Deliverable

D1.4 Deployment of PhenoTips custom forms according to the ERNs specifications	
Version | Status	V1 | final
Work package	WP1
Lead beneficiary	CNAG-CRG (Sergi Beltran)
Due date	31.12.2018 (M12)
Date of preparation	18.12.2018
Target Dissemination Level	Public
Author(s)	Ricky Joshi (CNAG-CRG), Sergi Beltran (CNAG-CRG)
Reviewed by	Holm Graessner (EKUT), Alexander Hoischen (RUMC)
Approved by	Ana Rath (INSERM-Orphanet)

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:

In this task we will make the necessary adaptations to PhenoTips allowing for the collection of phenotype data from unsolved cases coming from ERNs. The requirements for the templates will be prepared by the ERNs.

Abstract:

We describe here the adaptations and customisations we have carried out in PhenoTips to allow for the collation of phenotype data tailored to each ERN and disease group. Users can enter data individually for each proband and family member through one of the customized templates, import information using JSON schemas or provide a filled-in Excel template for bulk upload.

Introduction:

Solve-RD is aiming to collate 19,000 exome/genome datasets and associated phenotypic/clinical information from unsolved rare disease patients with the goal to solve as many cases as possible. This data is currently provided by members from four European Research Networks (ERNs). The ERNs are: ERN-ITHACA (rare congenital malformation syndromes and intellectual disability), ERN-NMD (neuromuscular disorders), ERN-RMD (rare movement and cognitive disorders) and ERN-GENTURIS (highly heritable rare genetic tumor risk syndromes). For Solve-RD to reach this goal, comprehensive phenotypic data acquisition is paramount. Prior to Solve-RD, researchers could only upload phenotypic data through the RD-Connect's PhenoTips instance graphical user interface. However, this process can be somewhat time-consuming and repetitive, especially for large datasets and may discourage in-depth phenotypic data entry. We have therefore made developments to create customized PhenoTips forms tailored to the ERN needs, have enabled import of data in JSON format and have created and Excel template for bulk upload.

Report:

We have made a series of adaptions to allow for individual data entry through disease-customised PhenoTips forms, to import data in JSON format and to bulk upload data by means of an Excel template.

1) Customised disease group PhenoTips templates tailored to ERN needs

In order to facilitate data entry through the PhenoTips graphical user interface, we have created customized entry form templates tailored to the needs of the ERNs and the diseases they address. These forms contain pre-defined sections and sets of Human Phenotype Ontology (HPO) terms provided by ERN experts. The HPOs that can be easily clicked by the user to be present or absent in the patient. However, the user is not limited to the HPOs shown in first instance, and can easily navigate to lower categories/levels and search for and select any other HPO term available in the whole ontology. We created a document to prepare customized forms and the mandatory sections and fields to be included (working document available at https://drive.google.com/file/d/1Dc6jJVyTcvvOEVSi0uU5WYy15Y3Xxyum/view).

Researchers can select a customized PhenoTips template easily in PhenoTips (https://platform.rd-connect.eu/phenotips/), clicking "create" and then "new patient". This will invoke a pop-up where researchers can select the disease relevant for each patient (see Fig1). Each ERN can select from the following forms:

- ERN-ITHACA:

ID_epilepsy_ataxia_HSP

- ERN-NMD:

Congenital muscular dystrophy and congenital myopathy

Congenital myasthenic syndrome

Muscular dystrophy

Muscular channelopathies

Spinal muscular atrophy and lower motor neuron disease

Titinopathies

- ERN-RND:

Ataxia

Frontotemporal lobe dementia

Hereditary motor neuropathies

Hereditary spastic paraplegia

- GENTURIS:

Tumour risk syndromes

Some of the pre-determined terms have been used in other projects and meticulously modified and standardized for Solve-RD. The ERN-GENTURIS template however, was created especially for Solve-RD by experts in the field.

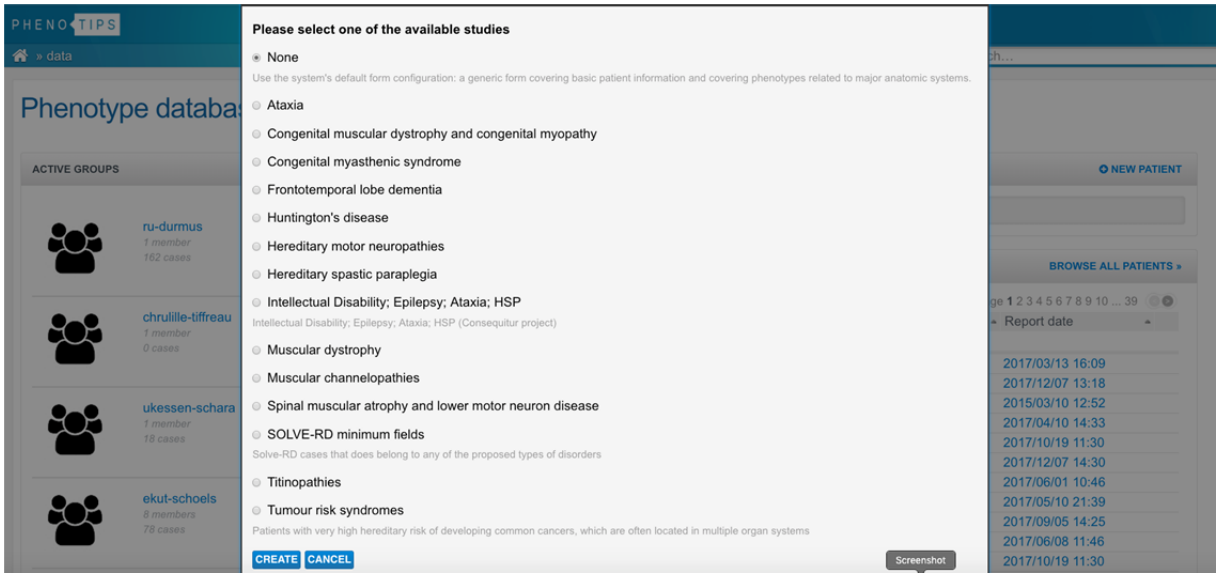


Fig1. Screen shot of ERN customized forms for phenotyping of Solve-RD cases.

Fig.2 shows a screenshot of the Clinical Symptoms section for the PhenoTips template for Ataxia, displaying the pre-selected HPO terms (ERN-NMD).

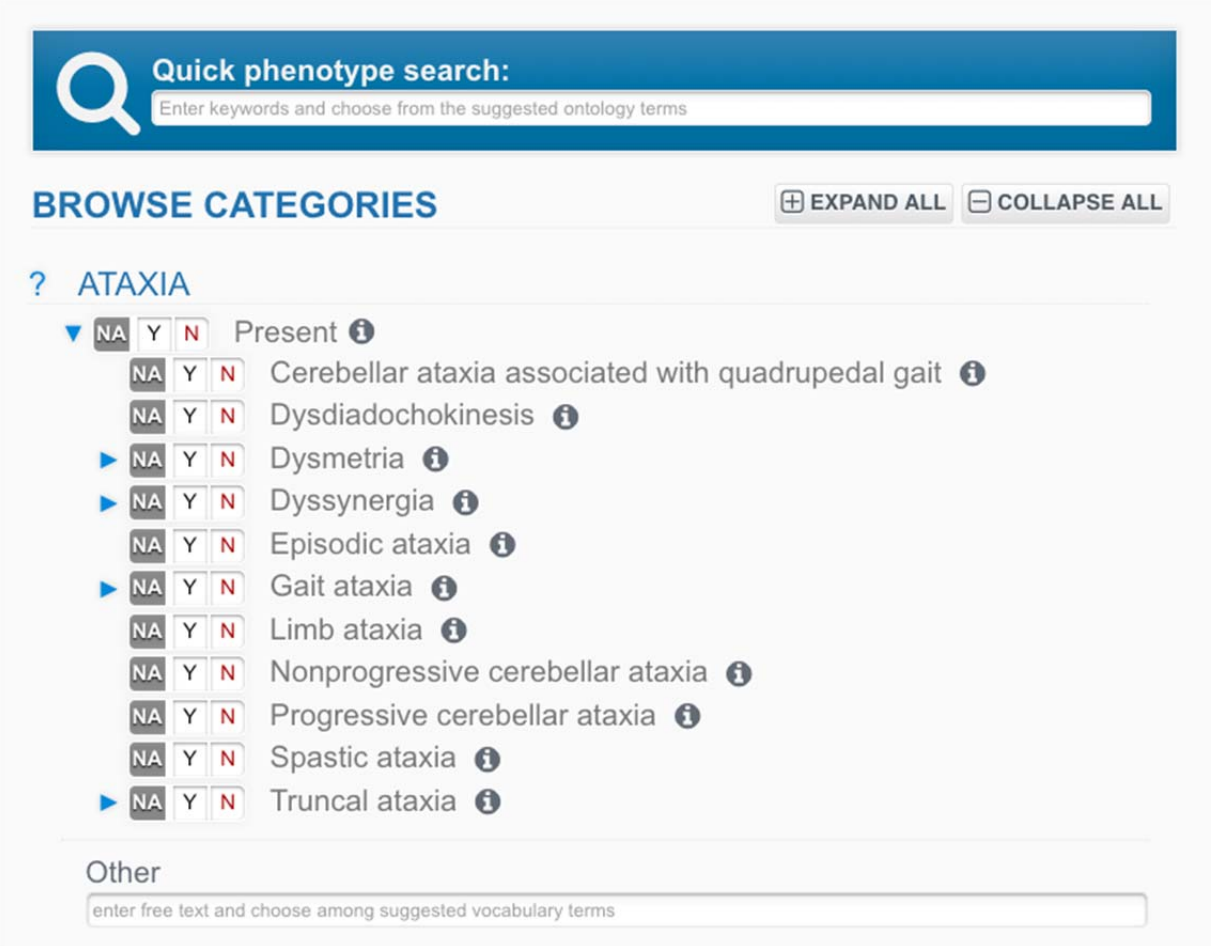


Fig2. Pre-determined clinical phenotypes for patients with Ataxia.

If Solve-RD partners do not find their particular group of diseases, they have the option of selecting “SOLVE-RD minimum fields”. This will return a default template for Solve-RD. It’s worth noting that Solve-RD requires six mandatory fields to be completed (local family ID, Local ID, sex, clinical status, Date of birth and Positive HPO terms - minimum 5) that are essential for providing high-quality phenotypic data.

We will be able to periodically update these sets of pre-selected terms according to the needs of the researchers and the Solve-RD project.

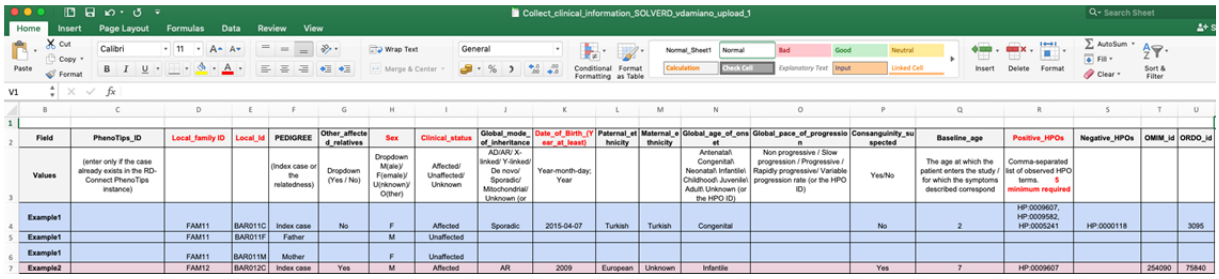
2) Import existing phenotypic data through JSON format

To further assist the ERNs, researchers who have their data already in another PhenoTips database or in PhenomeCentral can easily export their data from those instances in JSON format and send it to us to import in the RD-Connect’s PhenoTips instance.

3) Bulk data upload using an Excel template

We have also created an excel spreadsheet for the submission of phenotypic data for all Solve-RD partners. Phenotypic data collated in Excel is converted to JSON format by our in-house bioinformatic pipeline and rapidly uploaded into PhenoTips instance. Users are subsequently provided with PhenoTips ID’s for all their cases.

For correct excel usage, we require at least six mandatory fields to be completed to ensure high quality phenotypic data is being collected (see above). Below, we provide an example of the clinical data required for Excel upload.



Field	PhenoTips_ID	Local_family ID	Local_id	PEDIGREE	Other_affected_relatives	Sex	Clinical_status	Global_mode_of_inheritance	Date_of_Birth_(Year_at_least)	Paternal_ethnicity	Maternal_ethnicity	Global_age_of_onset	Global_pace_of_progression	Consanguinity_suspected	Baseline_age	Positive_HPOs	Negative_HPOs	OMIM_id	ORDO_id
Values	(enter only if the case already exists in the RD-Connect PhenoTips instance)			(Index case or the relatedness)	Dropdown (Yes / No)	Dropdown M(ale)/ F(emale)/ U(nknown)/ O(ther)	Affected/ Unaffected/ Unknown	AD/AR/ X-linked/ Y-linked/ De novo/ Sporadic/ Mitochondrial/ Unknown (or	Year-month-day; Year			Antenatal\ Congenital\ Neonatal\ Infantile\ Childhood\ Juvenile\ Adult\ Unknown (or the HPO ID)	Non progressive / Slow progression / Progressive / Rapidly progressive/ Variable progression rate (or the HPO ID)	Yes/No	The age at which the patient enters the study / for which the symptoms described correspond	Comma-separated list of observed HPO terms. 5 minimum required				
Example1		FAM11	BAR011C	Index case	No	F	Affected	Sporadic	2015-04-07	Turkish	Turkish	Congenital		No	2	HP:0009607, HP:0009582, HP:0005241	HP:0000118		3095
Example1		FAM11	BAR011F	Father		M	Unaffected												
Example1		FAM11	BAR011M	Mother		F	Unaffected												
Example2		FAM12	BAR012C	Index case	Yes	M	Affected	AR	2009	European	Unknown	Infantile		Yes	7	HP:0009607		254090	75840

Fig3. Data required for PhenoTips. Note, all fields in red denote mandatory fields.

For both online and bulk data upload through Excel, we provide online training modules, guidance documents and help tools. Online PhenoTips guidance can be found here: (https://rd-connect.eu/phenotips-guide/) and a YouTube tutorial here: (https://www.youtube.com/watch?v=wvTJXxtuI8Y). Webinars are also organized to train users on providing phenotypic data, and support is available at help@rd-connect.eu. See deliverable D1.3 for further details.

Conclusion:

We have implemented a number of modifications to PhenoTips in line with the needs of each ERN and the Solve-RD project. We provide user-friendly customized templates and the ability to provide data in JSON format or in an Excel template. We will continue to make changes to improve the phenotypic entry experience and motivate scientists to provide as profound phenotypic detail for their cases as possible.