Artificial intelligence for personalised medicine in depression - analysis and harmonization of clinical research data for robust multimodal patient profiling for the prediction of therapy outcome

ICPerMed RECOGNITION 2022
Data Sharing in Personalised Medicine Clinical Research

Julia Stingl

Institute of Clinical Pharmacology, University Hospital Aachen

Maria Giulia Baccalini

IRCCS, Istituto delle Scienze Neurologiche di Bologna

ERA PerMed

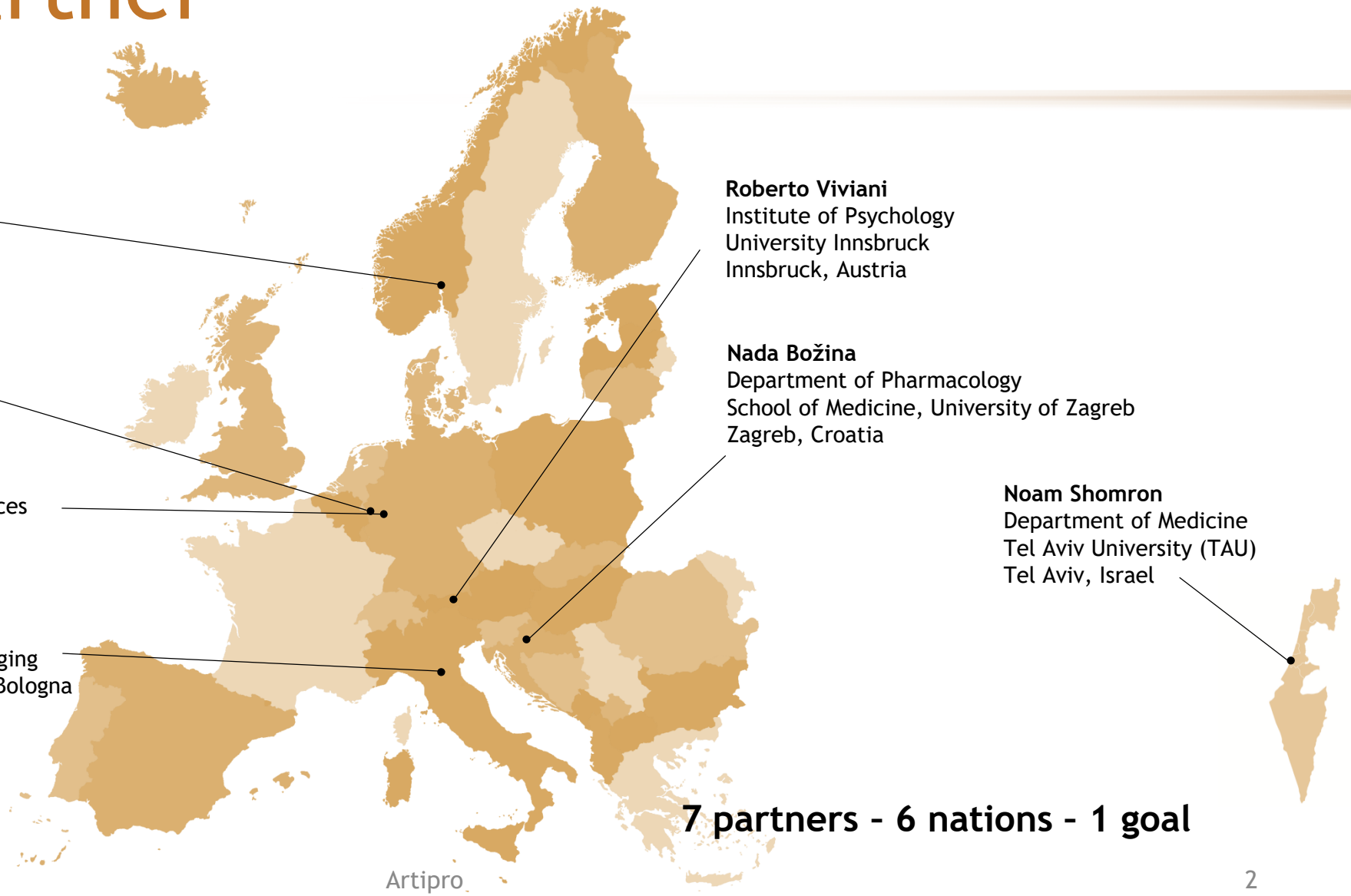
Project Partner

Espen Molden
Center for Psychopharmacology
Diakonhjemmet Hospital,
Oslo, Norway

Julia Stingl
Clinical Pharmacology
University Hospital of RWTH Aachen
Aachen, Germany

Catharina Scholl
Research Department
Federal Institute for drugs and medical devices
BfArM
Bonn, Germany

Maria Giulia Bacalini
Direzione Scientifica – Laboratorio di Brain Aging
IRCCS Istituto delle Scienze Neurologiche di Bologna
Bologna, Italy



Roberto Viviani
Institute of Psychology
University Innsbruck
Innsbruck, Austria

Nada Božina
Department of Pharmacology
School of Medicine, University of Zagreb
Zagreb, Croatia

Noam Shomron
Department of Medicine
Tel Aviv University (TAU)
Tel Aviv, Israel

7 partners - 6 nations - 1 goal

Goal – Redrawing the boundaries of mental illness

- identify biomarker signatures that can serve to stratify patients for symptom-related outcome and therapeutic response
  - a multifactorial biomarker profile that can be used in clinical decision support systems to personalise patient health care, including prognosis definition and treatment

  - Clinical implications
    - →health care personalisation
    - → changes in regulatory practice

Multimodal approach for identification of biomarkers

- Establishment of an artificial intelligence platform
  - Combining material from clinical studies

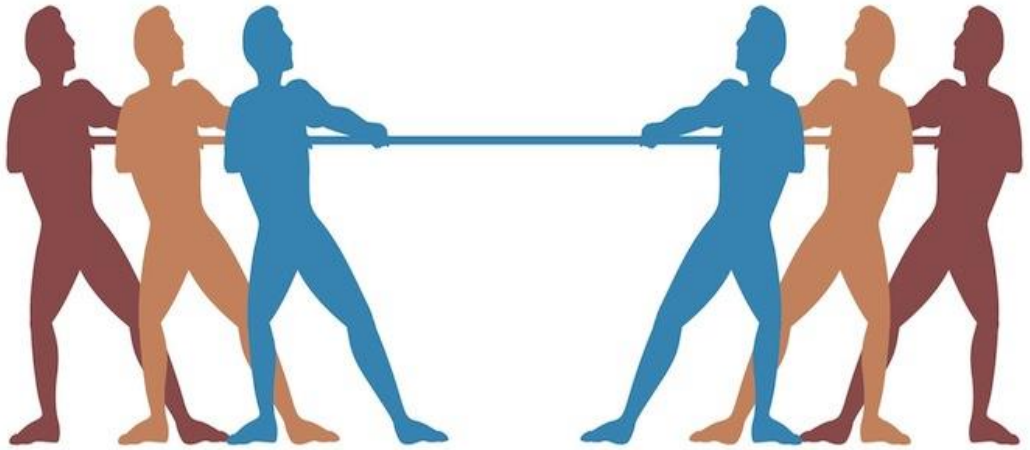
→Combining of multiple modalities

→Fill biomarker gaps across studies and complement biomarkers with the newest molecular analytical technologies

ArtiPro ethical and legal challenges

ArtiPro will **re-use data**
that have been previously collected or generated in the framework of distinct and unrelated projects and that will be shared and transferred across the Consortium.

WHO
EU
FAIR



Ethics committees
GDPR
National laws

ArtiPro ethical and legal challenges

The example of Italy

The analysis of the Decree reveals a legislative technique aimed at developing most of the *'open clauses'* contained in the GDPR, *i.e.* those allowing member states to specify further certain aspects of the provisions concerned (for instance, those related to children's consent and the processing of special categories of personal data).

Herein you will find a summary of the main provisions of the Decree.

- **A. Children's consent in relation to information society services** Children who have reached the age of 14 years can validly express their consent to data processing in relation to the offer of information society services. The holder of parental responsibility over the child must give consent where the child is below the age of 14 years. In this regard, the GDPR sets the minimum age of 16 years old, but allows member states to fix a lower age for granting valid consent. This is an example of how the Italian legislator benefitted from a GDPR open clause. Considering such age limit, online content requests will have to be devised so that they are appropriate for children of at least 14 years of age. This, without prejudice to the specific law provisions concerning the required age for the execution of contracts.
- **B. Safeguard measures for the processing of biometric, genetic and health-related data** Biometric, genetic and health-related data can be processed if specific safeguard measures (including security measures, such as encryption and pseudonymization) are implemented. The Italian DPA (*Garante*) will establish such safeguards at least on a two-yearly basis (see also lett. G below). This is a further specification of Article 9 (4) of the GDPR, concerning the processing of special categories of personal data, allowing member states to introduce additional conditions and limitations with regard to the special categories of personal data. This means that, under Italian law, the processing of such data may be subject to particularly strict requirements. Such a provision enhances the protection of biometric, genetic and health-related data, but it creates further burdens for data controllers and processors.
- **C. Judicial data** The processing of judicial data is allowed based only on a law or regulatory provision providing for appropriate safeguards for data subjects. Lacking these law provisions, the requirements for the lawful processing of judicial data shall be determined through a Decree of the Ministry of Justice. That said, there are specific cases—newly introduced in the Italian Data Protection Code—in which law provisions should broaden, in the near future, the possibility to process judicial data lawfully (e.g., in the

The legislative decree no. 101 of August 10, 2018 (Decree), amends and adapts the Italian Data Protection Code (Legislative decree no. 196/2003, Data Protection Code or DPC) to the GDPR.

The Data Protection Officer of Partner 1 did not recognize existing consents as legal basis for ArtiPro data sharing and re-use

We prepared a D.P.I.A. (Data Protection Impact Assessment) to submit a prior consultation request to the Italian Privacy Authority

ArtiPro ethical and legal challenges: methodology

A step-by-step process has been undertaken to guide ArtiPro researchers to efficiently and consciously share existing data in the framework of the project

- Data Management Plan
- Partners with different expertise within the Consortium
- Questionnaire templates shared within the Consortium to check the characteristics and specific requirements of each dataset

Licence issues (Refer to the Consortium Agreement of the existing dataset)
Question
Who is the owner of this dataset?
Anonymised demographic/clinical data
Can the dataset be re-used by the owner in the framework of ArtiPro project?
Can the dataset be re-used by the other researchers belonging to ArtiPro Consortium?
Is there any restriction for the re-use of this dataset in ArtiPro?
Is the existing dataset "open access" at present?
If yes, how can it be accessed?
If not, how can ArtiPro researchers access the dataset?
What is the estimated volume of the data?
If not already open access, can the existing dataset become completely open in the future?
If yes, when?
If not, why?
Existing biomarkers (genetics, metabolomics, etc); if necessary, please copy this section
Can the dataset be re-used by the owner in the framework of ArtiPro project?
Can the dataset be re-used by the other researchers belonging to ArtiPro Consortium?
Is there any restriction for the re-use of this dataset in ArtiPro?
Is the existing dataset "open access" at present?
If yes, how can it be accessed?
If not, how can ArtiPro researchers access the dataset?
What is the estimated volume of the data?
If not already open access, can the existing dataset become completely open in the future?
If yes, when?
If not, why?

Committee)
Question
Does the existing consent cover re-use of available data?
Does the existing consent cover re-use of available samples?
Does the existing consent cover transfer of available samples between ArtiPro partners?

DATA HARMONIZATION

Datasets tend to be organized differently and to express variables with diverse formats.

e.g., different indicator for sex over three providers.

A	B	C
0	M	Male
1	F	Female

e.g., different scales for QoL over two providers.

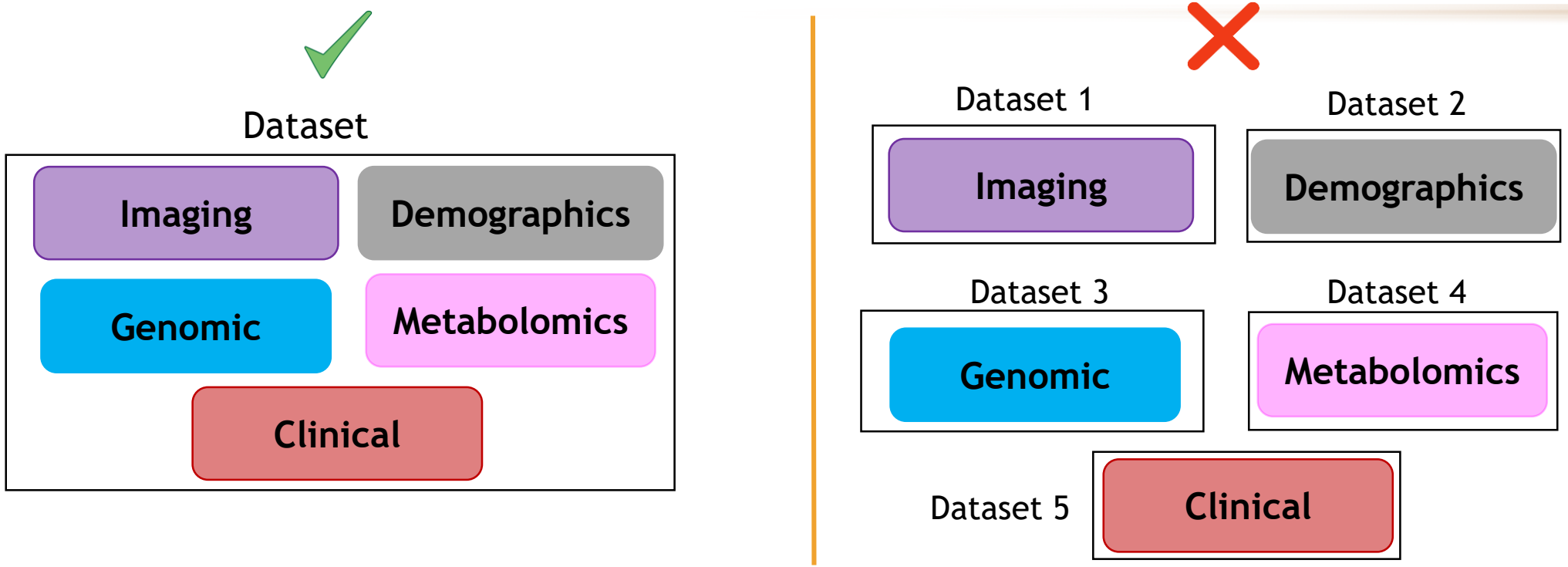
	Value	Scale
Sample from A	100	HDRS
Sample from B	4	QIDS
Sample from B	2	QIDS
Sample from A	30	HDRS

AI-based

Conversion table needed.

QIDS	HDRS
0	70
1	65
	60
2	55
	50
3	45
	40
4	30
	20
	10
5	0

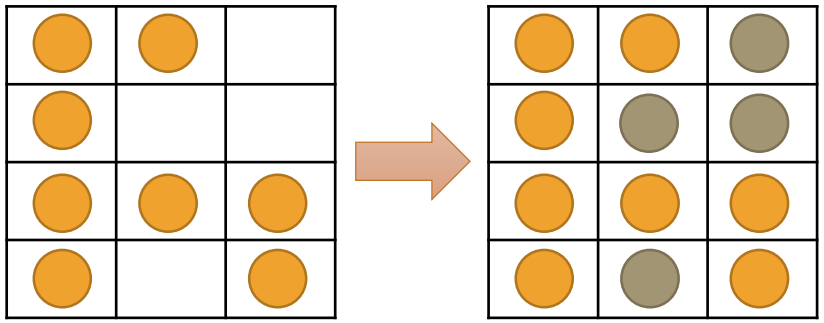
MULTI-MODAL INTEGRATION



It is better to have a few multi-modal datasets (even one) than many single-modal ones.

Filling gaps — Imputation

Instead of removing samples and variables with unavailable data, the idea is to replace such missingness with likely values.

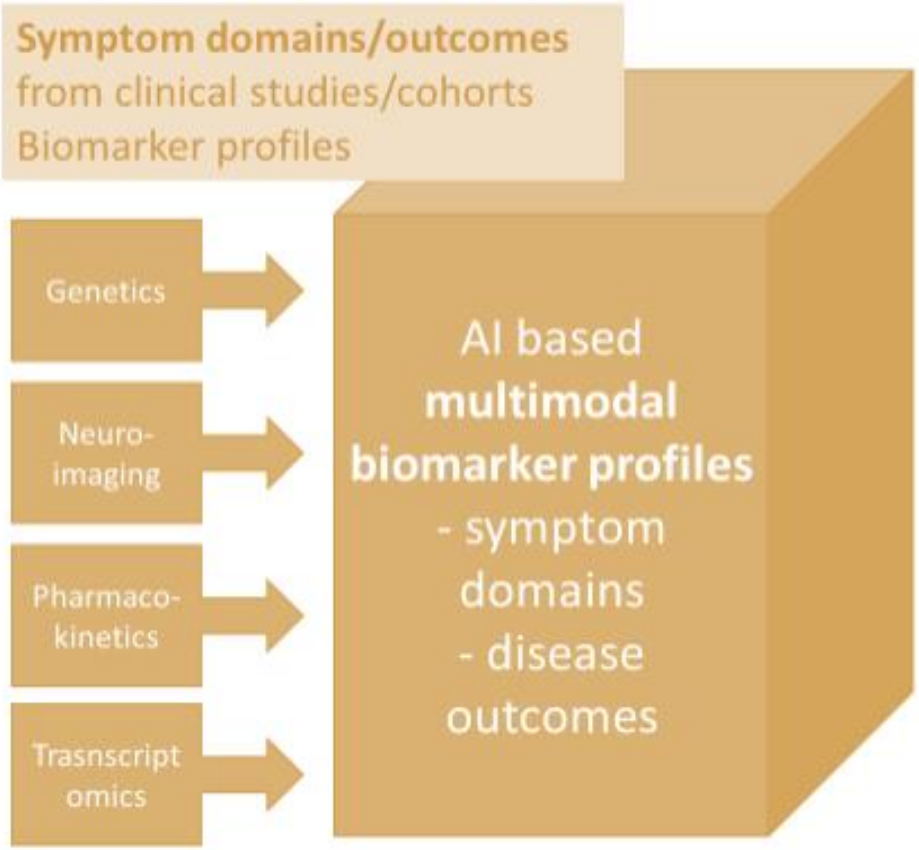


AI-based

How the missing data distribute strongly influences imputation techniques.

A general rule of thumb is: the more ***at random*** they are, the better it is.

Benefit of ArtiPro



Benefit of ArtiPro