-omics and multi-scale modelling, for example integrating ELMUMY transcriptomics and proteomics with single-cell networks and drug-repurposing signatures, coupling GENIAL's biopsies with genomics, combining LUCIA's CT and omics and extending DISCERN's pipelines to additional omics layers.
- Prospectively validating and benchmarking AI models in independent cohorts and clinical settings, including progression predictors in MGUS/SMM, liver-cancer risk models, lung-cancer risk and EHR models, CT nodule classification and future melanoma and exposome-related tools.
- Deliver cluster-level demonstrators, such as shared dashboards or pipelines that operate across multiple projects, showcasing how harmonized platforms and trustworthy AI can support cancer prevention, early detection and personalized management on a European scale.

2.4 Cross-comparison and integration of risk stratification/early diagnosis tools

All projects show that non or minimally invasive biomarkers can provide powerful prognostic information. MELCAYA and LUCIA use volatile organic compounds from breath and skin, captured by nanomaterial-based sensors, to detect disease-specific signatures in melanoma and lung cancer. DISCERN demonstrates that large-scale blood proteomics can reveal hundreds of prognosis-related proteins across several cancers, GENIAL combines circulating proteins, genetics, imaging and clinical variables to refine HCC surveillance, while ELMUMY uses routine MGUS/SMM clinical data with survival ML to predict progression. There is strong convergence on AI as a driver of superior risk models: ELMUMY's models outperform existing scores, DISCERN's analyses uncover shared pathways of poor prognosis, GENIAL integrates clinical, genetic, imaging and serum markers for better HCC risk stratification, while LUCIA and MELCAYA translate sensor signals into accurate classifiers. All projects underline that successful translation requires rigorous validation, harmonised protocols, and integration into real care pathways, including prospective trials and inter-site standardisation.

The common recommendations proposed to advance work in this area are the following:

1. **Scale and diversify cohorts while standardising pre-analytical conditions:** all teams recommend completing recruitment and increasing diversity (centres, ages, stages) while enforcing strict protocols for sampling (fasting, fragrance restrictions, handling times, storage) to ensure that VOC, proteomic and clinical signals are robust and comparable across sites.
2. **Favour models and tests that are clinically feasible and explainable:** the most promising tools rely on accessible inputs such as routine blood tests and bone marrow metrics (ELMUMY), standard clinical variables plus simple scores (GENIAL), peripheral blood proteomics (DISCERN) and quick breath/skin patch measurements (LUCIA, MELCAYA). Models should provide clear risk categories and interpretable drivers (e.g. key proteins, sensor features), to support adoption in guidelines rather than remaining black-box experiments.
3. **Build shared analytical pipelines and cross-cancer biomarker panels:** since many prognostic proteins and systemic processes appear common across cancers, DISCERN's findings argue for cross-disease panels rather than isolated markers. Similarly, the same VOC technologies are already being piloted in melanoma and lung cancer. The projects collectively recommend shared pipelines, code and quality metrics so that methods proven in one disease can be rapidly tested in others.

4. **Integrate economic and organisational aspects early:** GENIAL's linkage of risk stratification with MRI cost-effectiveness, and the emphasis on portable, low-cost devices in LUCIA and MELCAYA, underline the need to consider resource use and workflow integration as part of development.

Regarding next steps, the following points are proposed:

- Complete datasets and refine models, including the completion of recruitment and follow-up in MELCAYA's melanoma VOC monitoring study, expanding DISCERN's proteomics and metadata (particularly for pancreatic cancer), adding further patients and omics layers to ELMUMY, completing inclusion and biomarker collection in GENIAL's fast-track trial and scaling up LUCIA's VOC and spectrometry-on-card datasets.
- Perform a prospective and external validation in real-world settings, as all groups plan to test their signatures and devices in independent cohorts or new sites, comparing performance against current standard-of-care tools for risk stratification, surveillance and early detection across cancers.
- Integrate the developed tools into concrete clinical pathways, such as risk-adapted HCC surveillance, MGUS/SMM follow-up or lung cancer screening refinement and melanoma monitoring.

2.5 Sharing of best practices on implementation of healthcare policies

All three projects pointed to important support gaps for patients and families. MELCAYA's work on the CAYA melanoma journey shows poor early recognition, fragmented referral pathways and limited psychological and social support, leading to a substantial emotional and financial burden on families. DISCERN's global patient surveys similarly reveal unmet needs around guidance, participation in decisions and access to expert centres and trials in kidney and digestive cancers. LUCIA's inequalities activities emphasise that disadvantaged groups are least likely to benefit from new screening and early-detection tools without deliberate policy action. A second shared conclusion is that AI and digital technologies bring great promise but also ethical, legal and social risks. MELCAYA has developed a structured framework to assess digital health and AI tools for melanoma across clinical and non-clinical dimensions, including transferability along the technology life cycle. LUCIA is piloting an AI impact assessment tool that automatically analyses project documentation to flag risks related to bias, explainability and transparency and propose mitigation actions. DISCERN demonstrates how structured evidence (e.g. patient surveys) can be translated into scientific and policy briefs used in dialogues with European and national authorities.

The common recommendations proposed to advance work in this area are the following:

1. **Systematically embed the patient voice in policy and technology design:** MELCAYA's patient-led tasks and focus groups and DISCERN's use of global surveys, show that patient organisations should be formal partners in guideline development, priority setting and evaluation of new tools.
2. **Develop clear, shared standards for AI and digital health governance:** building on MELCAYA's assessment framework and LUCIA's AI impact assessment, the cluster should adopt common criteria on fairness, transparency, accountability, data protection and acceptability for vulnerable groups such as CAYA patients.
3. **Address structural inequalities and workforce capacity:** all different projects point to a weak attention to rare and early-onset cancers in national plans, insufficient training of GPs and dermatologists and uneven access to innovation, calling for targeted professional education and equity-oriented policy action.

As future work, the main goal will be to develop shared policy and practice tools. For instance, MELCAYA is finalising a Delphi process structured in thematic sections (information and awareness, legal/ethical aspects, digital health, organisation of care, inequalities) to generate consensus recommendations for CAYA melanoma policies and technology use. DISCERN is updating and expanding scientific and policy briefs on kidney and early-onset digestive cancers to be used with public authorities and professional societies as a basis for adapting screening strategies, follow-up and survivorship care. LUCIA will also further develop and test its AI impact-assessment tool in multiple projects, supporting anticipatory governance of AI-enabled screening and early detection. Together, these efforts aim to deliver a coherent European agenda where digital innovation, AI and new models of care are co-designed with patients, implemented fairly across regions and age groups, and continuously evaluated for their real impact on cancer prevention, early diagnosis and quality of life.

3 Recommendations for addressing inequalities and citizen engagement

The cluster working group on these topics provided evidence from several sources (European Commission, EFPIA or EuroHealthNet) showing major disparities in access to prevention, screening, innovative diagnostics, treatments, survivorship services and even smoking prevention, strongly influenced by social determinants and workforce shortages. At the same time, patient and citizen engagement is seen as indispensable to address these gaps. Shared decision-making improves

adherence, survival and quality of life. Patient experience should also be considered to shape relevant research questions, study designs and policy priorities, while engaging citizens can support sustainable funding for cancer research. Coordinated communication has proven to be a strategic tool against inequalities. Cluster videos, brochures, project summaries, podcasts and social-media collaborations help make complex research and policy debates accessible and visible to wider audiences, especially when aligned with key events such as World Cancer Day or European Week Against Cancer.

The common recommendations proposed to advance work in this area are the following:

1. **Structure policy work around the full cancer pathway:** the inequalities cluster has distilled three overarching policy areas, (1) equal access to prevention and early detection, (2) equal access to diagnostics and innovation and (3) equal access to treatment, care pathways and survivorship (including workforce and the role of AI). These axes should guide future recommendations and joint activities.
2. **Systematically integrate patient and citizen voices:** shared-decision initiatives, advocacy workshops and closer collaboration with patient organisations across projects should become standard, ensuring that research priorities, clinical pathways and communication products reflect real needs.
3. **Professionalise and coordinate communication across the cluster:** regular coordination between dissemination officers, common social-media roadmaps, shared KPIs (posts, views, engagement), and reusable formats like podcasts and short video interviews can maximise impact and avoid fragmented messaging.

Regarding future work, planned actions include performing a social media campaign on inequalities in November 2025, followed by three thematic webinars (with accompanying campaigns) between 2026–2027 on awareness in prevention/screening, breaking the diagnostics divide and access to treatment and survivorship for all. These will feed into a consolidated policy-recommendation report in 2027. As for citizen engagement, the cluster will expand multimedia outreach (with new podcasts for each project plus a patient-organisation episode, updated videos and a brochure) and pilot or scale up patient-centred workshops and social-media training models across consortia. Finally, the working group is planning to integrate additional evidence sources (such as new IARC work on cancer inequalities) and to strengthen the collaboration between patient organisations and research projects, creating a sustained European platform where equity and engagement are embedded in every phase of cancer research and care.

4 Recommendations for the collaboration with the European federated cancer research data hub initiatives

All three initiatives aim to overcome fragmentation and make project legacy data reusable. EUCAIM is tackling the proliferation of separate imaging projects by creating a federated "atlas" of cancer images, with reference nodes and different clear tiers of compliance levels for data holders. UNCAN-Connect is building a decentralised network with a central platform and national cancer data nodes that connect to thematic and institutional data providers through shared governance and technical services. CANDLE is focusing on the national layer, helping member states design and implement functional cancer data nodes aligned with the European Health Data Space (EHDS). All these initiatives show that governance, legal and ethical frameworks are as important as technology. EUCAIM combines detailed technical onboarding (handbooks, maturity questionnaires and checklists) with strong legal support to handle data transfer, federated sharing, data-sharing agreements and ethical compliance. UNCAN-Connect is setting up an interim governance model that will evolve into a full framework for data governance, compliance and operations, driven by real use cases in six cancer types. CANDLE explicitly positions itself as an implementer of EU policy (EHDS and related initiatives), with close links to Commission policy officers and joint actions on personalised medicine, comprehensive cancer centres and cancer registries. All projects stress that user needs and co-creation with existing projects are essential. UNCAN-Connect repeatedly emphasised that use cases and data providers should drive requirements and validation, not the other way around. CANDLE builds on the practical Dutch national node experience, where quarterly meetings bring users real-life challenges together with infrastructure experts. The "Understanding (Risk Factors & Determinants)" cluster projects are explicitly invited to act as extra use cases, helping shape guidelines for data processing at source, as well as data deposition pathways and interoperability.

The common recommendations proposed to advance work in this area are the following:

1. **Use a layered hub-and-node model instead of isolated project repositories:** projects should plan to deposit or expose their data through recognised infrastructures (e.g. EUCAIM for imaging, EGA and other domain repositories, future UNCAN platforms) and, where possible through national cancer data nodes.
2. **Converge on shared standards, metadata and FAIR workflows:** EUCAIM already enforces imaging formats (such as DICOM), a common data model, DCAT-AP-based metadata and a rich

"hyperontology", plus tools for de-identification, harmonisation and FAIRness assessment. UNCAN-Connect and CANDLE will provide resource kits and guidelines for data processing at source, interoperability across member states and EHDS-compatible metadata catalogues. Cluster projects are encouraged to align with these emerging profiles from the outset.

3. **Invest in skills, communication and support for data holders:** setting up local or national nodes requires specialised technical, legal and organisational capacity. EUCAIM responds to this demand with detailed handbooks, onboarding workflows and training (e.g. Moodle environment). CANDLE and UNCAN-Connect stress the need for clear technical requirements, common language between clinicians and data experts, and national communities that bring registries, hospitals, infrastructures and ministries into the same room.
4. **Preserve and integrate legacy data and project outputs:** rather than starting from scratch, the initiatives recommend systematically onboarding legacy cohorts, biobanks and project datasets (e.g. EOSC4Cancer, Understanding cluster projects) into the new infrastructure, via either transfer to reference nodes or federated exposure through national nodes.

5 Conclusions

The third annual "Understanding (Risk Factors and Determinants)" cluster meeting confirms that the projects are collectively building a coherent, forward-looking agenda for cancer prevention, early detection and policy translation in Europe. Taken together, their work shows that interoperable data ecosystems, integrated risk models and trustworthy AI can transform heterogeneous clinical, omics, imaging and exposomic resources into actionable tools for both research and care. The meeting also highlighted that many cancers and precursor conditions are, in principle, preventable or interceptable if modifiable exposures, early lesions and high-risk groups are systematically identified and managed.

At the technical level, the cluster converges on secure platforms, FAIR data practices, shared standards and federated analytics as foundations for sustainable impact. The recommended shift from isolated project repositories towards layered national and European nodes (aligned with initiatives such as EUCAIM, UNCAN-Connect and CANDLE) will help preserve legacy data, support reuse and reduce duplication of effort. In parallel, the projects prioritise clinically feasible biomarkers and AI models that rely on routine data, minimally invasive samples and explainable outputs, increasing the chances of real-world uptake.

Equally important, the meeting reaffirmed that innovation must be embedded within equitable health systems and robust governance. Patient and citizen engagement, attention to inequalities, and clear frameworks for digital health and AI are recognised as essential conditions for responsible deployment. By combining technical harmonisation, cross-project demonstrators, shared policy tools and close collaboration with European data infrastructures and patient organisations, the cluster is well positioned to deliver long-lasting benefits for cancer research, prevention, early diagnosis and survivorship across diverse populations in Europe.