RD IT & Data Systems: Meeting the needs of ERNs
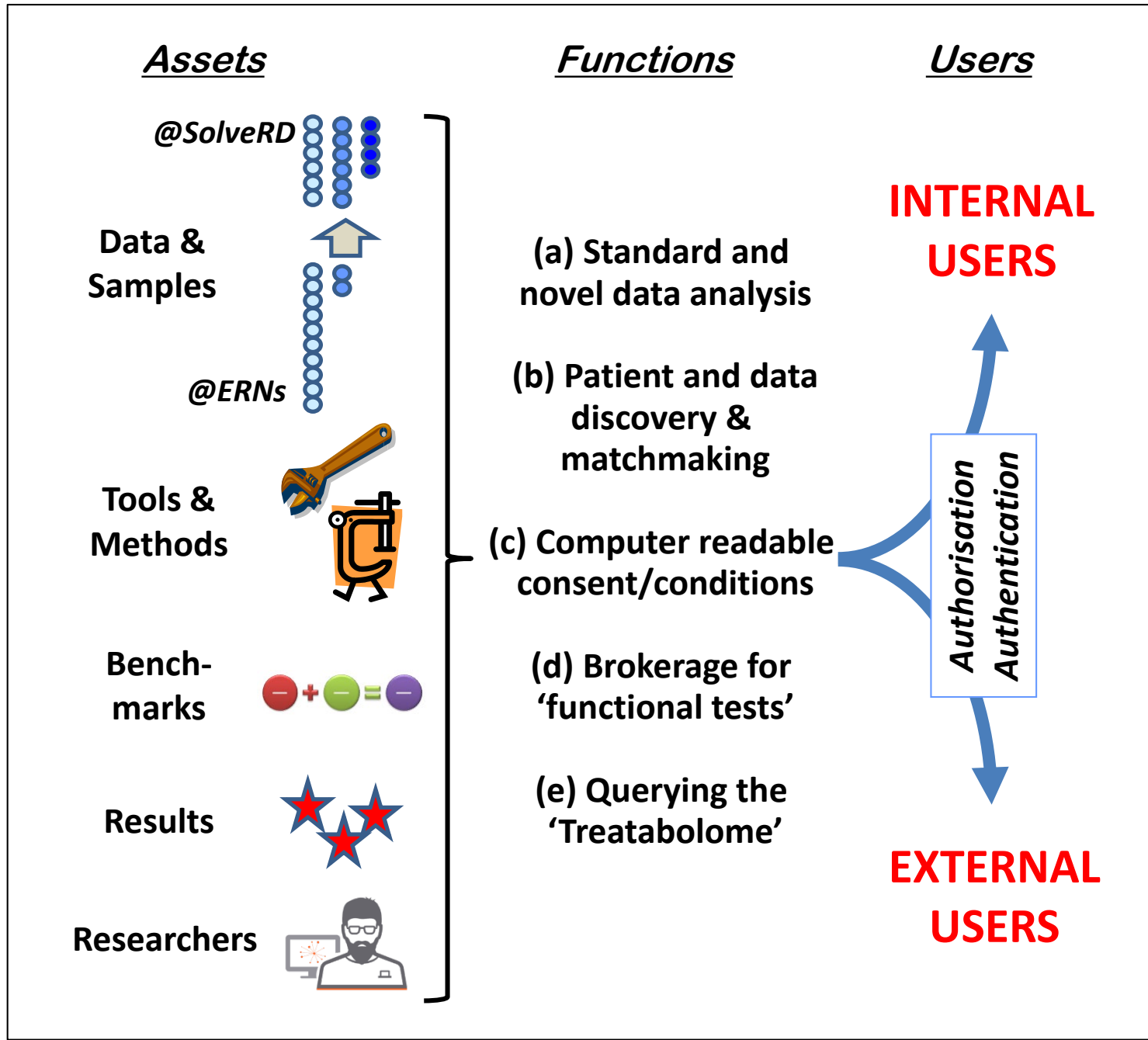
Anthony Brookes
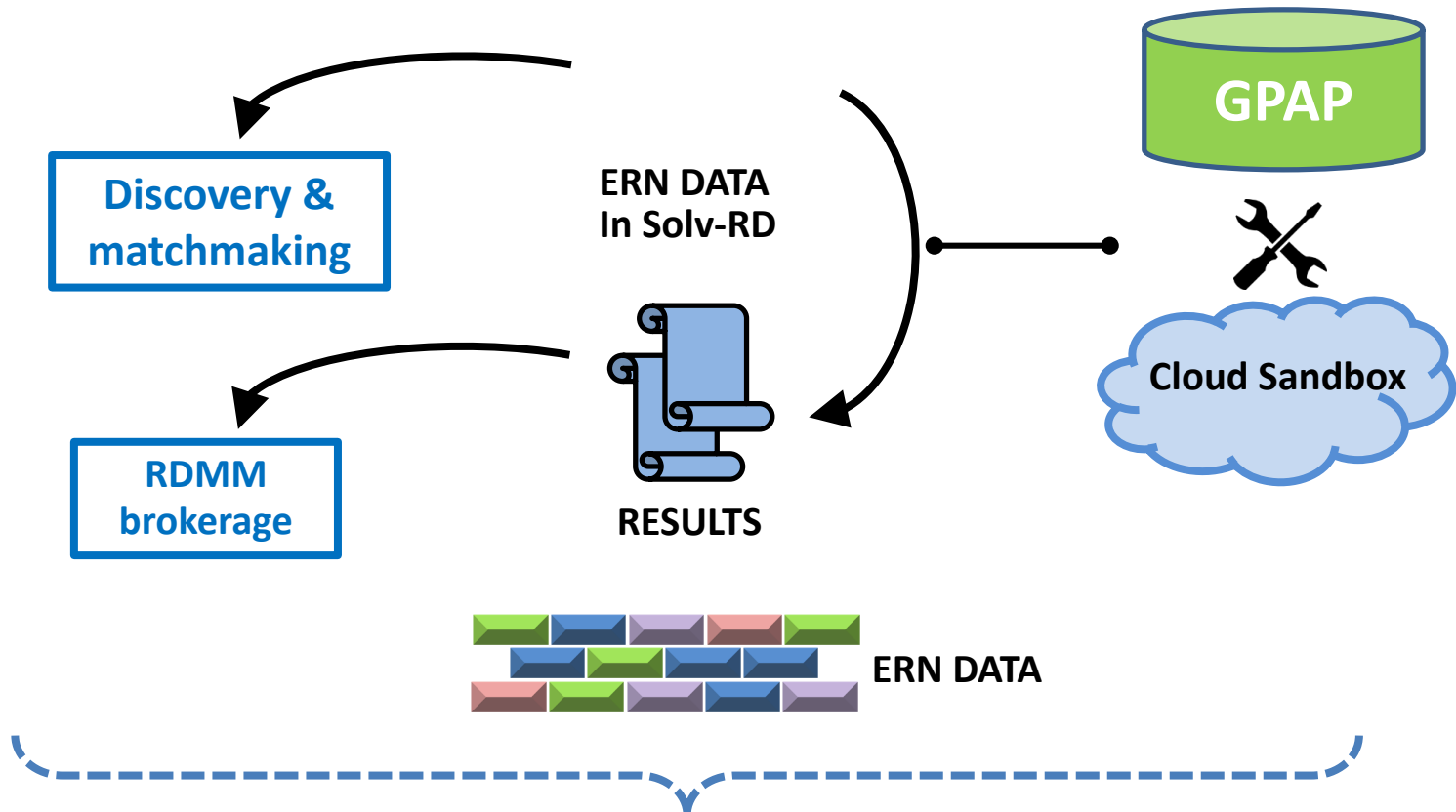
University of Leicester, UK

UNIVERSITY OF
LEICESTER

Conflict of interest statement:

I have no actual or potential conflict of interest in relation to this program/presentation
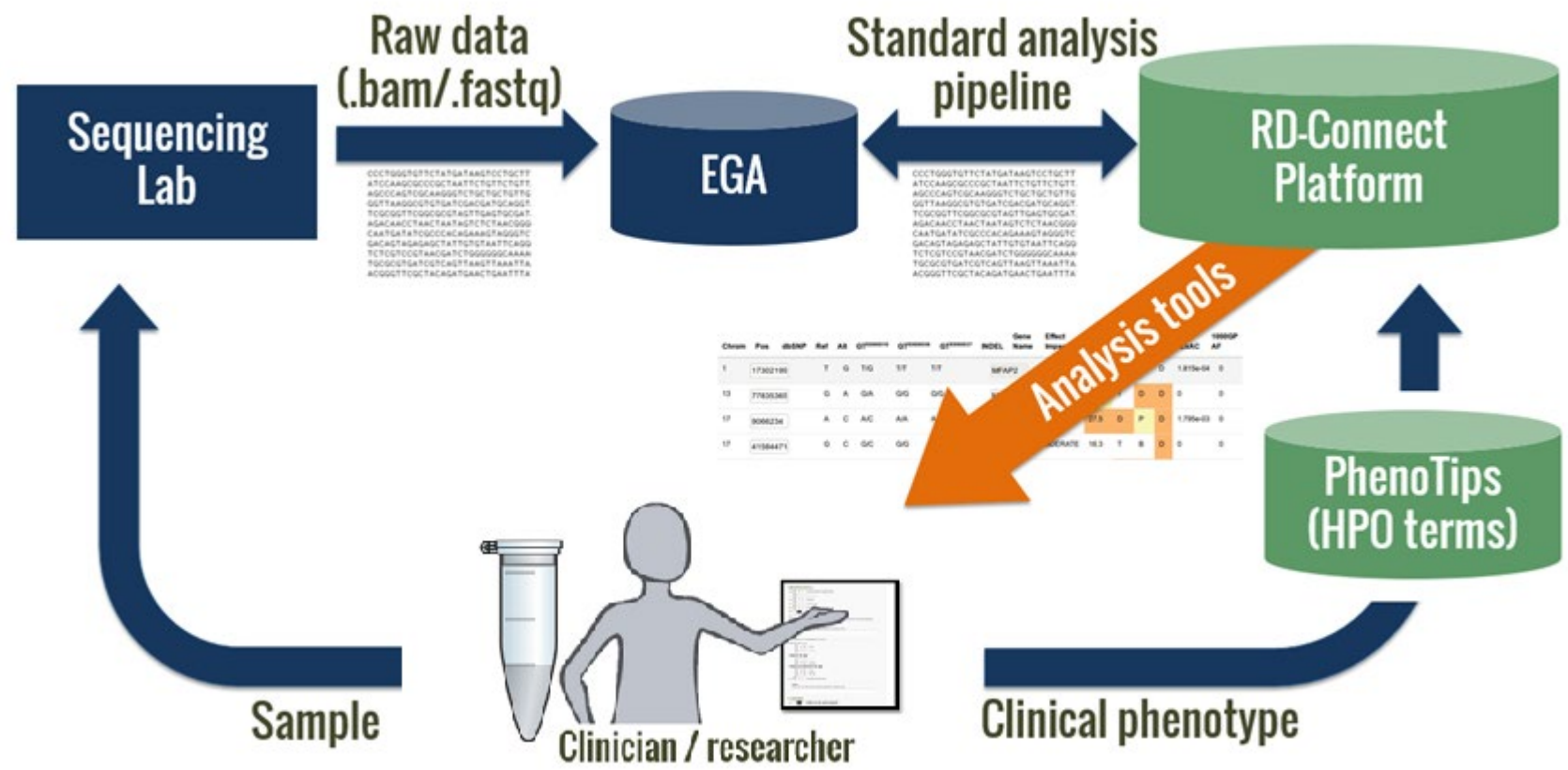
Assets

@SolveRD

Data & Samples

@ERNs

Tools & Methods

Bench-marks

Results

Researchers

Functions

(a) Standard and novel data analysis

(b) Patient and data discovery & matchmaking

(c) Computer readable consent/conditions

(d) Brokerage for 'functional tests'

(e) Querying the 'Treatabolome'

Users

INTERNAL USERS

Authorisation Authentication

EXTERNAL USERS

GPAP

Cloud Sandbox

ERN DATA
In Solv-RD

RESULTS

Discovery &
matchmaking

RDMM
brokerage

ERN DATA

- **Where are the data at any time ?**
- **Data ownership, provenance [metadata] ?**
- **Permissions for discovery/sharing /use ?**
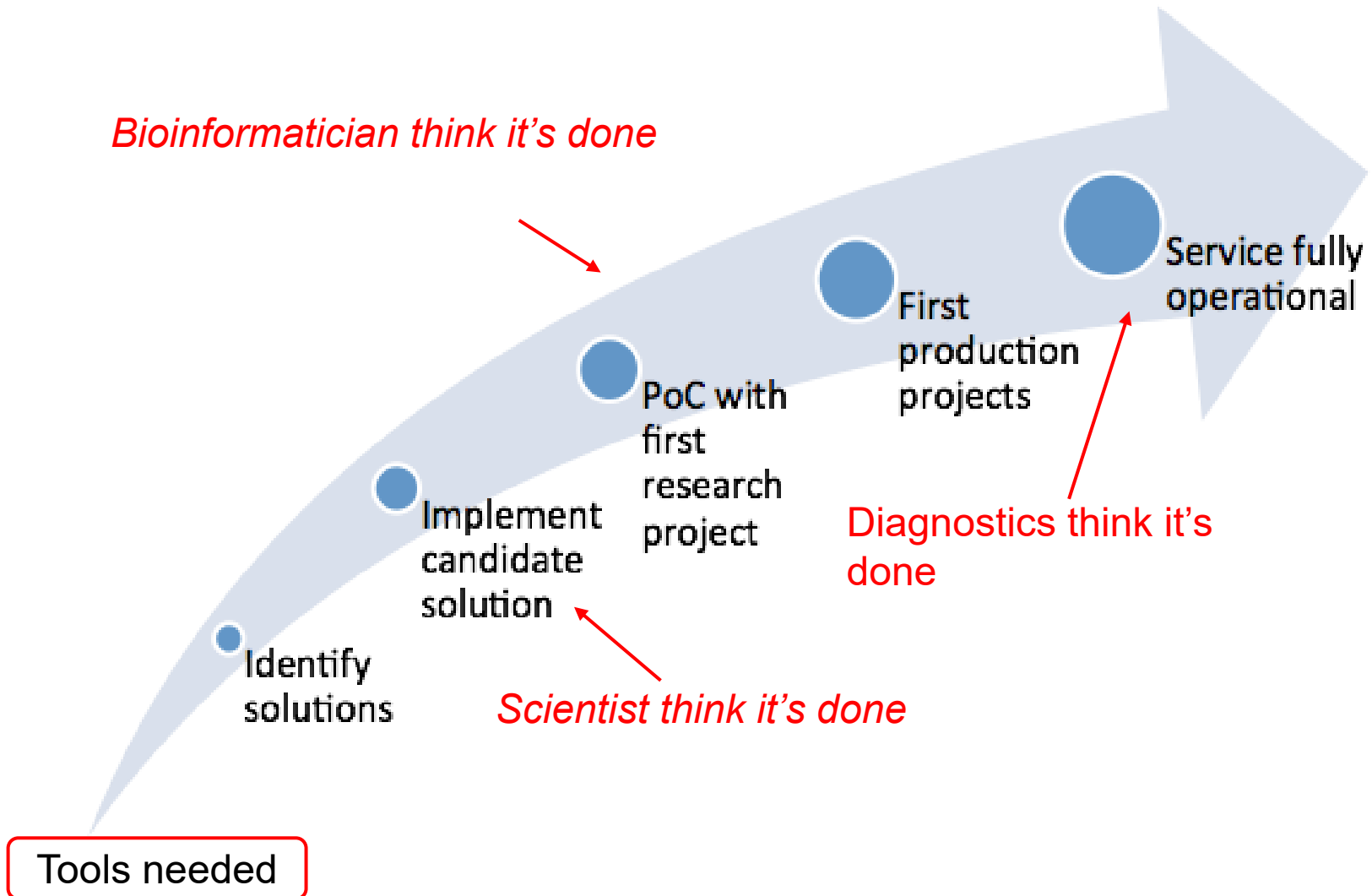- **Which data are 'findable' by whom [Solve-RD, ERN, Others] ?**

**Under
ERN
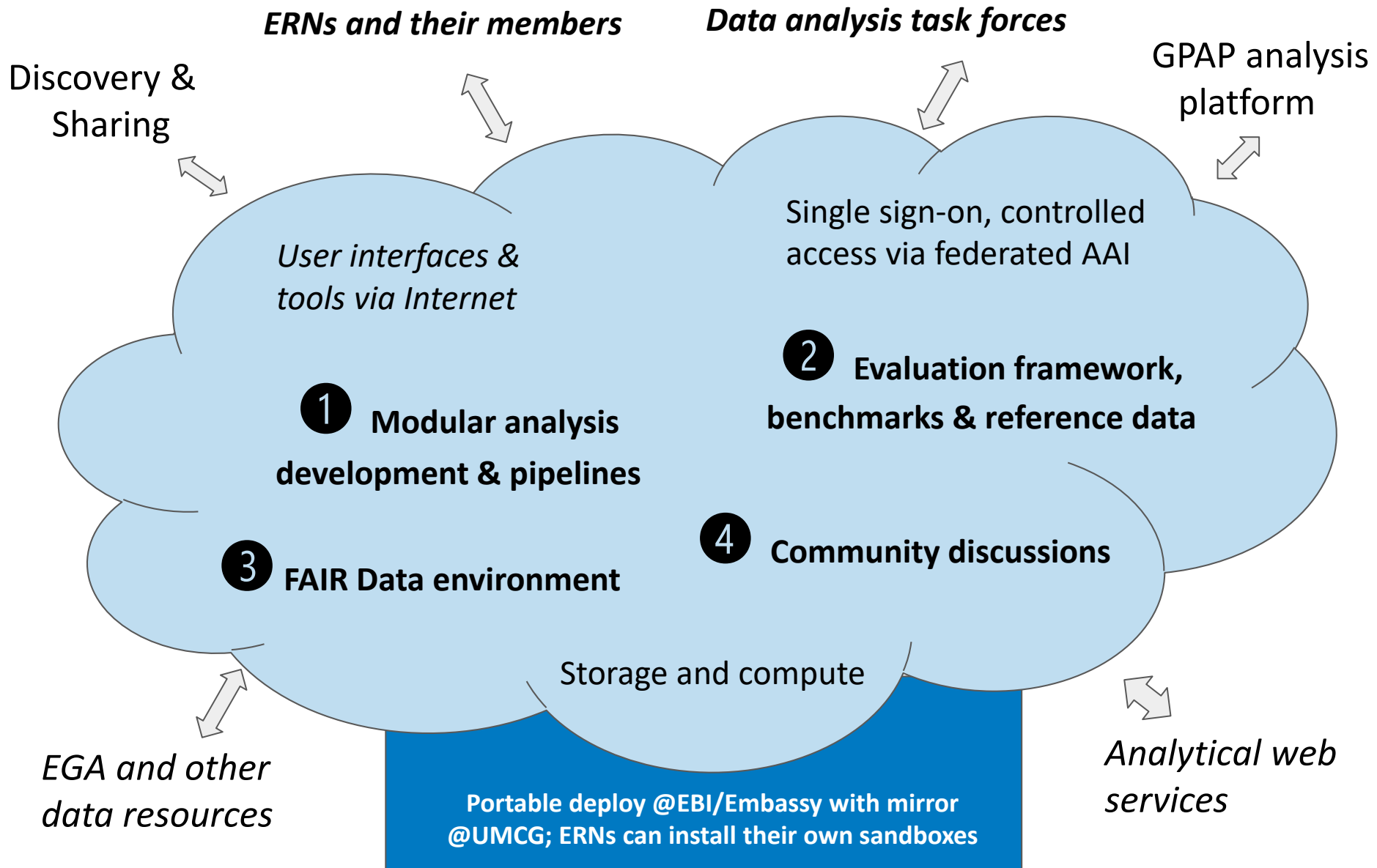Control**

Genome-Phenome Analysis Platform (GPAP)

Ethics Committee Approval, GDPR Compliance, Code of Conduct,
Data Access Committee, Security audited, User Activity logged

Genome informatics

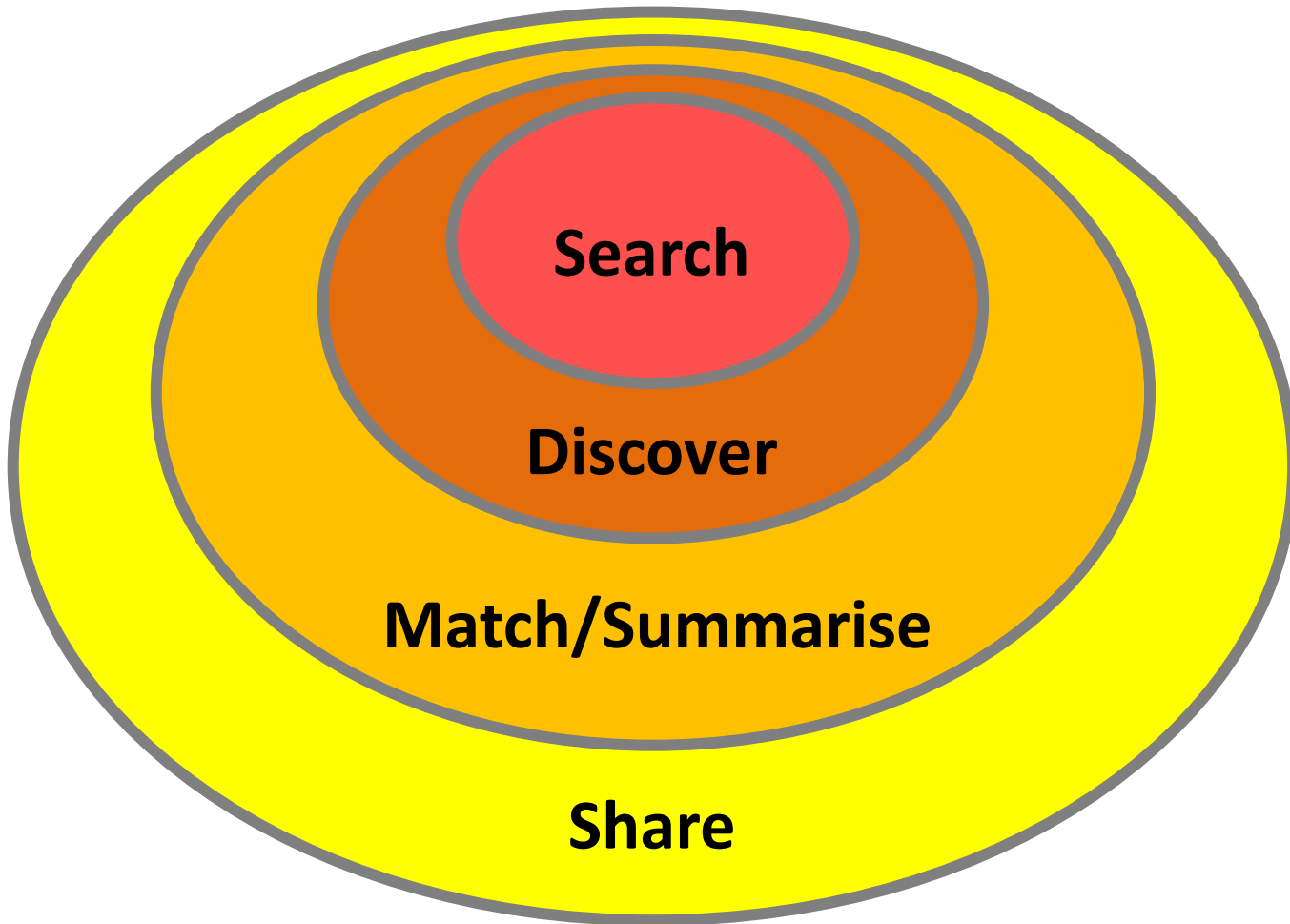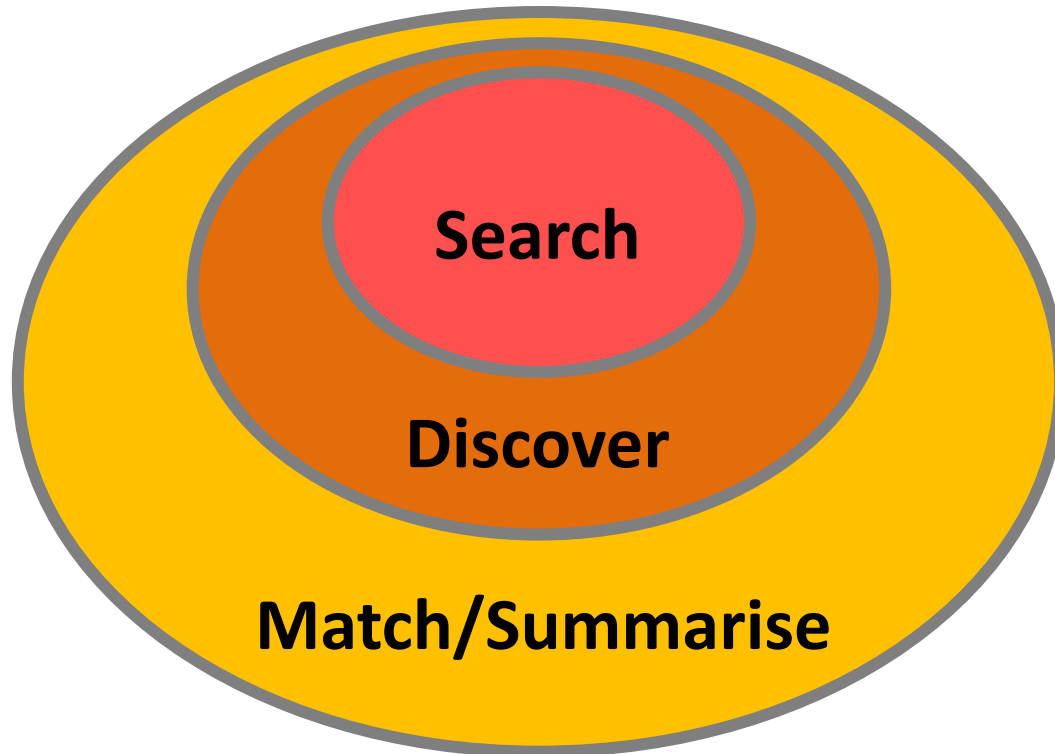*Bioinformatician think it's done*

Service fully operational

First production projects

PoC with first research project

Diagnostics think it's done

Implement candidate solution

Identify solutions

Scientist think it's done

Tools needed

Collaborative Cloud 'Sandbox'

ERNs and their members

Data analysis task forces

Discovery & Sharing

GPAP analysis platform

User interfaces & tools via Internet

Single sign-on, controlled access via federated AAI

❶ **Modular analysis development & pipelines**

❷ **Evaluation framework, benchmarks & reference data**

❸ **FAIR Data environment**

❹ **Community discussions**

Storage and compute

EGA and other data resources

Analytical web services

Portable deploy @EBI/Embassy with mirror @UMCG; ERNs can install their own sandboxes

Controlled data discovery

Controlled data discovery

Search

Discover

Match/Summarise

Café Variome software

Enable RD-related data/samples/patients to be (co-)discovered

at the METADATA &/or RECORD levels, ACCURATELY & SAFELY

...finding data & patients for sharing/collaboration

...unusual case matchmaking (extending MME)

...'RD Models & Mechanisms' (RDMM) brokerage

mtDNA Disease: Patient Discovery

DEMOGRAPHIC ⌄

| Sex ▾ | | Select/Input value ▾ |
| Age ▾ | Select operator ▾ | Select/Input value ▾ |

HAPLOGROUP ⌄

filter by keyword 🔍

```
A2q
H1
H15a1b
H1c1
H1e
H4a1a
H7
H9
I3
N1b1b
```

Add ›

‹ Remove

TISSUE ⌄

| Tissue ▾ | IS ▾ | Select/Input value ▾ |

GENOTYPE ⌄

| Select/Input position ▾ | ALT ▾ | ALT AF ▾ | Select... ▾ | Select/Input val... ▾ | ➕ |

PHENOTYPE ⌄

filter by keyword 🔍

```
HP:0000001 (All)
HP:0000005 (Mode of inheritance)
HP:0000009 (Functional abnormality of t
HP:0000014 (Abnormality of the bladder
HP:0000107 (Renal cyst)
HP:0000277 (Abnormality of the mandib
HP:0001298 (Encephalopathy)
HP:0001507 (Growth abnormality)
HP:0002401 (Stroke-like episode)
```

Add ›

‹ Remove

Solve-RD RDMM

Automatable Discovery & Access Matrix (ADA-M)

Co-Chairs: Anthony Brookes; John Wilbanks

Coordinator: Emily Kirby

Vartika Agrawal; Alessia Baretta; Moran Cabili; Knox Carey; Virginia Chavis; Ed Conley; Primavera De Filippi; Stephanie Dyke; Luca Emili; David Fajgenbaum; Clara Gaff; Chiara Garattini; Jack Goldblatt; Peter Goodhand; Elli G. Gourna; Tudor Groza; Reece Hart; James Hazard; Anneliene Jonker; Petra Kaufmann; Alastair Kent; Warren Kibbe; Bartha Maria Knoppers; Ray Krasinski; Lilian Lau; Josh Leslie; David Lloyd; Hanns Lochmuller; Alex Mankovich; Mauricio Moldes Quaresma; Francis Ouellette; Anthony Philippakis; Barbara Prainsack; Jordi Rambla; Fabien Richard; Olaf Riess; Gregory Rushton; Andrea Saltzman; Mahsa Shabani; Anne Marie Tassé; Adrian Thorogood; Mihaela Ulieru; Susan Wallace; Tomoyuki Yamada; Andreas Zankl; Vagelis Ladas; Patrick Woolley; Francis Jeanson

+ Substantial contributions from IMI(EMIF), P3G and Intel.

Journal List › NPJ Genom Med › v.3; 2018 › PMC6056554

NPJ Genom Med. 2018; 3: 17.   PMCID: PMC6056554
Published online 2018 Jul 23. doi: 10.1038/s41525-018-0057-4   PMID: 30062047

Responsible sharing of biomedical data and biospecimens via the "Automatable Discovery and Access Matrix" (ADA-M)

J. Patrick Woolley,[1] Emily Kirby,[2] Josh Leslie,[3] Francis Jeanson,[4] Moran N. Cabili,[5] Gregory Rushton,[5] James G. Hazard,[6] Vagelis Ladas,[7] Colin D. Veal,[7] Spencer J. Gibson,[7] Anne-Marie Tassé,[2] Stephanie O. M. Dyke,[8] Clara Gaff,[9,10] Adrian Thorogood,[8] Bartha Maria Knoppers,[8] John Wilbanks,[11] and Anthony J. Brookes[7]

Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer

Abstract   Go to: ⊡

Given the data-rich nature of modern biomedical research, there is a pressing need for a systematic, structured, computer-readable way to capture, communicate, and manage sharing rules that apply to biomedical resources. This is essential for responsible recording, versioning, communication, querying, and actioning of resource sharing plans. However, lack of a common "information model" for rules and conditions that govern the sharing of materials, methods, software, data, and knowledge creates a

+ API
+ Support software

Solve-RD Data Pathways

UDP + ERN network

Identifiable Individual level data

Pseudomised individual level data.

Consent
(ADA-M)
+
Deep phenotype data
+
NGS data
WES/WGS

Biosamples

GPAP

Pseudonymised individual level data

WES/WGS reanalysed via RD-connect pipelines and annotated

Deep phenotype data combined with WES/WGS

Context added using information from additional tools available in RD-Connect

ERN user analysis of datasets using RD-Connect pipelines and associated tools.

Cloud based "Sandbox"

Pseudonymised raw data for method testing and development.

European Genome-Phenome Archive

Individual level anonymous Metadata
Individual level pseudonymised/anonymised
FASTA / BAM files
and phenotype data

Federated Data Discovery Environment

Anonymised metadata and associated data on subjects and samples

EuroBioBank + Registry

Psudonymised / anonymous
Sample data