

Large Data for Personalised Screening

The EUCanScreen Approach

Maiju PANKAKOSKI
Finnish Cancer Registry

On behalf of the WP5 team

European-Taiwan
Cancer Screening Conference 2026

Riga 7th Oct 2026



Co-funded by
the European Union

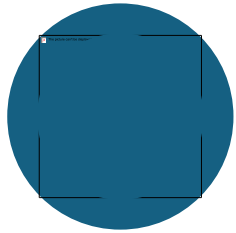


*Data harmonisation for reliable
multi-country analysis is important
even
(or especially?) when things don't
go according to the plan.*

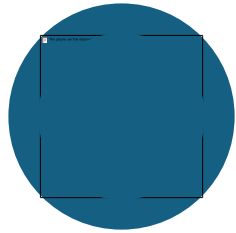


Data for Personalized Screening

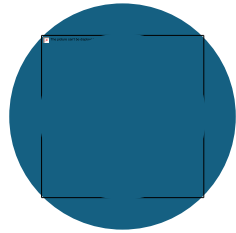
Personalized cancer screening requires a multi-layered data framework, including variables about



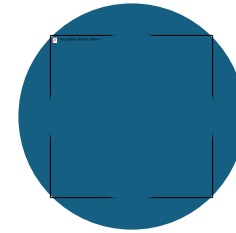
Traditional
demographics



Genetic risk
markers



Clinical / Screening
history



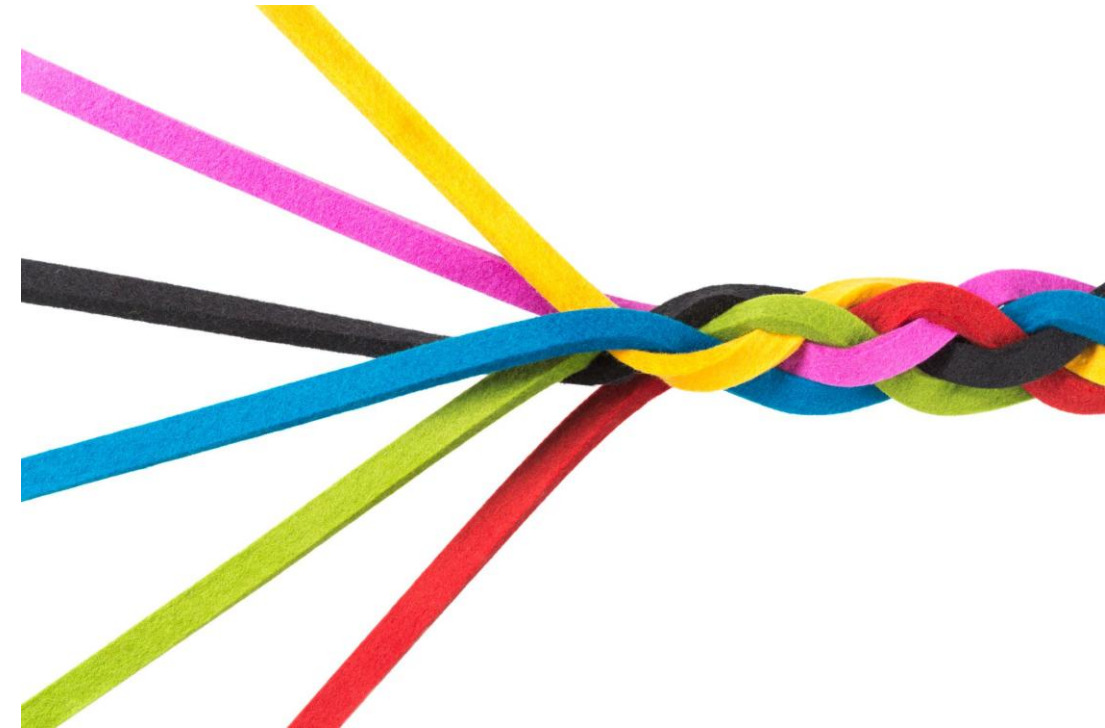
Lifestyle factors

Information is important for tailoring screening intervals and modalities

Data for Personalized Screening

Multinational studies would strengthen the evidence base:

- Although screening datasets are large, cancer outcomes are rare
 - 1-2 cancers per 1 000 screened in CRC screening
- Detailed stratification reduces subgroup size and statistical power
- However, legal changes have made cross-border secondary data sharing more difficult



EUCanScreen WP5 aims to

- Improve and develop data sharing and standardisation
- Improve monitoring and comparison of screening programs
- Enable **cross-border research** using screening data

T5.1: Ethical, Legal, and Social implications (ELSI) of data sharing (Dypedes, Greece)

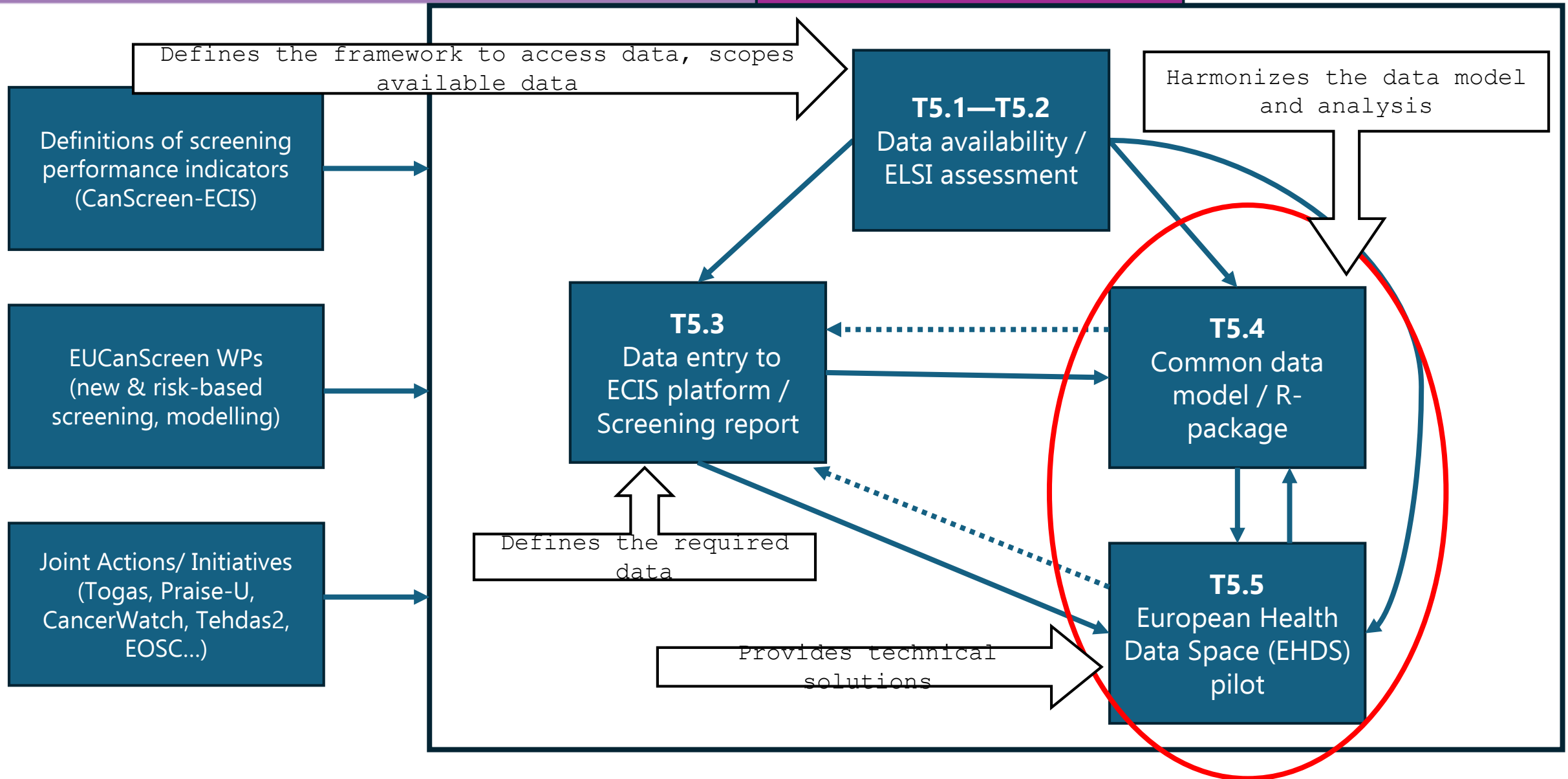
T5.2: Assessment of data availability (CPO, Italy)

T5.3: Extended pilot of data entry to the ECIS platform (CPO, Italy)

T5.4: Software and standardised data model for screening (CSF, Finland)

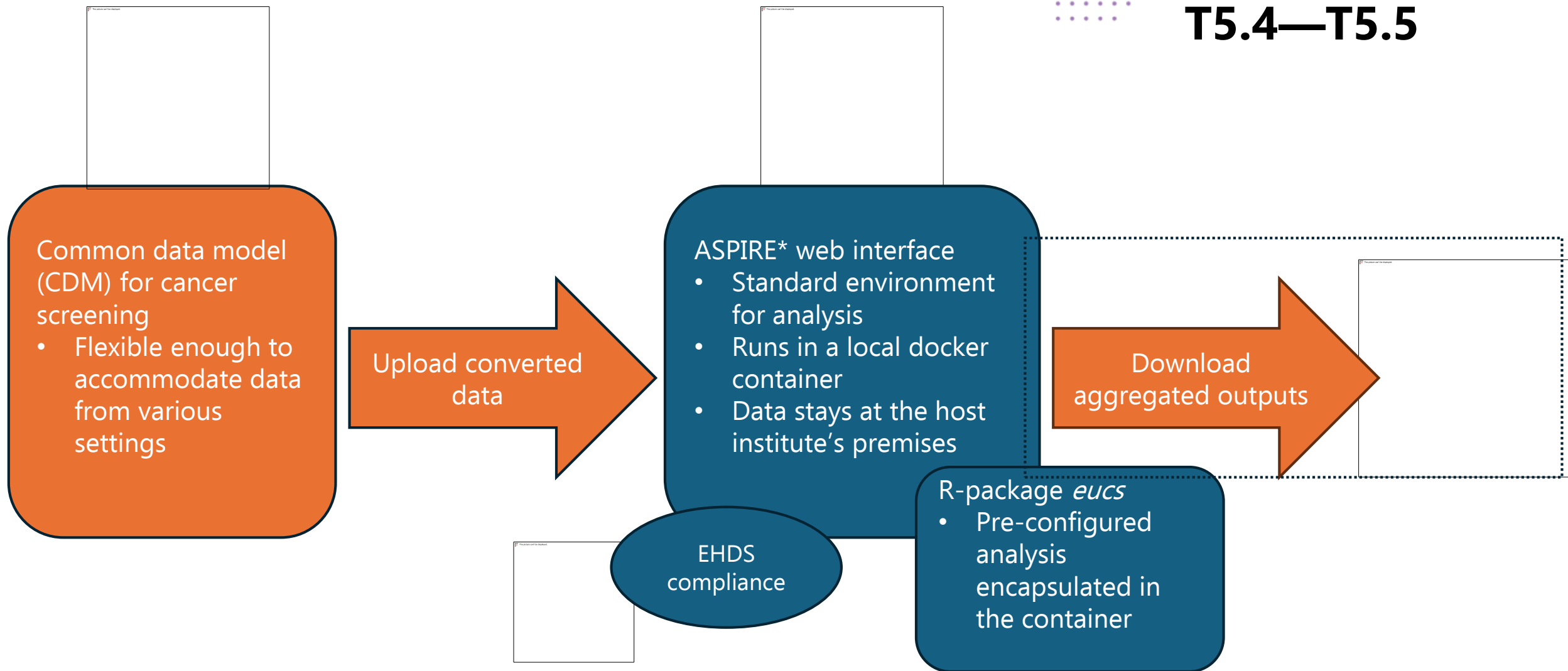
T5.5: Piloting for European Health Data Space (EHDS) (IACS, Spain)

WP5



Individual-level analysis

T5.4—T5.5



*ASPIRE = Analytic Software Pipeline Interface for Reproducible Execution: <https://github.com/cienciadedatosysalud/aspire>

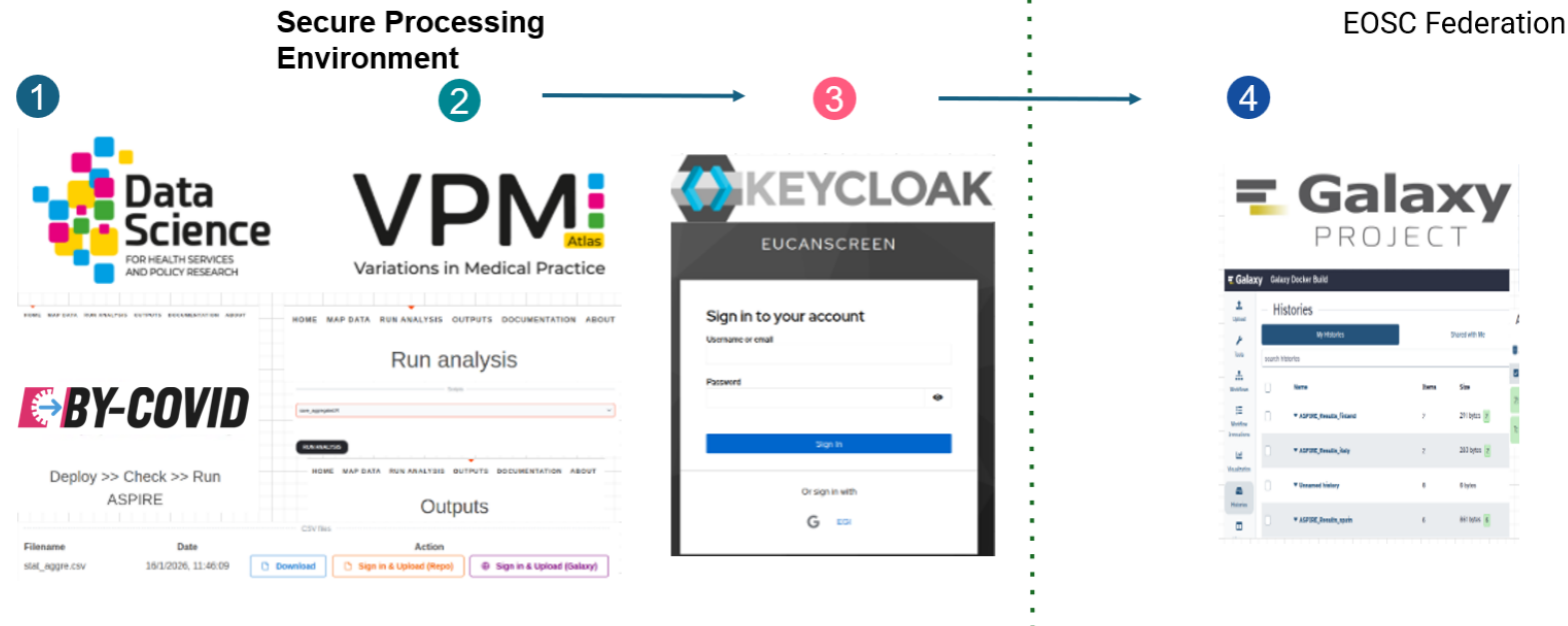
Future: EOCS EU node

Beyond EUCanScreen:
Authenticated automatic
push

- Send live pieces of information for federated learning

Scenario 2

Push data



 **Funded by
the European Union**
Grant agreement ID: 101188045
Grant agreement ID: 101215594

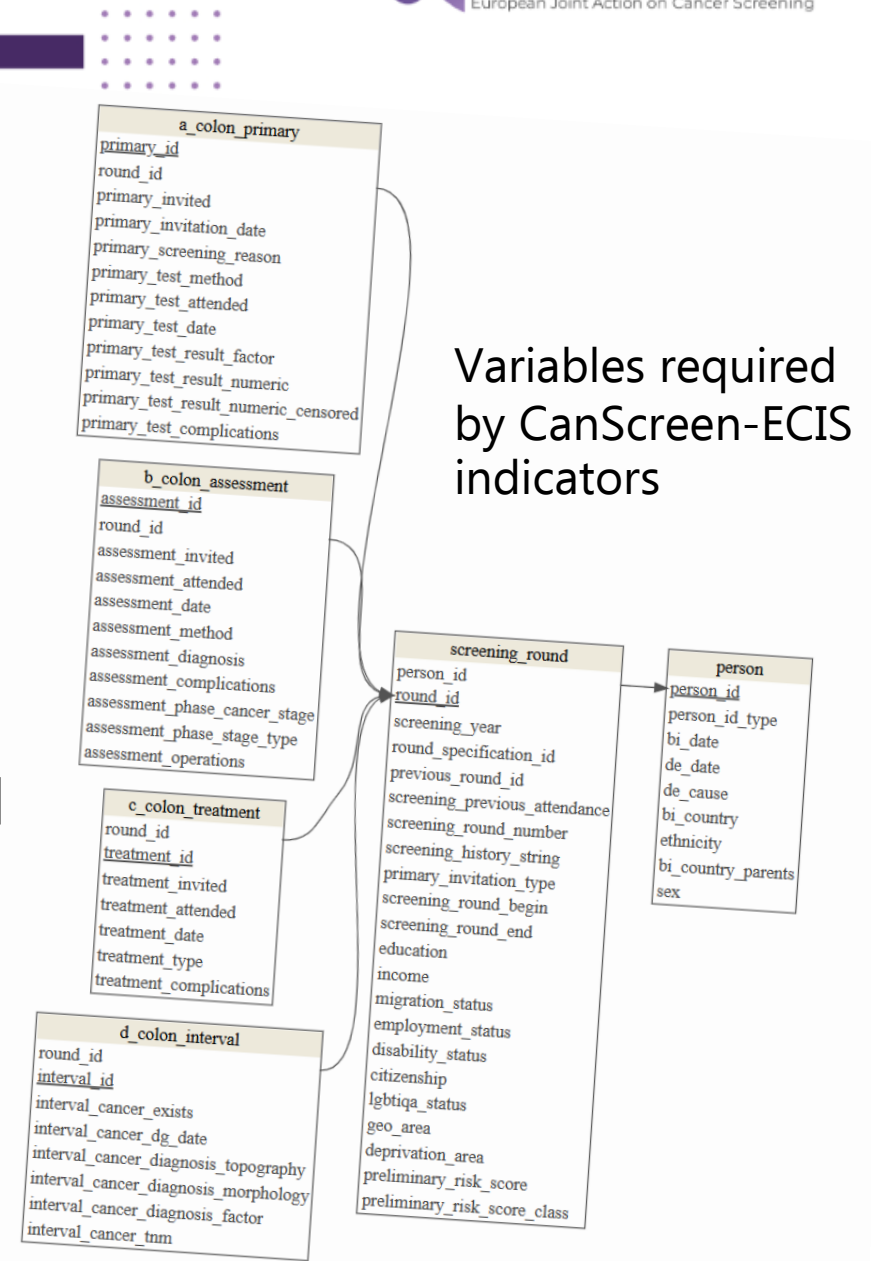
- 1 **Aspire: Analytic Software Pipeline Interface for Reproducible Execution**
- 2 **Aspire: Generate CSV with script results**

- 3 **Authenticated "Push-to-Galaxy"**
- 4 **Sending the results to Galaxy**

Credits: Juan González-García

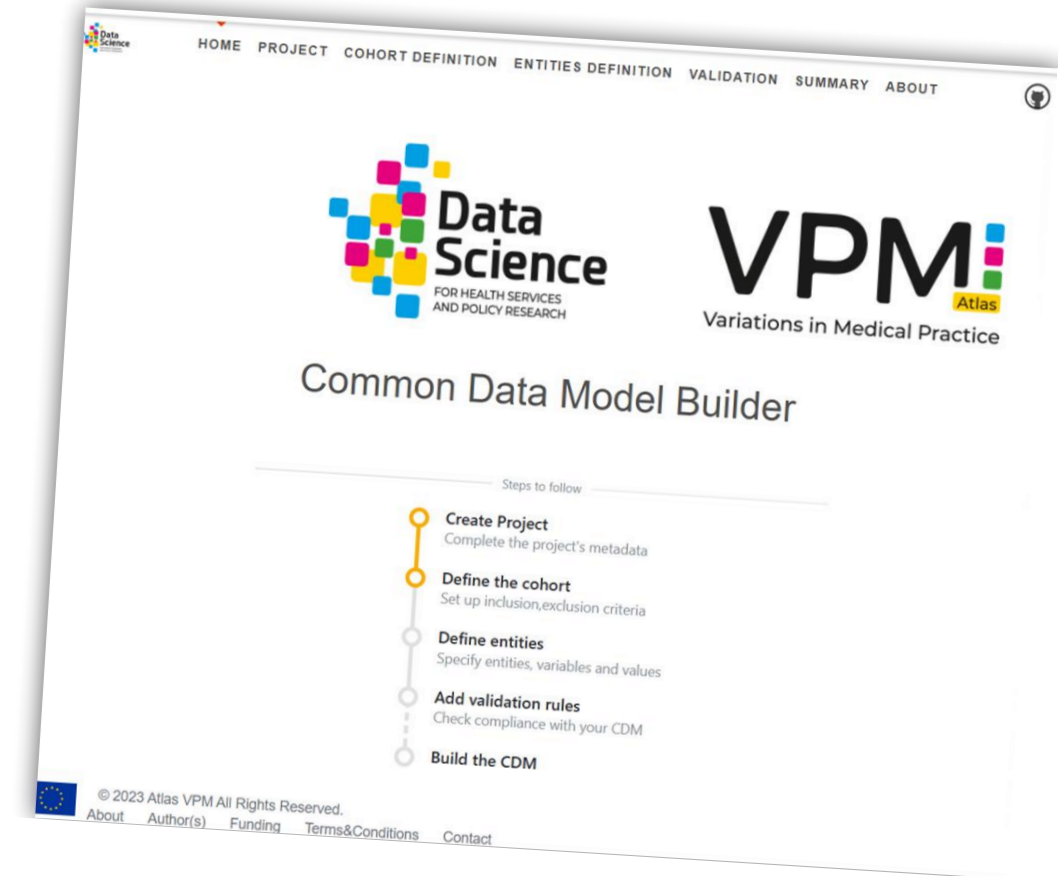
The data models

- Three CDMs: breast, cervix, colon.
- CDM based on the idea of five screening phases.
 - Primary
 - Triage
 - Assessment
 - Treatment
 - Interval
- Phases are linked together to form a screening round that is linked to person data
- Phases can be omitted or unlinked



Preparation by the WP5 team

- 1) Enter the data model to CDMB*
CDMB creates simulated data for testing and validation protocol for data
- 2) Plug CDMB output and R-package (EUCanScreen) into ASPIRE
ASPIRE is containerised using Docker, and the image will be shipped to data providers

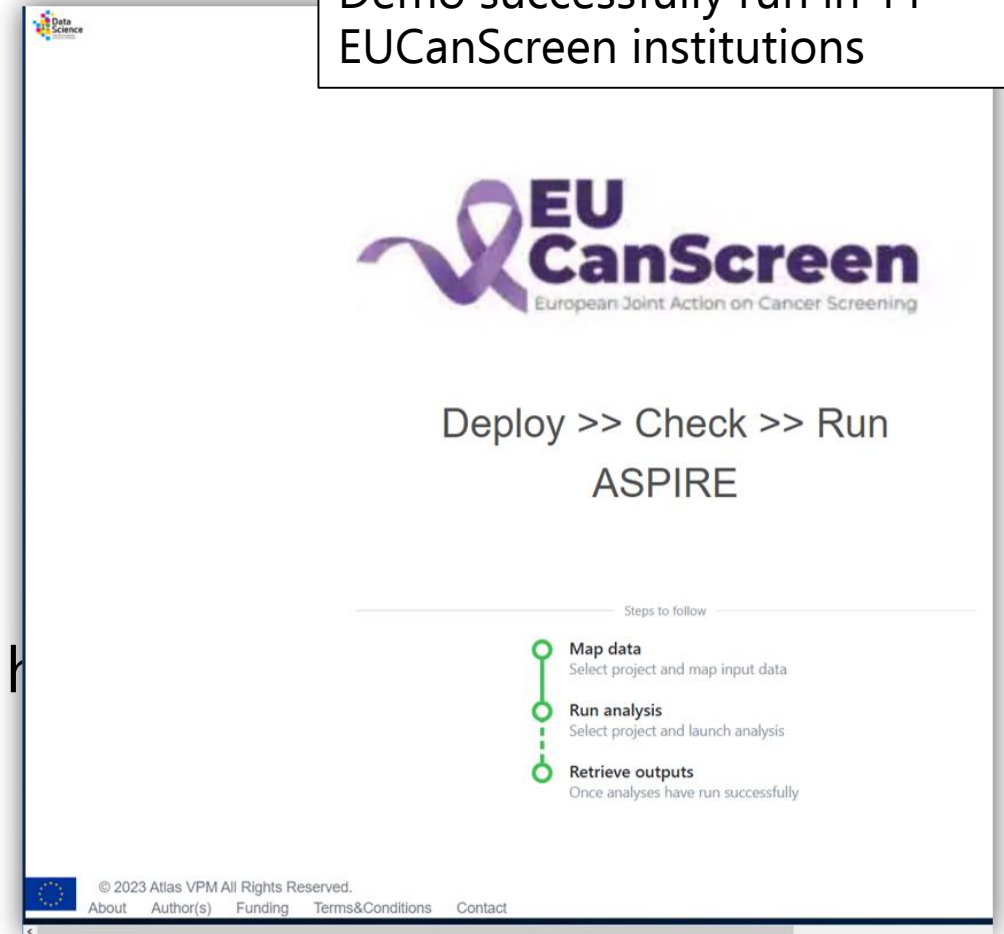


* CDMB=Common Data Model Builder: <https://github.com/cienciadedatosysalud/cdmb>

Execution by the end-user

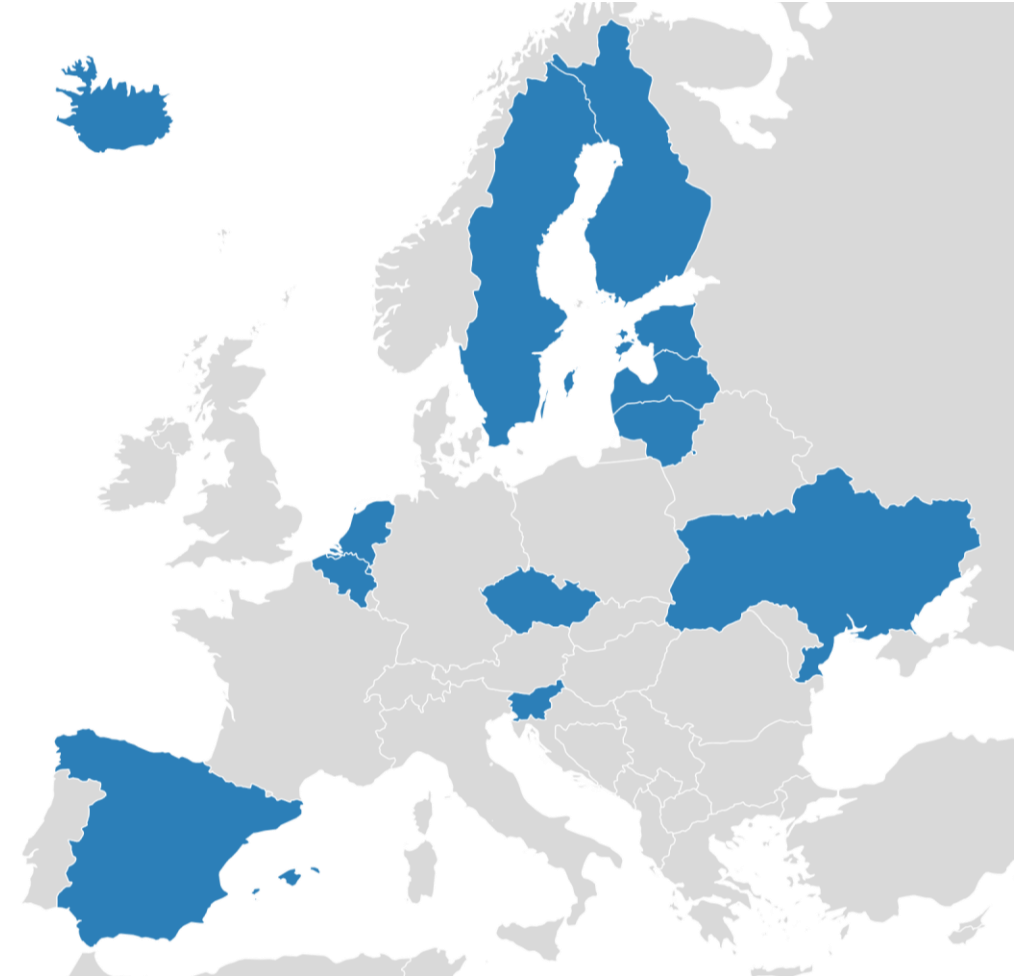
- Host institution enters CDM formatted data in ASPIRE using GUI
- ASPIRE opens to localhost in web browser
- Produces data validation reports and runs desired scripts/reports in standardised environment
- Only person preparing the analysis (WP5 team) has to do programming

Demo successfully run in 11 EUCanScreen institutions



Piloting the pipeline

- EUCanScreen WP5 partners from 15 institutions and 12 countries agreed to test the analysis pipeline with real screening data
 - breast, cervical, colorectal, prostate, and gastric cancer screening
- Feedback on experiences is expected by the end of this year



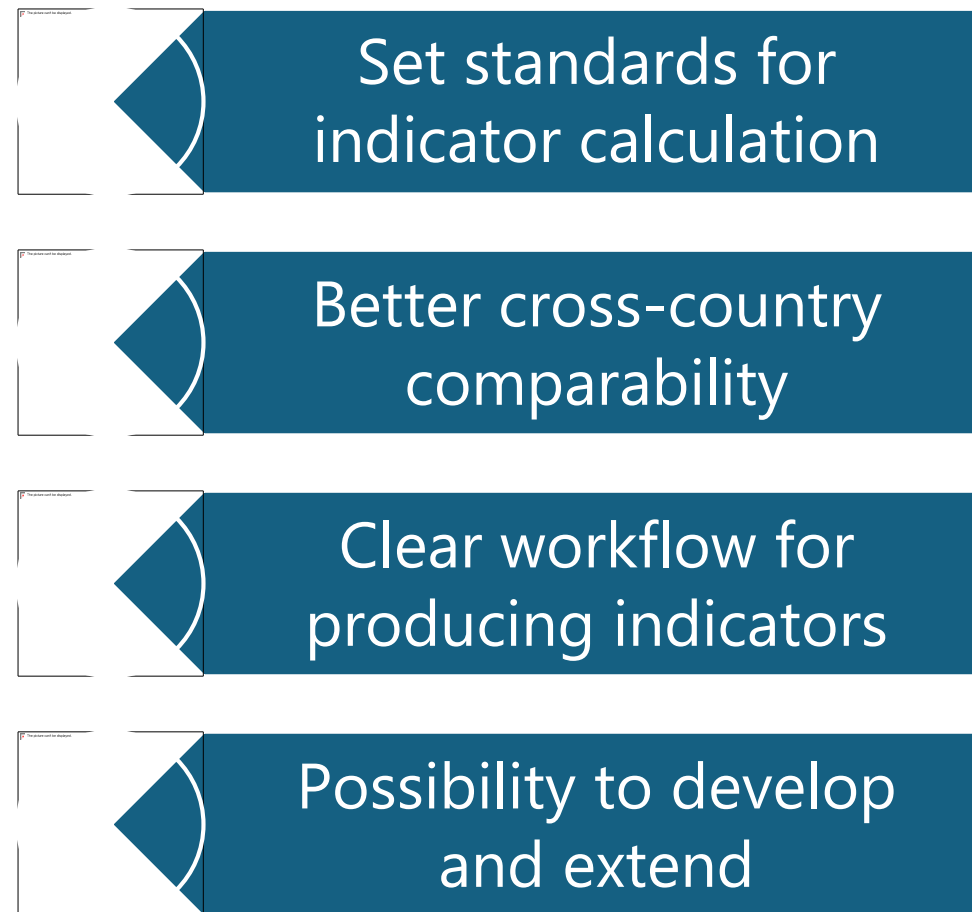
Piloting the pipeline

- Small subset of the total planned variable set
- Variables mostly site-agnostic
- Indicators include invitation/screening coverage, participation rate and test result
- The results can be benchmarked against existing national statistics



Template for monitoring

The CDM and standardized analysis offer a template for monitoring



Template for monitoring



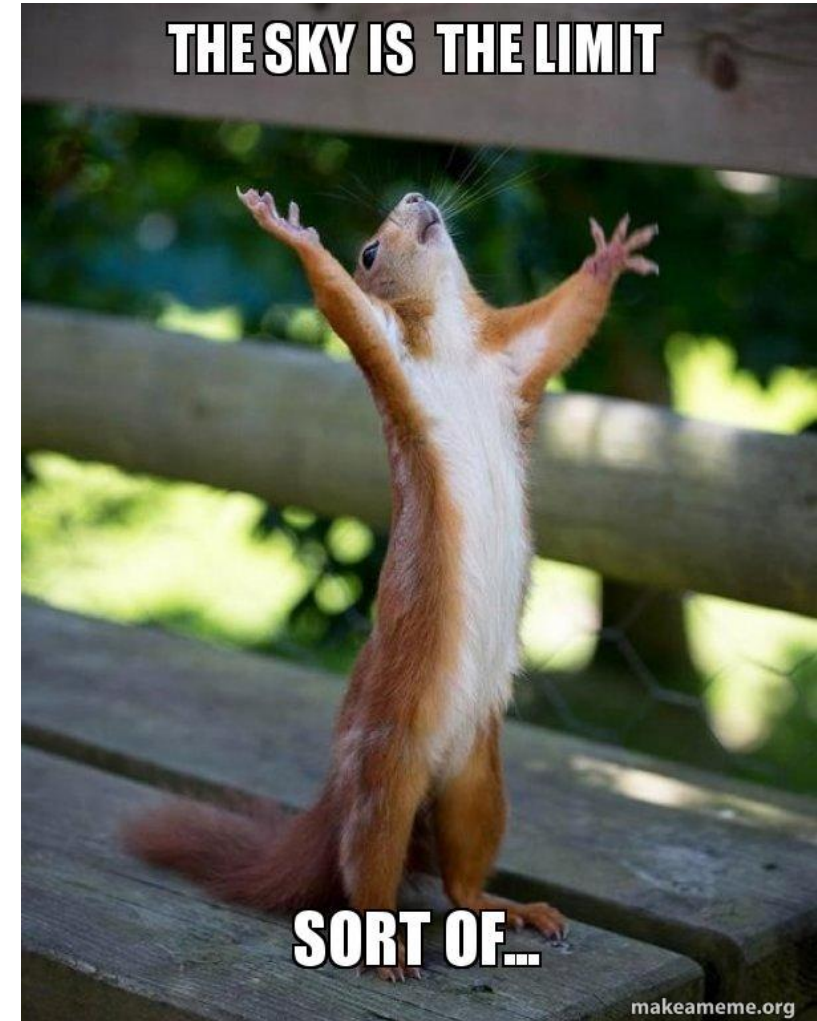
- Example: Screening attender characteristics relevant for targeting personalized screening, using the EUCanScreen CDM variables
- Indicators stratified by
 - Screening history (participation, positivity)
 - Sociodemographic factors (SES, education, immigration status)
 - Risk status (based on, e.g., risk score)
- Screening history could also be derived from the data, if sufficient follow-up exists

sex	age group	education	previous attendance	previous positivity	total (N)	test positive (N)	test positive (%)
1	50–54	2	1	1	100	30	23
1	50–54	2	1	0	900	90	9
1	50–54	2	0	NA	500	75	13
...	...	...	...	...	...	...	...

Future: template for research



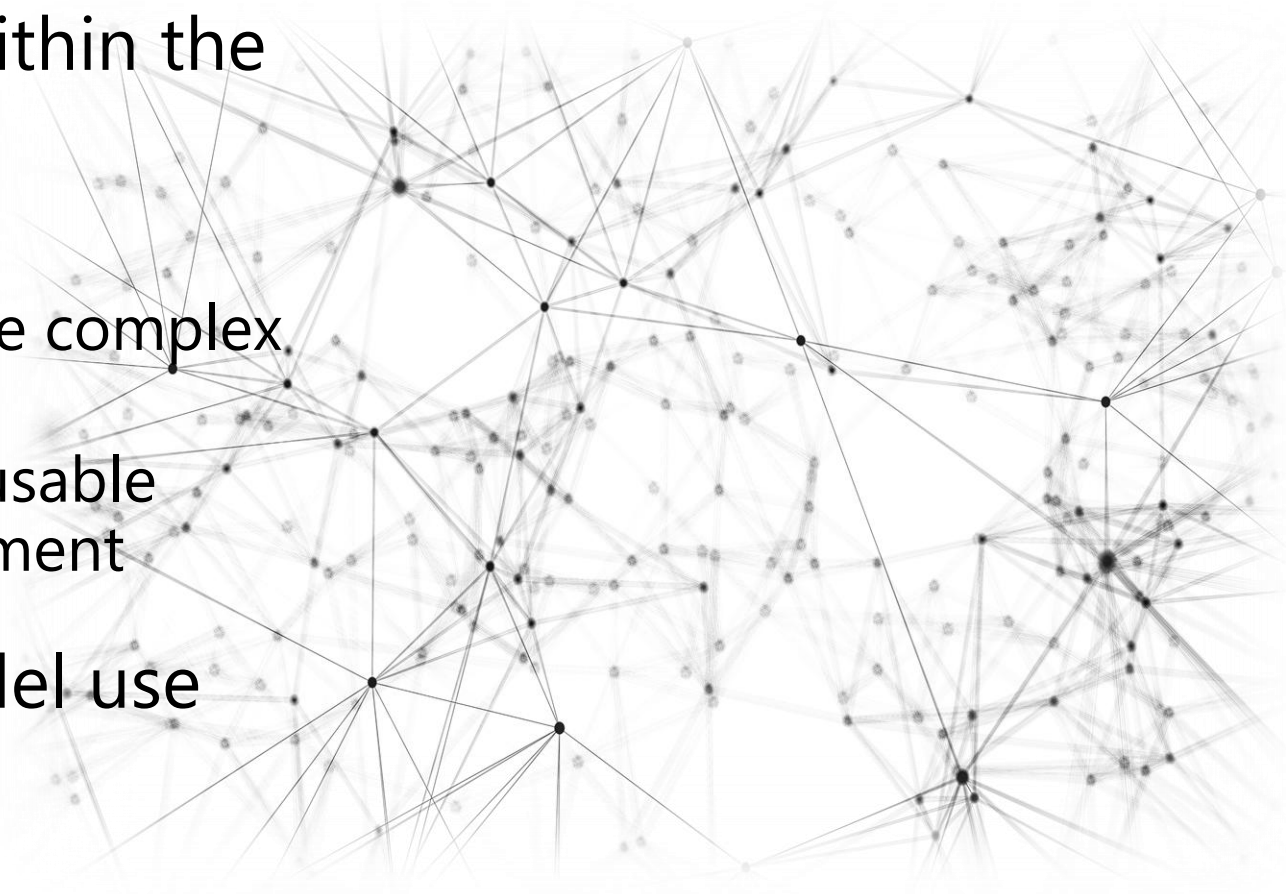
- Any data model can be added to ASPIRE with the Common Data Model Builder (CDMB)
- Common data model + standard analysis scripts
 - Template for practically any federated analysis research project



Future: template for research



- Standardise complex modelling within the same pipeline:
 1. Convert data to the CDM
 2. Replace monitoring scripts with the complex modelling script (or program)
 3. Generate analysis outputs, e.g., reusable model objects for further development
- Reduces barriers for complex model use and scale models better



Challenges

- Converting data: One size does not fit all
 - Need to have a flexible CDM and iterative implementation
- Coordination and development
 - Need continuous resources
- Software permissions and processes vary across hospitals, SPEs, and institutions

Biggest challenge:

Get screening data holders to run the system

Synergies – Data sharing



- EOSC (European Open Science Cloud)
 - Presented push-to-central storage approach, others being considered
- CancerWatch
 - Data quality labels
- Tehdas2
 - Metadata, EHDS compliance



Hot take: It is not that difficult, for personalized screening. The most difficult indicator may be the screening coverage, which is not (that) relevant for most personalised pathways.





Thank you

Aapeli Nevala, Maiju Pankakoski, Carlo Senore, Paola
Armaroli, Alberto Montolío Marco, José Javier Rabal Otal,
Tytti Sarkeala, Juan González García

aapeli.nevala@cancer.fi



Co-funded by
the European Union

Disclaimer

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor the granting authority can be held responsible for them.