Deliverable

D2.5 Report on new matchmaking strategies	
Version | Status	V1 | final
Work package	WP2
Lead beneficiary	CNAG-CRG
Due date	31.12.2021 (M48)
Date of preparation	24.02.2022
Target Dissemination Level	Public
Author(s)	Sergi Beltran (CNAG-CRG), Steven Laurie (CNAG-CRG), Luca Zalatnai (CNAG-CRG), Anthony Brookes (ULEIC), Lennart Johansson (UMCG)
Reviewed by	Gulcin Gumus (Eurordis), Hanns Lochmüller (UKLFR)
Approved by	Alexander Hoischen (RUMC)

The Solve-RD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779257.

Explanation according to GA Annex I:
Work out best practices for evaluation and reporting of results.

Abstract:

Matchmaking technologies cover a very specific data findability need by enabling the discovery of individuals with similar phenotypes and/or potential causative genetic variants, among others. Solve-RD has contributed to the development and implementation of several matchmaking strategies which are now available to the consortium but can also be deployed in other settings. The RD-Connect Genome Phenome Analysis Platform (GPAP) is connected to the MatchMaker Exchange network (MME), which enables users, including all Solve-RD partners, to query for individuals with similar symptoms and/or candidate disease genes within the GPAP (internal matchmaking), and against four external databases: PhenomeCentral, DECIPHER, GeneMatcher, and MyGene2. The PhenoStore module allows the users to find individuals with specific categories within the RD-Connect GPAP. The CohortApp module enables the users to create *in-silico* cohorts according to certain search criteria; such cohorts can then be used to launch genetic queries within the RD-Connect GPAP analysis module. Finally, powerful search, discovery and matchmaking capabilities are provided through RD3/Sandbox and Discovery Nexus.

Introduction:

Finding available patient data is of vital importance for rare disease research, as identifying cases with common phenotypes and disrupted genes can help in solving these cases. However, important considerations regarding privacy must be addressed prior to sharing potentially sensitive information from participants between different platforms. This is why the implementation of Matchmaking technologies, which allow for defined queries between different databases without revealing important sensitive identifiable data, are so crucial nowadays.

Matchmaking involves making it possible to remotely query a minimal amount of genotypic and phenotypic information that allows the discovery of patients with similar phenotypes and/or similar genetic variants. For instance, if a user of the RD-Connect GPAP identifies a potential causative variant(s) in a gene of interest for one of their cases, but they lack further information to indicate whether this variant/gene really may be the cause of the rare disease, they can query the other nodes available via the MME API to ask if they have a patient with similar phenotypic characteristics who may also have a variant in the same gene. A score for the match is calculated for all patients in the receiving database, and any score above a certain threshold results in a match being returned, and a contact email with the relevant details of the match being generated automatically. Once the match has been generated, it is then in the hands of the relevant researchers/clinicians to follow-up. The RD-Connect GPAP implementation of the MME API, only considers the data from those participants/individuals for which the data submitter has selected that they can participate in matchmaking.

Since the beginning of the Solve-RD project, the partners have greatly increased the number of strategies available to them for matchmaking. As well as matchmaking internally within the RD-Connect GPAP (with cases submitted by other users), they now have connections from the RD-Connect GPAP to several nodes of the MatchMaker Exchange network (MME;

https://pubmed.ncbi.nlm.nih.gov/26295439/), GPAP PhenoStore, GPAP CohortApp, the Rare Disease Data about Data (RD3) system and Discovery Nexus. In addition, Solve-RD has developed novel matchmaking functionalities, which are described in the confidential Deliverable D2.20.

Report:

RD-Connect GPAP connections to MatchMaker Exchange (MME)

The RD-Connect GPAP offers both internal and external matchmaking through the implementation of MME v1.1 API. As part of Solve-RD we have established new connections to two American nodes - myGene2 (University of Washington) and GeneMatcher (Johns Hopkins University; Figure 1) - and matchmaking has been performed in both directions. Together with previously existing connections to PhenomeCentral (Canada) and Decipher (UK), in addition to the internal matchmaking functionality within the RD-Connect GPAP, this has markedly increased the number of patients available to match with externally and hence the probability of finding a successful match with a remote rare disease case.

There are currently 7,511 affected individuals that are available for matchmaking via the MME network in the RD-Connect GPAP.

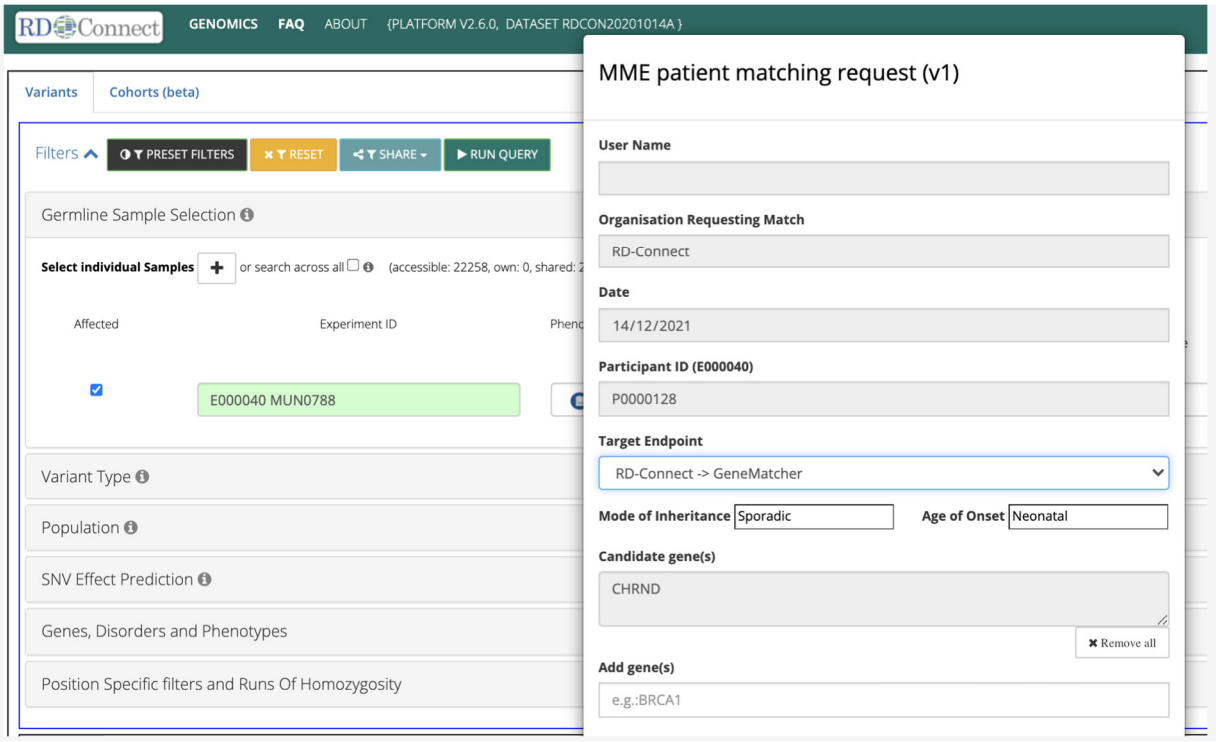

Figure 1. *Screenshot of the GPAP Genomics module showing the opportunity of external matchmaking through GeneMatcher.*

RD-Connect GPAP PhenoStore search capabilities

With the development of the RD-Connect GPAP PhenoStore module (developed through EJP-RD), the ease with which users can query their own data has been significantly improved in comparison to previous functionality available to RD-Connect GPAP users for phenotypic data.

Users can now search directly on a wide variety of fields, including by owner (for cases where data has been shared), user-supplied Family and Individual identifiers, the suspected mode of inheritance, etc (Figure 2). Furthermore, users can decide which fields they wish to be displayed in the main table view, and likewise when they export they can choose the subset of fields in which they are interested.

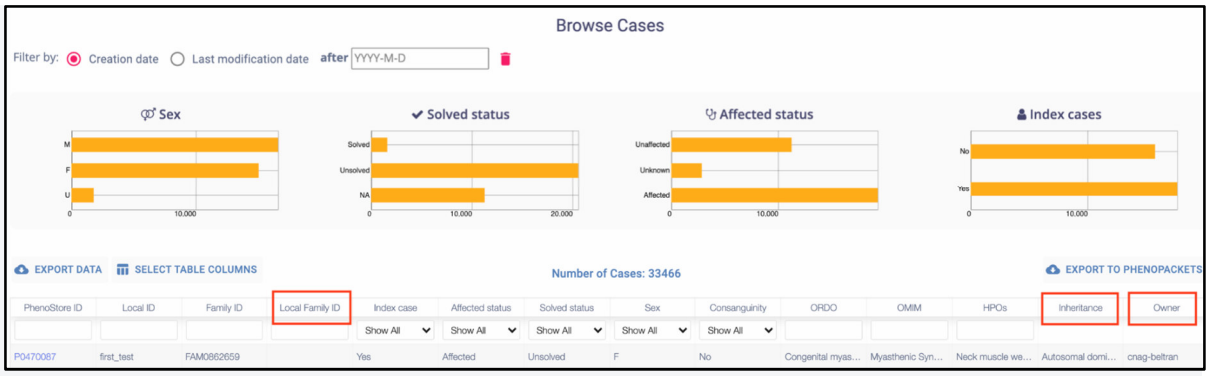

Figure 2. *Screenshot of the GPAP PhenoStore module showing the newly introduced search components.*

RD-Connect GPAP Cohort App

The Cohort App module has been released within the RD-Connect GPAP at the beginning of the year 2021. This module is designed to facilitate the exploration and construction of cohorts on experiments metadata and structured clinical data using standard ontologies. This means that a group of patients matching a certain set of criteria (e.g. variants in a specific gene or with a certain constellation of phenotypic features) can be constructed and saved for further analysis. Several improvements have been made to this feature. On the server side, we have improved the performance of the filtering functionality by harmonising the intersection of experiment metadata with the participants' phenotypic data. We have also created a dedicated API to create, save, and export cohorts in CSV format in an interactive way, allowing such cohorts to be used as the input to an analysis within the GPAP analysis module. The User Interface (UI) and the User Experience (UX) have been further adapted to enable filtering by multiple field values and by combinations of ontology terms (OMIM, HPO, ORDO).

The main development of the Cohort App module was done in the EJP-RD project, although we have added here the ability to filter the experiments by Solve-RD ERN (Figure 3).

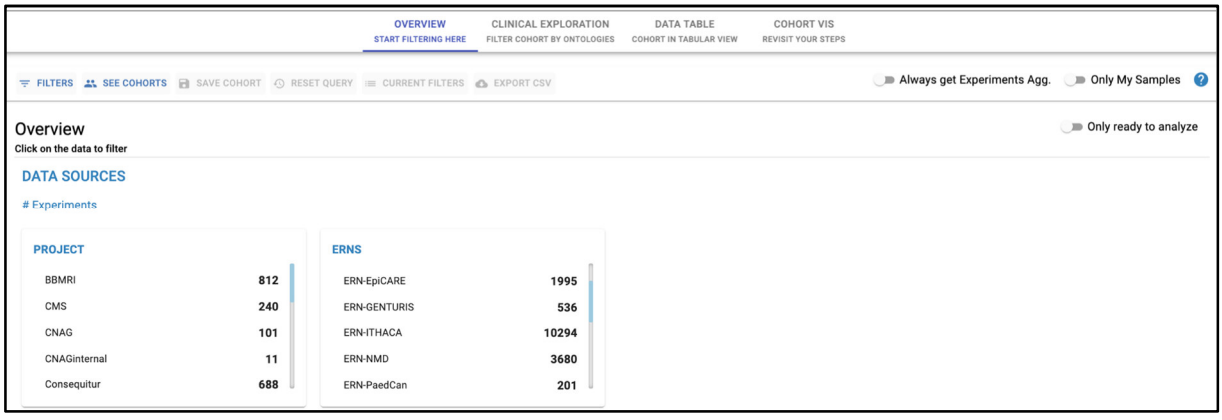

Figure 3. *Screenshot of the GPAP Cohort App module showing the ERN component.*

Matchmaking within RD3 and Discovery Nexus

Discovery services have been implemented in RD3/Sandbox using modules from the RD-Nexus platform, as explained in D2.20 (confidential report) and D4.8 (public report), with the combined discovery system being called Discovery Nexus. A joint login (via FusionAuth single-sign on login: https://fusionauth.io/) has been established which provides a convenient user experience that gives access to extensive search capabilities. When using this service, a user has to explicitly specify which ERN datasets are to be searched, in order to prevent unintended incidental findings. This approach is currently only available for Solve-RD partners, but the aim is to expand it beyond that.

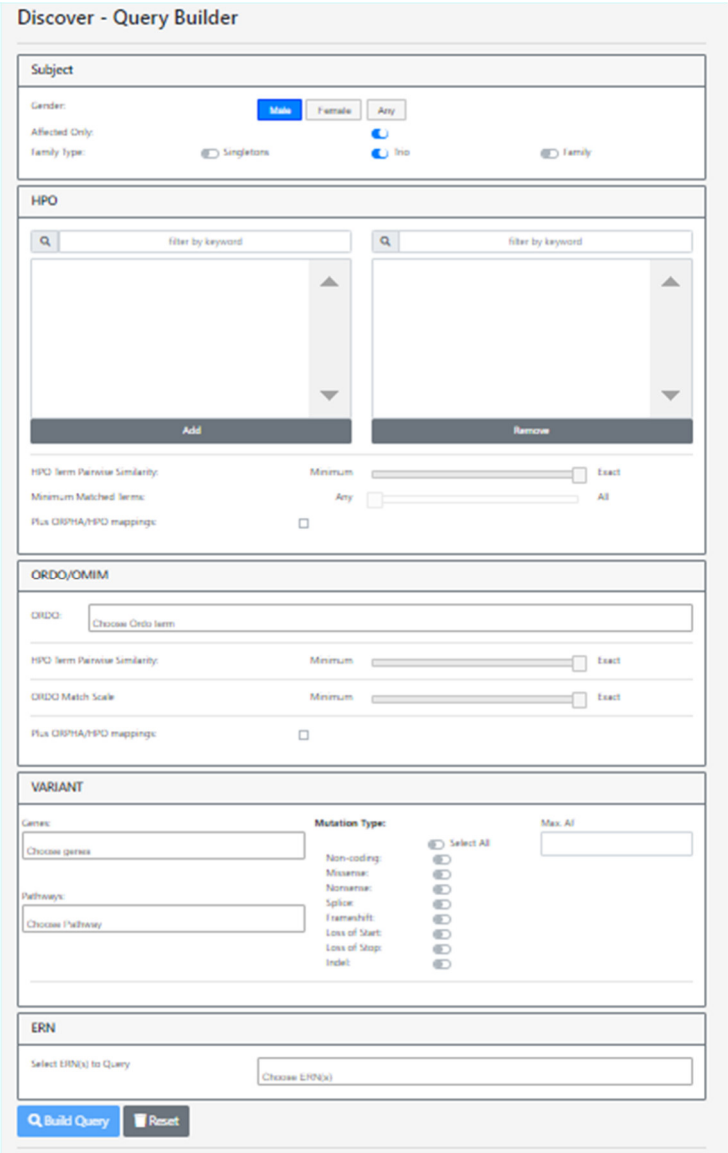

Figure 4. *Screenshot of the Discovery Nexus showing the search fields.*

This Discovery Nexus system can be used for many things, including RD patient matchmaking. Within RD3 itself, simple filtering options are also provided so that users can directly show or hide data columns or search for specific samples/records. More extensive search capabilities that can be used for cohort building are available via Discovery Nexus, using five fields to specify which are the samples of interest. First of all, lists of phenotypes and/or diseases can be entered. The subjects that will be included in the cohort will be determined by these phenotypes or diseases. Using two sliders it can be determined if a patient needs to have any

to all of these phenotypes or diseases and if they should have a more or less exact match with the entered phenotypes or diseases.

Additionally, detected variants can be used to build a cohort. For instance, if a subject contains a variant below a given allele frequency in a selected gene or pathway, it will be included in a cohort. Given the nature of the data in RD3 (i.e., variants do not have a classification) then these cannot be directly filtered upon (e.g., pathogenic variants in gene X). Nevertheless, categorical querying for variants is instead supported via the Discovery Nexus interface. Filtering by sex and by whether the case is a singleton, or part of a trio or a family, is also possible.

Conclusion:

Solve-RD has contributed to the development and the implementation of new matchmaking strategies, providing additional data discovery capabilities for the consortium and beyond. Some of these strategies, like the RD-Connect MME connections, PhenoStore Search and CohortApp are already available to both Solve-RD and non-Solve-RD partners, although not all users have access to exactly the same set of data. Other strategies, like Matchmaking within RD3 and Discovery Nexus are for now restricted to Solve-RD users within the Solve-RD infrastructure, but the technology is ready to be implemented for other initiatives or within other settings.