EasyChair Preprint

№ 44

# SARAEasy: A mobile app for Cerebellar Syndrome Quantification and Characterization

Haitham Maarouf, Vanessa López, Maria J Sobrido, Diego Martínez and Maria Taboada



April 5, 2018

# SARAEasy: A mobile app for Cerebellar Syndrome Quantification and Characterization

Haitham Maarouf[1], Vanessa López[1], Maria J Sobrido[2], Diego Martínez[3] and Maria Taboada[1]

[1]Department of Electronics & Computer Science, Campus Vida, University of Santiago de Compostela, Spain, haitham.maarouf@rai.usc.es, vanessalp16vlp@gmail.com, maria.taboada@usc.es

[2]Instituto de Investigación Sanitaria (IDIS), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Santiago de Compostela, Spain, ssobrido@gmail.com

[3]Department of Applied Physics, Campus Vida, University of Santiago de Compostela, Spain, diego.martinez@usc.es

**Abstract.** The assessment of latent variables in neurology is mostly achieved using clinical rating scales. Mobile applications can simplify the use of rating scales, providing a quicker quantitative evaluation of these latent variables. However, most health mobile apps do not provide user input validation, they make mistakes at their recommendations, and they are not sufficiently transparent in the way they are run. The goal of the paper was to develop a novel mobile app for cerebellar syndrome quantification and clinical phenotype characterization. SARAEasy is based on the Scale for Assessment and Rating of Ataxia (SARA), and it incorporates the clinical knowledge required to interpret the patient status through the identified phenotypic abnormalities. The quality of the clinical interpretation achieved by the app was evaluated using data records from anonymous patients suffering from SCA36, and the functionality and design was assessed through the development of a usability survey. Our study shows that SARAEasy is able to automatically generate high-quality patient reports that summarize the set of phenotypic abnormalities explaining the achieved cerebellar syndrome quantification. SARAEasy offers low-cost cerebellar syndrome quantification and interpretation for research and clinical purposes, and may help to improve evaluation.

**Keywords:** SARA, Health App, Rating Scales, Human Phenotype Ontology, Clinical Archetype.

# 1 Introduction

Clinical rating scales play a significant role in collecting standardized data, mainly in neurology. They are also used to measure the so-called latent variables, those that cannot be directly observed and must be inferred from other variables. An example of latent variable is ataxia, i.e., the lack of voluntary ability to coordinate muscle movements. The assessment of this latent variable is made indirectly through some questionnaire covering a set of clinical statements or items [1] that can be directly observed, such as abnormal gait or loss of balance. For example, the Scale for Assessment and Rating of Ataxia (SARA) [2] is a survey evaluating motor performance in patients suffering ataxia, by adding scores resulting from the assessment of eight items. Most rating scales used in neurology are ordinal scales, providing facilities to rank patients in degrees of disability according to certain external criteria. The strategy of this type of scales is to obtain a single score (total score) that characterizes an individual. The total score of SARA ranges from 0 (no ataxia) to 40 (most severe ataxia). This strategy is very attractive, although it is somewhat ambiguous in that two individuals with different clinical conditions may have identical scores through different combinations of items. Let's consider two patients with the same clinical stage, for example, a total SARA score of 15 (moderate cerebellar syndrome). This does not ensure that their functional situation is similar. For example, one of them could barely walk or sit unaided (notably midline ataxia), whereas the other patient could have compromised speech and limb coordination. In both cases, this functional differentiation could be inferred from the data collected into the rating scale, thought it is not reflected in the total SARA score. Therefore, in cases like this, the numerical result of a scale is not enough to describe the patient's condition accurately. Therefore, inferring qualitative descriptions for classifying patients with diagnostic implications from numerical scores collected using rating scales is a challenge. A solution to overcome this drawback is to incorporate the knowledge required to interpret the numerical assessments. These interpretations could be provided in various formats, depending on their use. If this is focused on the preparation of patient reports, its most appropriate format is a textual summary. SARAEasy provides numerical assessments of cerebellar syndrome following the Scale for Assessment and Rating of Ataxia (SARA), and it incorporates the clinical knowledge required to also provide accurate clinical interpretations of the achieved numerical assessments. The outcome provided by SARAEasy can be directly incorporated in the patient report.

# 2 Methods

Ensuring safety and quality of medical apps represents a challenge [3]. For example, current insulin dose calculators do not provide user input validation, they make mistakes at the dosage recommendation, and they are not sufficiently transparent in their operation (i.e., in the formulas they are using) [4]. With the aim of guaranteeing patient safety, SARAEasy was based on three basic premises: transparency of the used

algorithms and knowledge, data sharing, and data quality. In addition, patient data privacy represented a key principle on which our approach is based. For patient safety purposes, SARAEasy was substantiated by the Scale for Assessment and Rating of Ataxia (SARA), a well-validated instrument to assess the presence and severity of cerebellar ataxia [5]. Firstly, we modeled the SARA using openEHR archetypes [6], as they promote computational data standardization, comparison of results across studies [7] and integration of information from different sources. Archetypes support interoperability and can be re-used across many types of healthcare applications. Secondly, mapping the archetype data structures to some ontology facilitates the automated clinical interpretation of the patient status and it omits the ambiguity in interpretation, especially when several patients have the same total score. We decided to choose the Human Phenotype Ontology (HPO) [8] to represent the meaning of the SARA clinical items. For patient data privacy purposes, SARAEasy was designed to be patient data access free. It only uses a coded identity to incorporate into the final report. The code is not stored into the app.

### 2.1 Modeling the SARA by clinical archetypes

Currently, the main approach to develop a new archetype follows a classical methodology [9]: 1) analysis of the clinical domain and requirements, identifying the archetype content (clinical concepts and their organization) from different sources (literature, record forms, etc.), 2) selection of the archetype type, structuring the content according to the archetype type, and 3) filling of the parts of the archetype with the content. A knowledge engineer analyzed the SARA scale with the help of a neurologist, determining all the entities that should be represented, and both of them organized and structured the contents. All these entities are shown in the Data part in Fig 1. After the analysis process, the type of the archetype was chosen. An *Observation* archetype was selected as it can record directly measurable data. Then, the archetype was filled with metadata, including purpose, keywords, definition and authors, among other information, with the help of the openEHR Archetype Editor [10]. All the archetype entities were modeled using the *Element* structure. Then, the Elements were defined with proper data types, descriptions, comments, details, occurrences, constraints and possible values. We used two data types: *Quantity* for the items corresponding to arithmetic averages and total score, and *Ordinal* for the rest of items. The developed archetype, named SARA, was submitted to the Clinical Knowledge Manager (CKM) [11], a system for collaborative development, management and publishing. After revision and some modifications, the archetype was accepted for publication in the CKM, where it is publicly accessible [12]. Fig. 1 displays the mind map representation of the Observation archetype.

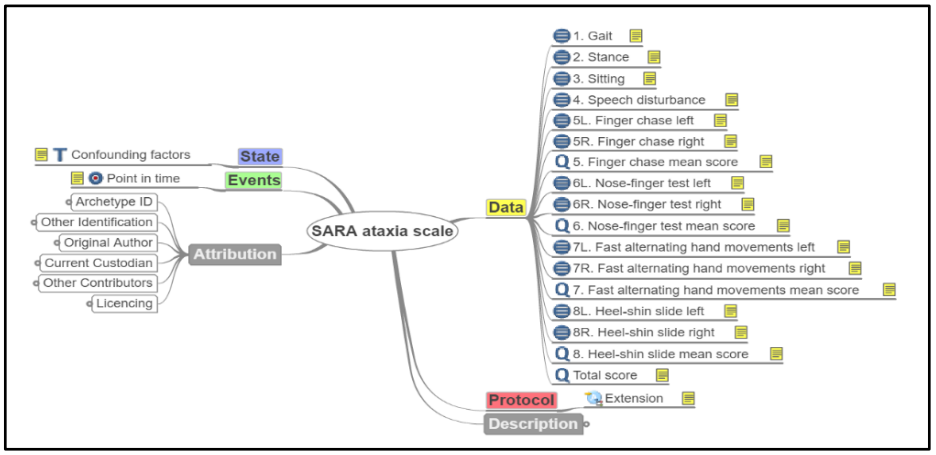


**Fig. 1.** The mind map representation of the SARA Observation archetype, which is publicly accessible in [12].

### 2.2 Mapping the archetype SARA to the HPO

Firstly, we used the concept recognition system OBO Annotator [13] to annotate all the relevant information about the SARA survey with the HPO classes. The annotated classes constituted the seed terms required to extracting the HPO subontology that was relevant to the SARA. Next, a neurologist revised the extracted subontology, proposing a minimal extension and reorganization of the subontology classes. Additional classes and subClassOf relationships are summarized in Fig. 2.

In addition to the classes that were directly related to the eight SARA items, the neurologist identified three classes especially relevant for clinical interpretation of cerebellar ataxia (Fig. 3): 1) *Truncal Ataxia*, which subsumes *Gait Ataxia*, *Standing Instability* and *Sitting Imbalance*, 2) *Appendicular Ataxia*, which subsumes *Dysdiadochokinesis*, *Intention Tremor* and *Limb Dysmetria*, and 3) *Dysarthria*. In order to determine the levels of severity that a patient has, we reused the HPO class called Severity and the subclasses *Borderline*, *Mild*, *Moderate*, *Severe* and *Profound*. New classes were created based on the severity levels of their superclasses. For example, *Borderline Sitting Imbalance*, *Mild Sitting Imbalance*, *Moderate Sitting Imbalance* or *Severe Sitting Imbalance*. We translated the HPO subontology to Protégé [14], and we checked the ontology consistency with the Hermit reasoner [15].

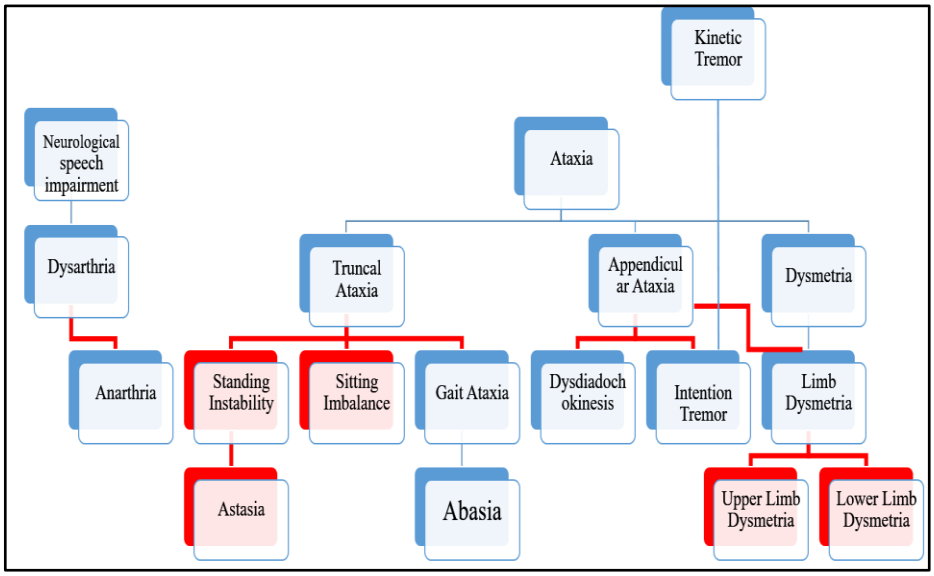


**Fig. 2.** Structure of the extended HPO Ontology Modules. The blue rectangles represent the original HPO classes, the blue lines represent the original relationships, the red rectangles represent the added classes and the red lines represent the added relationships.

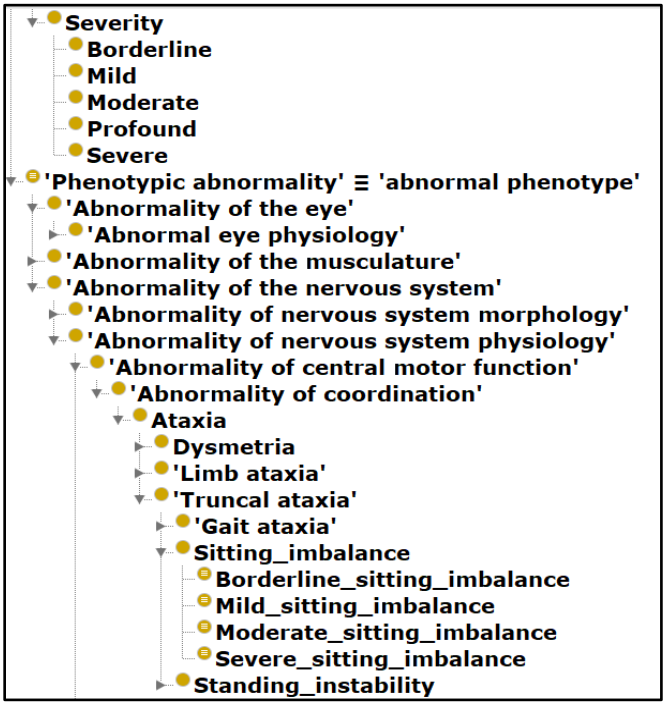


**Fig. 3.** Excerpt from the domain ontology

## 3 Result

To demonstrate the functionality of our approach, we developed SARAEasy, a mobile app for cerebellar syndrome quantification and clinical phenotype characterization (Fig 4). A proof of concept test was carried out to show that the SARAEasy app is able to identify and report cerebellar ataxia characteristics such as midline or appendicular ataxia in a straightforward and effortless way. SARAEasy faithfully mirrored the SARA archetype and was developed in Android Studio, the integrated development environment for Google's Android operating system (version 2.2). As with all Android apps, it comprised a set of interconnected activities, where most of them were presented to the user as full-screen windows. In order to protect the confidentiality of patient data, SARAEasy was implemented for not handle patient data and only request a coded identity to include into the final report plus an e-mail account for submitting the questionnaire data and the final report (Fig 5).

Four types of activities can be distinguished in SARAEasy: item entry, item query, questionnaire modification/deletion, and questionnaire submission via e-mail. A different activity was designed for each item in the rating scale (Fig 6), equipped with explanations and links to videos (Fig 7). As for some items, SARA distinguishes between right and left side (mirroring the archetype), hence twelve item entry activities were designed. Additionally, several activities were added to facilitate the navigation through the questionnaire. Finally, the questionnaire submission activity involves all the logic that is associated with the cerebellar syndrome quantification and characterization. It implements all the information-processing units required to automatically interpret the data obtained through the item entry activities. The modeling of these units can be revised in [16]. Both the final report (Fig 9) and the total score (Fig 8) can be visualized before submitting by e-mail. An SQLite database was designed for storing the collected and inferred data while the app is active. The database accesses are performed as background operations to avoid any slowdown.

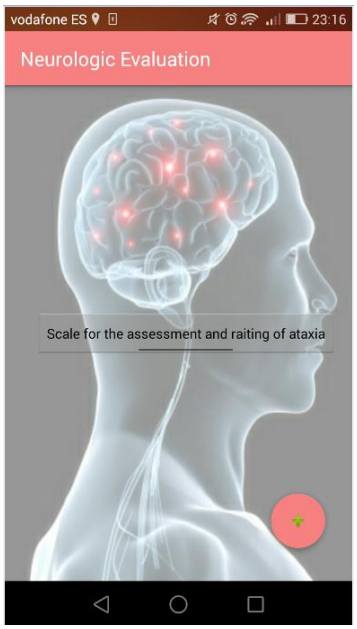


**Fig. 4.** Main Activity

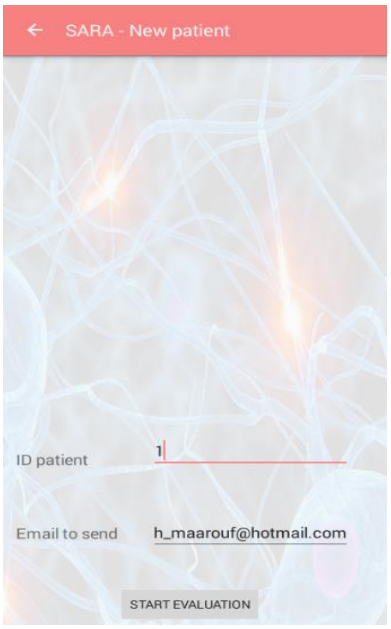


**Fig. 5.** Entry Activity for coded identification and e-mail account


**Fig. 6.** Entry Activity for the item *Sitting*

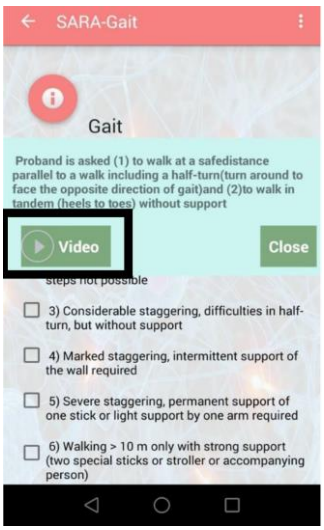


**Fig. 7.** Access to an explanation

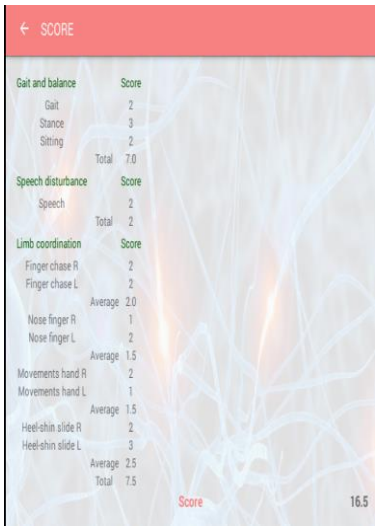


**Fig. 8.** Total score Activity

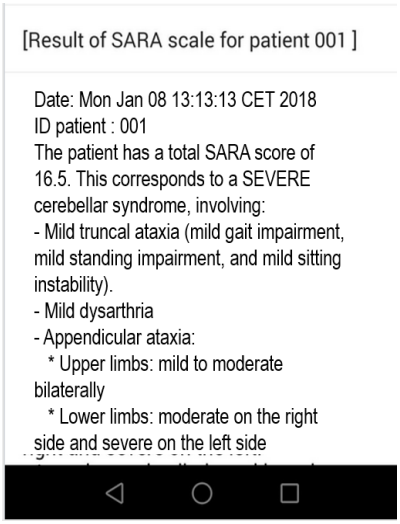


**Fig. 9.** Patient report automatically generated by SARAEasy

### 3.1 Dataset and Validation of SARAEasy

For validation purpose, two types of assessments were carried out. First, two independent neurologists validated the quality of the achieved results using data records from 28 anonymous individuals suffering from Spinocerebellar Ataxia Type 36 (SCA36). Additionally, a usability survey was designed to evaluate the functionality and design of the app.

The evaluation of the quality of the achieved results was carried out in three steps:

1. Inference of the scores and reports: We filled out the score data for each patient and extracted the following results: 1) the severity for each item, 2) the severity of cerebellar syndrome, 3) the severity of truncal ataxia, and 4) the severity of appendicular ataxia on the right and left sides. The results were translated to an excel file.
2. Interpretation by two independent neurologists: The total score data was sent to two neurologists, which used their expertise in ataxia to determine the severity of the cerebellar syndrome, and truncal and appendicular ataxia, if present, from the provided scores.
3. Comparison of results between the system and the human experts: The interpretations of the system and of the neurologists were imported into SPSS [17], and Weighted Kappa test [18] was executed 12 times to measure the strength of agreement between the implemented system and each neurologist, and between the two neurologists themselves. Weighted kappa scores ranged from 0.62 to 0.86.

**Usability surveys.** The assessment of the functionalities and design of the app was carried out through the development of a usability survey (Table 1). The design of its questions was based on software usability questionnaires [19], especially on the System Usability Scale (SUS) questionnaire [20], which was developed by John Brooke in 1986 as part of the introduction of usability engineering to the systems of Digital Equipment Co. Ltd. The survey consisted of 13 items. These items include the assessment of diverse aspects: language, colors, icons' images, terminology, speed, error messages and iteration. Each of these items has four possible answers, which are ranged from 1 to 4, according to the degree of agreement (1 Strongly disagree, 2 Disagree, 3 Agree, 4 Strongly agree). The maximum overall score for the survey is 52. The app was evaluated as a proof-of-concept study with two experts and five inexpert users. All users received detailed explanations about the goal and the functionality of the app. The obtained scores from expert and inexpert users were greater than 75% of the total score, which is therefore considered to have been successfully assessed.

**Table 1.** Usability Survey

| **Items** | **Achieved Average Rating** | | |
|---|---|---|---|
| | **Expert** | **Inexpert** | **Total** |
| 1. I think it is an easy application to use. | 3.5 | 3.6 | 3.6 |
| 2. Regarding the language, I could understand the application | 4.0 | 3.2 | 3.6 |
| 3. The application is too simple | 4.0 | 3.0 | 3.5 |
| 4. The representation of the icons of the application concerning their functions. | 3.0 | 3.0 | 3.0 |
| 5. The Structure and organization of the system | 3.5 | 3.8 | 3.7 |
| 6. I did not find useless buttons or tabs. | 4.0 | 3.8 | 3.9 |
| 7. The chosen color range is correct since the texts and the other elements of the application are clearly visible | 4.0 | 3.8 | 3.9 |
| 8. The presentation of the product is pleasant and not shabby. | 4.0 | 3.8 | 3.9 |
| 9. I did not need any help to manage the program | 2.5 | 2.8 | 2.7 |
| 10. I know at what stage I am in the application. | 3.5 | 3.6 | 3.6 |
| 11. The application does not make screen leaps pointless. | 4.0 | 3.8 | 3.9 |
| 12. The error messages are helpful and not confused | 4.0 | 3.8 | 3.9 |
| 13. The processing speed of the application is fast | 3.5 | 3.0 | 3.3 |
| **TOTAL** | 47.5 | 45.0 | 46.3 |

## 4 Discussion

In this study, a proof of concept test was carried out to ensure the feasibility of SARAEasy to easily register abnormalities of clinical phenotypes associated with the severity of cerebellar syndrome, such as truncal ataxia, appendicular ataxia or dysarthria. Cerebellar syndrome quantification and interpretation through phenotypic abnormalities can help to improve outcome measurements of evaluation of spinocerebellar ataxia. Mobile apps are portable devices facilitating the registration of measurements during the clinical examination with maximum flexibility and high cost-effectiveness. However, currently available apps in different domains face limitations on sharing measurement data and also information on the formulas they are using to calculate scores. SARAEasy can fill this gap by proposing an application providing full access to the item data while at the same time offering reporting capacities based on high-quality patient phenotype interpretation.

SARAEasy has been designed with European MEDDEV legislation in mind [21]. SARAEasy manipulates patient data as it calculates partial and total scores, and it generates a clinical report explaining the quantitative scores. Hence, the device should be considered a type of low risk. Even so, patient safety is guaranteed, as SARAEasy was based on a well-validated scale to assess the presence and severity of cerebellar ataxia and it has been modeled using a clinical archetype, which is now freely available in the OpenEHR CKM. Additionally, the clinical knowledge required to generate the final reports has been published and validated using data from SCA36. Thus, all data collected following the archetype can be shared between different healthcare systems. Regarding the clinical interpretations provided by SARAEasy in reporting, we used a combination of GDL (Guideline Description Language) and OWL (Web Ontology Language) to model the information-processing units [16]. In the current version, SARAEasy does not provide data in some format following the clinical archetype. This option is easily extensible and clearly feasible to reach interoperability with the current clinical information systems. With regard to the validation process, SARAEasy will be tested using more patient data that are affected by diverse cerebellar ataxias, and with other neurologists from different hospitals.

## 5 Conclusion

Our study shows that SARAEasy is able to automatically generate high-quality patient reports that explain the total score for the cerebellar syndrome quantification. This explanation is determined by phenotypic abnormalities as appendicular or midline ataxia. SARAEasy offers low-cost cerebellar syndrome quantification and interpretation for research and clinical purposes, and may help to improve evaluation.

## Funding

This work presented in this paper was supported by the National Institute of Health Carlos III [grant no. FIS2012-PI12/00373: OntoNeurophen], FEDER for national and European funding.

## Acknowledgment

The authors would like to thank Dr. Manuel Arias and Dr. Ángel Sesar for participating in the validation process to test the validity of SARAEasy.

## References

1. Martinez-Martin, P.: Composite rating scales. J Neurol Sci 289(1), 7–11 (2010).
2. Schmitz-Hübsch, T., Du Montcel, ST., Baliko, L., Berciano, J., Boesch, S., Depondt, C., et al.: Scale for the assessment and rating of ataxia Development of a new clinical scale. Neurol 66(11), 1717-1720 (2006).
3. Wicks, P., Chiauzzi, E.: 'Trust but verify'–five approaches to ensure safe medical apps. BMC medicine 13(1), p.205 (2015).
4. Huckvale, K., Adomaviciute, S., Prieto, J.T., Leow, M.K.S., Car, J.: Smartphone apps for calculating insulin dose: a systematic assessment. *BMC medicine*, *13*(1), p.106 (2015).
5. Saute, J.A.M., Donis, K.C., Serrano-Munuera, C., Genis, D., Ramirez, L.T., Mazzetti, P., Pérez, L.V., Latorre, P., Sequeiros, J., Matilla-Dueñas, A., Jardim, L.B.: Ataxia rating scales—psychometric profiles, natural history and their application in clinical trials. The Cerebellum 11(2), pp.488-504 (2012).
6. Beale, T., Heard, S.: openEHR - Release 1.0.2. 2016. http://www.openehr.org/programs/specification/releases/1.0.2, last accessed 2018/01/04.
7. Min, H., Ohira, R., Collins, M.A., Bondy, J., Avis, N.E., et al.: Sharing behavioral data through a grid infrastructure using data standards. Journal of the American Medical Informatics Association 21(4), p. 642-649 (2014).
8. Köhler, S., Vasilevsky, N.A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., Baynam, G., Bello, SM., Boerkoel, CF., Boycott, KM.: The human phenotype ontology in 2017. Nucleic acids research 45(D1), p. D865-D876 (2017).
9. Braun, M., Brandt, A.U., Schulz, S., and Boeker, M.: *Validating archetypes for the multiple sclerosis functional composite.* BMC medical informatics and decision making 14(1): p. 64 (2014).
10. openEHR Archetype Editor, http://www.openehr.org/downloads/archetypeeditor/home, last accessed 2017/12/04.
11. Clinical Knowledge Manager, http://openehr.org/ckm/, last accessed 2017/11/26.
12. SARA observation archetype, http://openehr.org/ckm/#showArchetype_1013.1.2661, last accessed 2017/11/26.
13. Taboada, M., Rodríguez, H., Martínez, D., Pardo, M., and Sobrido, M.J.: Automated semantic annotation of rare disease cases: a case study. Database **2014,** p. bau045 (2014).
14. Protégé, http://protege.stanford.edu/products.php#desktop-protege, last accessed 2017/12/02.
15. HermiT OWL Reasoner, http://www.hermit-reasoner.com/, last accessed 2018/01/02.

16. Maarouf, H., Taboada, M., Rodriguez, H., Arias, M., Sesar, Á., Sobrido, M.J.: An ontology-aware integration of clinical models, terminologies and guidelines: an exploratory study of the Scale for the Assessment and Rating of Ataxia (SARA). *BMC medical informatics and decision making*, *17*(1), p.159 (2017).
17. IBM SPSS Software, https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software, last accessed 2017/12/20.
18. Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin **70**(4), p. 213 (1968).
19. Cortes, A.F.: Manual de Técnicas para el Diseño Participativo de Interfaces de Usuario de Sistemas basados en Software y Hardware, http://www.disenomovil.mobi/multimedia_un/trabajo_final/03_cuestionarios_modelo_usabilidad_web.pdf, last accessed 2018/01/12.
20. Brooke, J.: SUS-A quick and dirty usability scale. Usability evaluation in industry 189(194), 4-7 (1996).
21. Whitepaper medical apps, https://www.nictiz.nl/publicaties/infographics/infographic-medical-apps-is-certification-required, last accessed 2018/01/15