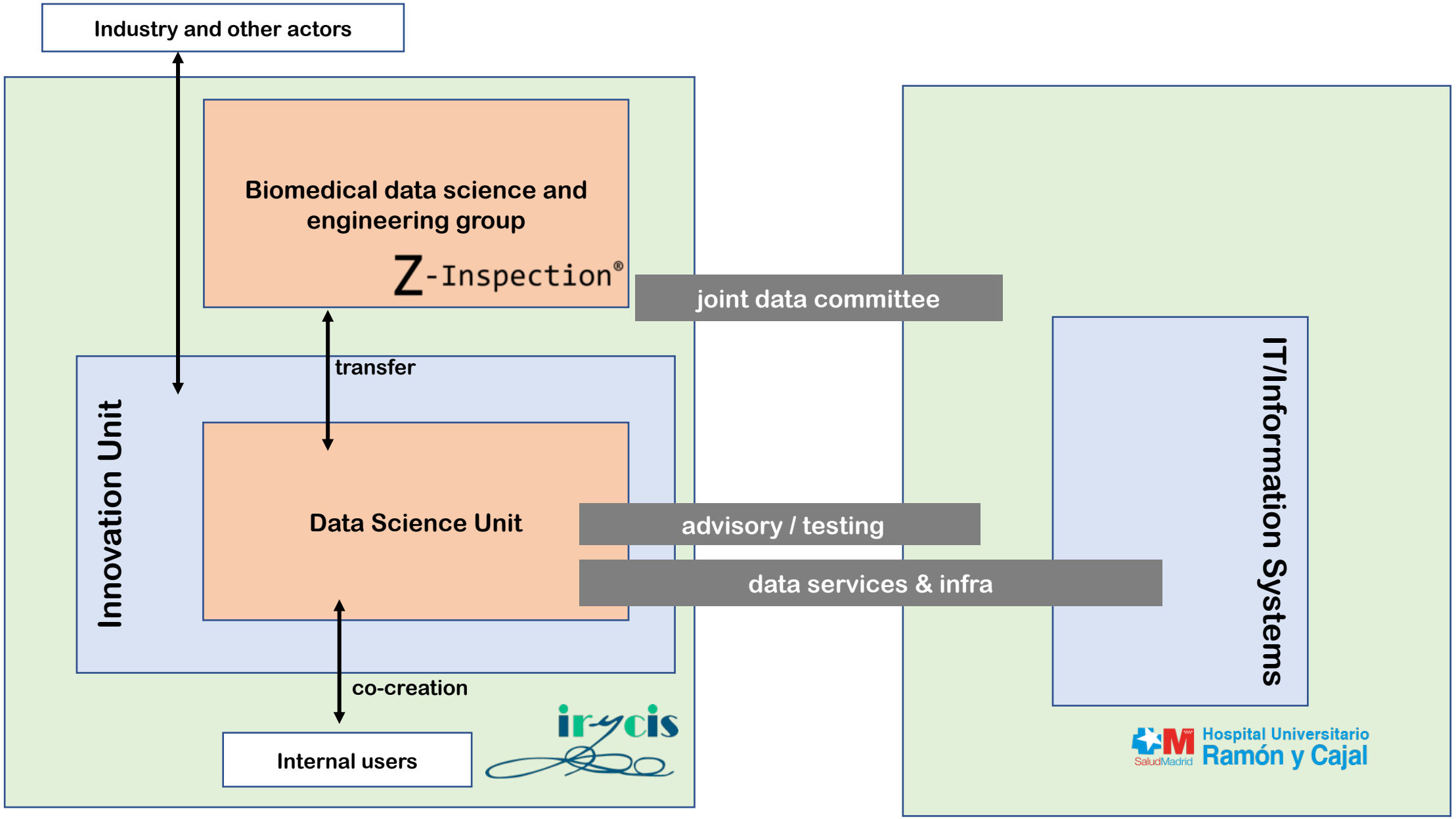
Trustworthy AI and Z-inpection@IRyCIS

Miguel-Angel Sicilia
msicilia@uah.es

Data Science Unit Leader
Instituto Ramón y Cajal de Investigación Sanitaria (IRyCIS)

About me

- Prof. in Computer Science, University of Alcalá
- Since 2022, coordinator of the Data Science Unit of the Research Institute at the “Ramón y Cajal” Hospital in Madrid.
- Co-founder In2AI
- Established the Z-Inspection Lab@IRyCIS early 2024.


Industry and other actors
Biomedical data science and engineering group
Z-Inspection®
joint data committee
transfer
Innovation Unit
Data Science Unit
advisory / testing
data services & infra
IT/Information Systems
co-creation
Internal users
irycis
Hospital Universitario Ramón y Cajal
SaludMadrid

Initial focus @ IRyCIS of using Z-inspection

- Given that AI is considered a crosscutting priority in the organization, the focus is on planning in advance.
- Trustworthy-first design of clinical decision making interfaces.
  - Preparing for transition to actual practice.
  - Avoiding late findings of pitfalls or barriers.
- Current emphasis on key enablers of AI based projects:
  - Adapting socio-technical scenario modeling.
  - Collecting best practices on traceability, especially and starting from research data.

Large case study: AI and pancreatic cancer

- Focus: risk, early detection and diagnostic tools for PC
- Z inspection considered as the basis for the case across institutes in Spain first, then in EATRIS overall.
- Digital capabilities with emphasis on predictive AI.
  - Includes Trustworthy AI as key element.
  - Materials with best practice in a Digital HUB

Demonstrate that the **DH is useful.**	Best practice **sharing** inside EATRIS.	**Supporting** EATRIS healthcare orgs in adoption.

① IRYCIS

⑥ IIS Aragón

② I+12

⑦ IDIVAL

③ IIS LA FE

⑧ ibsGranada

④ IDIS

⑨ VHIO (WP1)

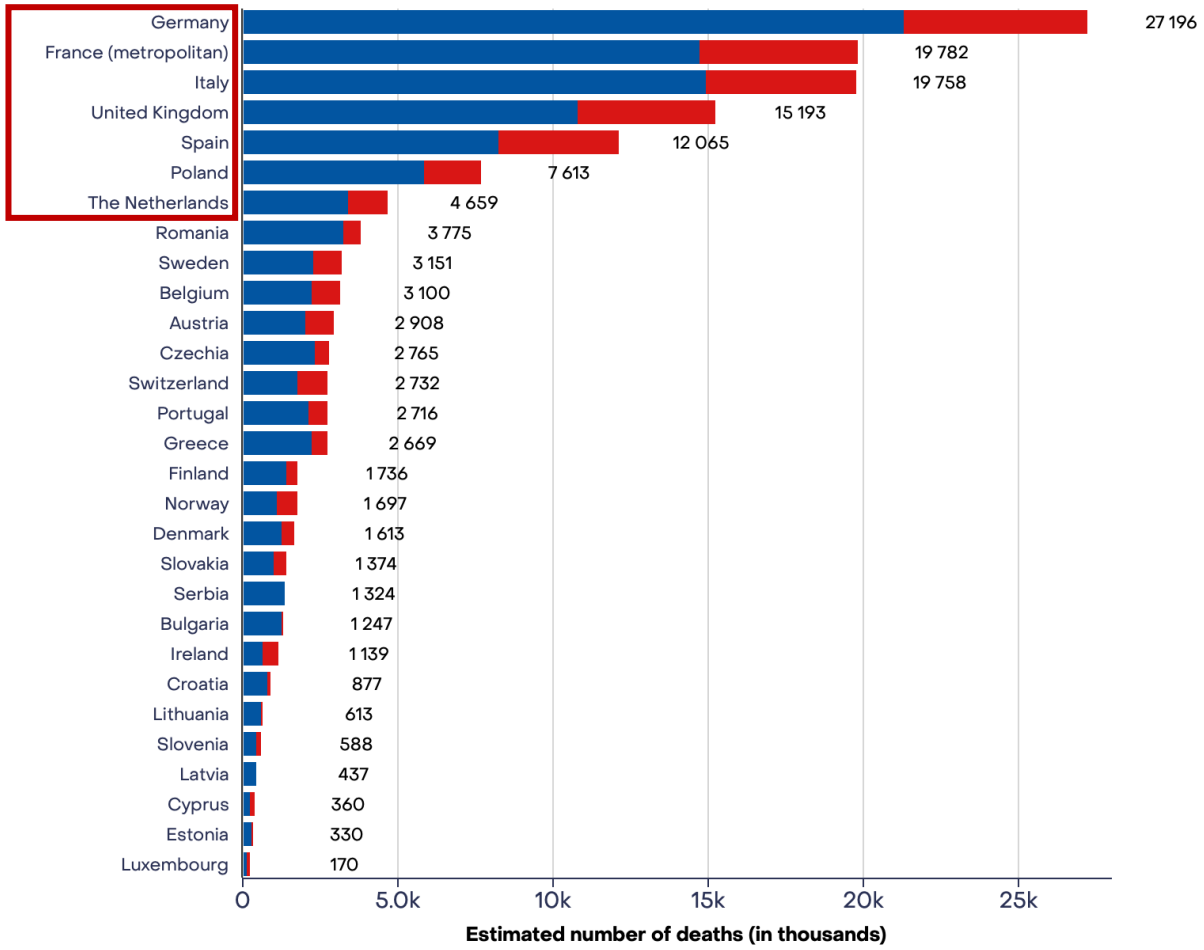
⑤ IIS Biogipuzkoa

UNIVERSITY OF HELSINKI

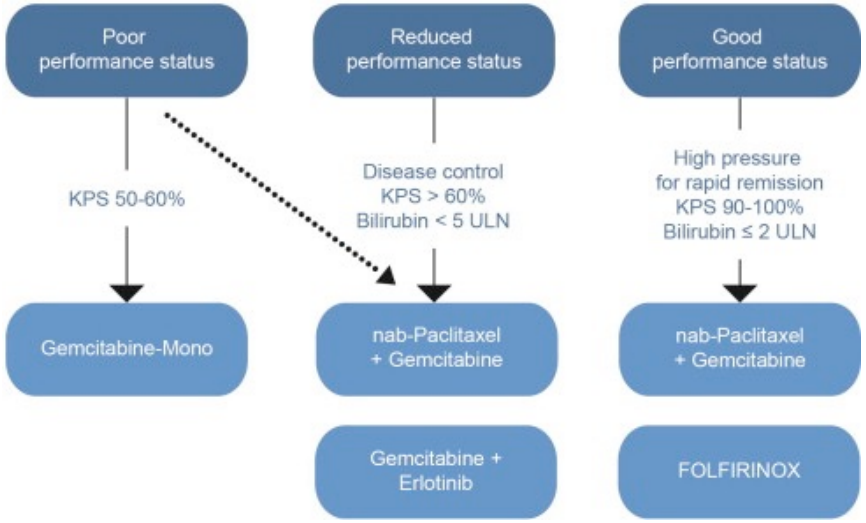
Palacký University Olomouc

Radboudumc university medical center

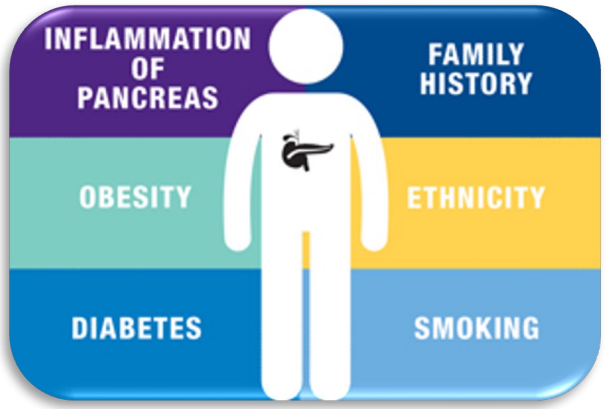
DEATHS ESTIMATED IN EUROPE 2022- 2045 IN EUROPE



Current treatments

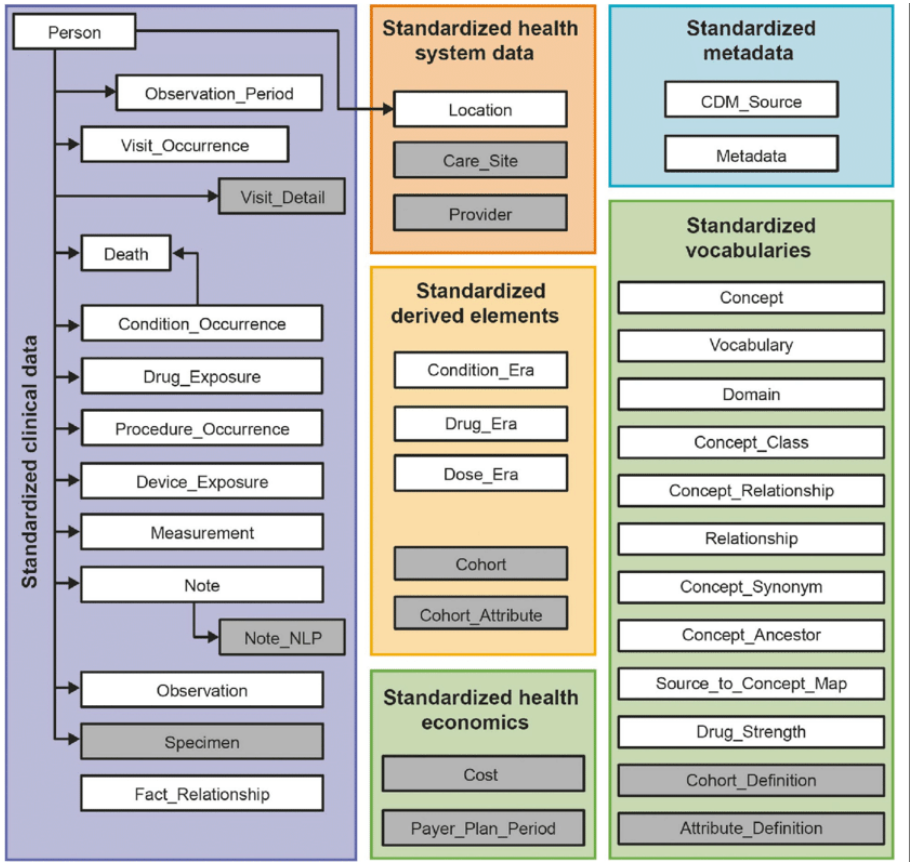


Risk Factors

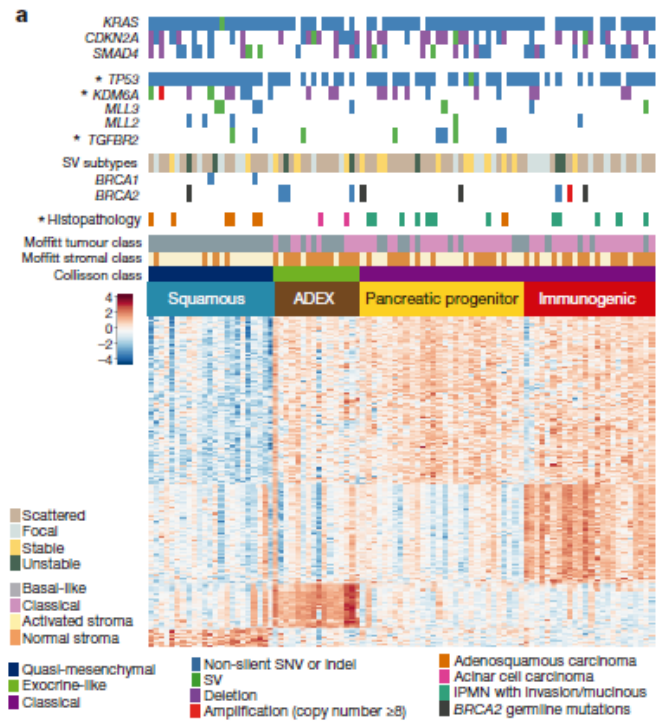


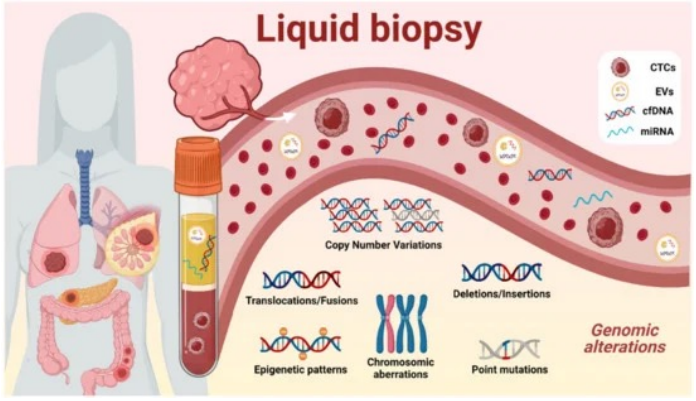
AVAILABLE INFORMATION AT HOSPITALS

Electronic Health Records (EHR) (standards as *enablers* for accountability)

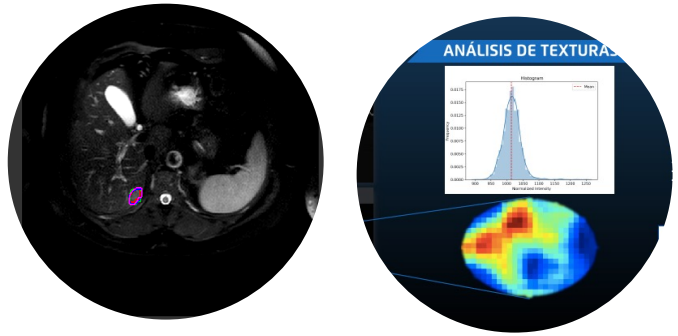


Molecular Markers

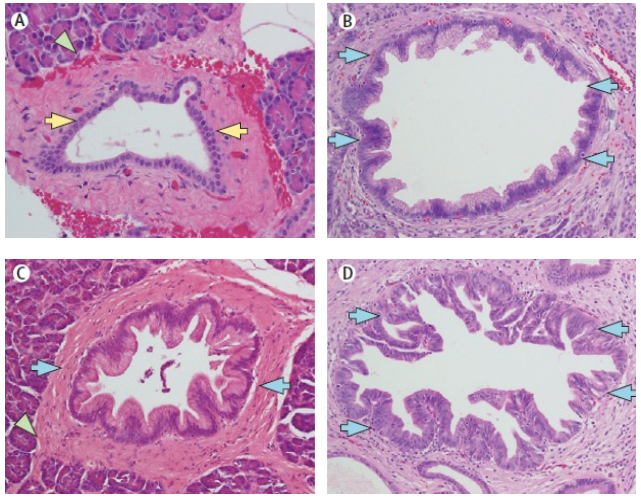




Imaging: CT scan, MRI



Histopathology



IRYCIS/ES TASK FORCE – multidisciplinary team

- M Laura García Bermejo, **WP6 coordinator, Biomarkers & Therapeutic Targets Leader**
- Miguel Ángel Sicilia**, Data Science Unit Director**
- Elena García Barriocanal, **Data Science Unit member**
- Ana García-García de Paredes**, Gastroenterologist (pancreatic diseases section leader)**
- Mercedes Rodriguez Garrote, **Medical Oncologist (GI tumor section)**
- Carolina de la Pinta**, Radiation Oncologist (GI tumor section) /Radiomics**
- Elisa Conde, **Biomarkers &Therapeutic Targets Unit**
- Julie Earl, **Pancreas Cancer Patient registry responsible at IRYCIS, BIOPAC Group**
- Val Fernández, **Bioinformatics (Trans-BIONET, ELIXIR)**
- Enrique Aracil, **Hospital Deputy Director, Responsible of EHR at Ramon y Cajal**
- Sonia Camaño, **HRYC Biobank coordinator, (BBMRI), Responsible IRYCIS Quality**
- Itziar de Pablo, **Ethical Committee of Clinical Investigation (SCREEN/ ECRIN), AI expert**
- Richard Buck, **Patient representative**
- Fuensanta Bellvis, **Quibim Development Support**

Clinical variables Definition and sources: initial Proposal

	Structured?	Where to find it?
Patient characteristics		
sex	Yes	HCIS
age at diagnosis	No	HCIS- Gastroenterology
race	No	HCIS
smoker	No	HCIS
alcohol	No	HCIS
cannabis	No	HCIS
exposure to chemical compounds	No	HCIS
Obesity: BMI <30	No	HCIS
blood group	No	HCIS
sedentary lifestyle	No	HCIS
history of helicobacter pylori infection	No	HCIS- Gastroenterology
history of chronic HBV	No	HCIS- Gastroenterology
history of chronic HCV infection	No	HCIS- Gastroenterology
history of other tumor	No	HCIS
DM 1 or 2	No	HCIS
date of DM diagnosis	No	HCIS
history of acute pancreatitis	No	HCIS
history of chronic pancreatitis	No	HCIS
type of pancreatitis	No	HCIS- Gastroenterology
familiar history of cancer	No	HCIS-Oncology
familiar pancreatic cancer	No	HCIS-Oncology
hereditary syndromes: familial breast and ovarian cancer (BRCA1, BRCA2, PALB2), Peutz-Jeghers syndrome (STK11), Lynch syndrome (MLH1, MSH2, MSH6), familial atypical melanoma syndrome with multiple mole (CDKN2A), ataxia-telangiectasia syndrome (ATM)	No	HCIS-Oncology
Precursor lesion: cystic mucinous neoplasia, intraductal papillary mucinous neoplasia and pancreatic intraepithelial neoplasia (PanIN).	No	HCIS
high-risk stigmata	No	HCIS- Gastroenterology
worrisome features	No	HCIS- Gastroenterology
Date of diagnosis of the precursor lesion	No	HCIS
Suspicious findings according to Fukuoka criteria.	No	HCIS
Surgical history of cholecystectomy or gastrectomy, treatment with statins, vitamin D, aspirin, NSAIDs, other diseases or complications	No	HCIS
fasting blood glucose in year prior to diagnosis	Yes	Laboratory- HCIS
glycosylated hemoglobin in year prior to diagnosis	Yes	Laboratory- HCIS
weight loss in year prior to diagnosis	No	HCIS
abdominal pain in year prior to diagnosis	No	HCIS
meteorism in year prior to diagnosis	No	HCIS
diarrhea in year prior to diagnosis	No	HCIS
nutritional status at diagnosis	No	HCIS
exocrine pancreatic insufficiency at diagnosis	No	HCIS
Tumor characteristics		
ECOG	No	HCIS
diagnosis: incidental, symptomatic	No	HCIS
date of diagnosis	No	HCIS
stage	No	HCIS
histology	No	GESPAD reports
grade	No	GESPAD reports
location	No	HCIS
size	No	HCIS
Presence of metastases at diagnosis, type of metastases: hepatic, pulmonary, bone	No	Radiology reports- HCIS
PET/CT at diagnosis	No	Nuclear medicine reports- HCIS
liver MRI at diagnosis	No	Radiology reports - HCIS
exploratory laparotomy at diagnosis	No	HCIS- General surgery
resectability assessed by a multidisciplinary committee	No	HCIS - Multidisciplinary committee reports
type of tumor (resectable, bordeline, unresectable)	No	HCIS - Multidisciplinary committee reports
Ca 19.9 at diagnosis	Yes	Laboratory- HCIS
fasting blood glucose at diagnosis	Yes	Laboratory- HCIS
glycosylated hemoglobin at diagnosis	Yes	Laboratory- HCIS
CEA at diagnosis	Yes	Laboratory- HCIS
hemoglobin at diagnosis	Yes	Laboratory- HCIS
PCR at diagnosis	Yes	Laboratory- HCIS
Total Bilirubin at diagnosis	Yes	Laboratory- HCIS
alkaline phosphatase at diagnosis	Yes	Laboratory- HCIS
GGT at diagnosis	Yes	Laboratory- HCIS
GOT at diagnosis	Yes	Laboratory- HCIS
GPT at diagnosis	Yes	Laboratory- HCIS
Treatment characteristics		
endoscopic biliary drainage		
type of surgery	Yes	HCIS- General surgery
pTNM	Yes	Pathology reports
date of surgery	Yes	Pathology reports – General surgery
neoadjuvant chemotherapy and type	No	HCIS -Medical oncology
date of start of chemotherapy	Yes	HCIS -Medical oncology- Pharmacy
number of cycles	Yes	HCIS -Medical oncology- Pharmacy
response to treatment according to RECIST	No	Radiology reports
response to treatment according to Ca 19.9		
ypTNM in surgical specimen	Yes	Pathology reports
adjuvant chemotherapy	No	HCIS -Medical oncology
radiation therapy (neoadjuvant, radical, adjuvant)	No	HCIS -Radiation Oncology
type of radiotherapy (SBRT or conventional radiotherapy)	No	HCIS -Radiation Oncology
doses of radiotherapy	No	HCIS -Radiation Oncology
unresectable or metastatic patient	No	Radiology reports
local progression	No	Radiology reports
distant progression	No	Radiology reports
first line of chemotherapy	Yes	HCIS -Medical oncology- Pharmacy
second line of chemotherapy	Yes	HCIS -Medical oncology
status	No	HCIS
Exploratory variables related to analytical parameters, other medical history and treatments prior to diagnosis will be included		
Radiomics: GLSZM, NGLDM, energy, entropy,...	NA	DICOM

Traceability – from DMPs to Trustworthy AI

- Translational medicine:
  - Often described as an effort to carry scientific knowledge “from bench to bedside,” builds on basic research advances and uses them to develop new therapies or medical procedures.
- Research planned in advance and data processes and products documented first in Data Management Plans (DMPs)
- This sets the need to advance the data infrastructure and data engineering practices to trace from DMPs to all downstream data artifacts.

Processes and trustworthy AI

planning/preparation
data collection
data extraction
data transformation

>

analytics/development
modelling
pipelining
tooling

>

productioning/deployment
testing
deploying
assessing/cert

accountability(7)
DMP integration ①
data lineage/pipelines ①②③

robustness/safety(2)
adversarial testing
data lineage/pipelines ①②③
causal experiments
standards/interop/FAIR ①

human agency (1)
HITL/HOTL/HIC
explainability tools ①②

privacy/data gov(3)
privacy risk assessment ①
PPML /PPA ①②
anonymization/diff. privacy ①②
metadata curation ①②
DMP integration ①

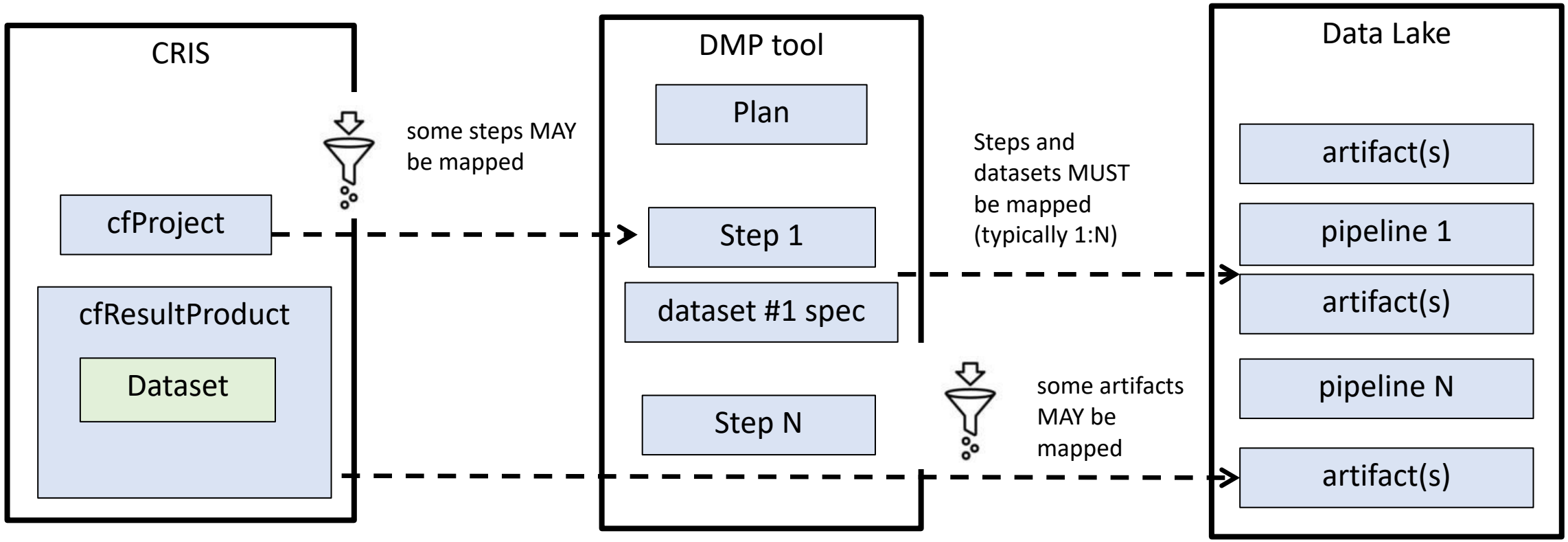
fairness(5)
fairness tools ①

social/env impact(6)
footprint tracking
data lineage/pipelines ①②③

transparency(4)
data lineage/pipelines ①②③
metadata curation ①②
explainability tools ①②

CRISes, DMPs and Data Lakes

- The Data Management Plan (DMP) is the specification that mediates between the Project and the artifacts being generated in the Data Lake.



Bridging the DMP, the metadata catalogue and actual data