

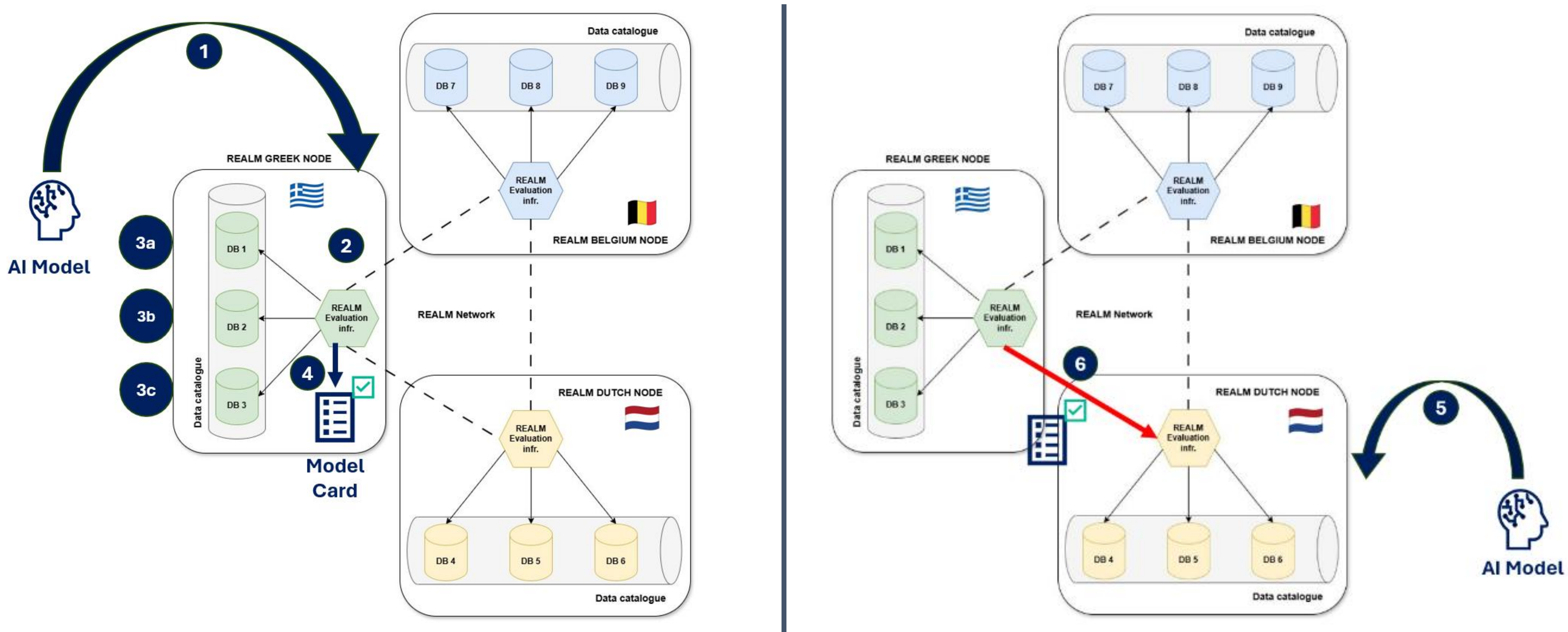
From Federated National Data Nodes to trustworthy medical AI evaluation

Title: The REALM Data Catalogue and Federation Model for Cohort-Driven Evaluation of Medical Device Software



Background: The evaluation of Medical Device Software (MDSW) and in particular those using AI increasingly require access to diverse, Real-World Data (RWD) across multiple data providers. The OHDSI community has established mature federation networks such as DARWIN EU, EHDEN, and FHIN based on distributed analytics and local execution with proven track record of efficiency and robustness for observational research and real-world evidence generation. However, the evaluation of AI-based MDSW introduces additional requirements, including repeated execution of computationally intensive models, scenario-based testing, and detailed performance characterization, which motivate complementary evaluation infrastructures tailored to these workflows and regulatory needs. The REALM project addresses this gap by designing an evaluation-oriented data infrastructure tailored to evaluate AI from MDSW that enables controlled access to harmonized clinical data for the purpose of verifying manufacturer claims, benchmarking model performance, and supporting regulatory decision-making in Europe.

Figure 1: REALM national node & evaluation workflow



 Belgium Node	 Greek Node	 Dutch Node
 Gokhan Ertaylan	 Vangelis Sakkalis	 Michel Dumontier
 Frederic Jung	 Emmanouil (G) Spanakis	 Gokhan Ertaylan
	 	

- 1 Model developers import AI medical software in a REALM node
- 2 Model developers submit their claims
- 3 A cohort definition is created and submitted against the data sources available (a, b c)
- 4 Model and evaluation results are described in a Model card
- 5 Model is submitted to another node, steps 1,2,3,4 are repeated
- 6 New results are written in the model card

Data protection and governance are central to the REALM federation model. All patient-level data remain secure within the national nodes. Patient-level data are never transferred outside the evaluation environment, and AI models submitted for evaluation are always pre-trained and executed locally without any retraining, ensuring that data does not leave the node at any stage. Only high-level evaluation results such as performance metrics, provenance information and high-level data distributions are shared externally through structured model cards. This approach aligns conceptually with OHDSI's federated principles while enabling workflows that necessitate continuous access to multiple data sources.



EU Project n°101095435



Frédéric Jung¹, Stef Rommes¹, Gökhan Ertaylan¹

1. VITO, Vlaamse Instelling voor Technologisch Onderzoek, Mol, Belgium

