UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CENTRO INTERNACIONAL DE ESTUDOS DE DOUTORAMENTO E AVANZADOS DA USC (CIEDUS)

TESE DE DOUTORAMENTO

Implementing electronic scales to support standardized phenotypic data collection - the case of the Scale for the Assessment and Rating of Ataxia (SARA)

HAITHAM MAAROUF

ESCOLA DE DOUTORAMENTO : Ciencias e Tecnoloxías da USC

PROGRAMA DE DOUTORAMENTO EN : Investigación en Tecnoloxías da Información

SANTIAGO DE COMPOSTELA

2018

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DO AUTOR DA TESE

Implementing electronic scales to support standardized phenotypic data collection - the case of the Scale for the Assessment and Rating of Ataxia (SARA)

D. **Haitham Maarouf**

Presento miña tese, seguindo o procedemento adecuado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) No seu caso, na tese se fai referencia as colaboracións que tivo este traballo.
3) A tese é a versión definitiva presentada para a súa defensa e coincide ca versión enviada en formato electrónico.
4) Confirmo que a tese non incorre en ningún tipo de plaxio de outros autores nin de traballos presentados por min para a obtención de outros títulos.

En Santiago De Compostela,18 de Decembro de 2017

Firmado. **Haitham Maarouf**

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR / TITOR DA TESE

Implementing electronic scales to support standardized phenotypic data collection - the case of the Scale for the Assessment and Rating of Ataxia (SARA)

Dna. **Prof. María Jesús Taboada Iglesias**

Dna. **Dr. María Jesús Sobrido Gomez**

INFORMAN:

Que a presente tese, correspóndese co traballo realizado por D. ***Haitham Maarouf****, baixo a nosa dirección, e autorizamos a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director desta non incorre nas causas de abstención establecidas na Lei 40/2015.*

En Santiago de Compostela,18 de Decembro de 2017

Asdo. **María Jesús Taboada Iglesias** Asdo. **María Jesús Sobrido Gomez**

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DEDICATION

This thesis is dedicated to the memories of my father and uncles: Prof. Adel Maarouf, Abdul Rahim and Mahmoud. I also dedicate it to my mother, wife, brothers, sisters and Nabil for their endless love, support and encouragement.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ACKNOWLEDGMENT

Writing an acknowledgment is the hardest part of writing a thesis. This work would not have been possible without the support and encouragement of several people. Please accept my sincere gratitude, irrespective of your name appearing in the acknowledgment.

First and foremost, I would like to sincerely thank my supervisor, Prof. María Jesús Taboada Iglesias for her great support, untiring help, patience, commitment of time, guidance and being an exceptional supervisor through my long Ph.D. process. There are no words that can describe how much I am grateful to her. Thank you, Prof. Taboada, for giving me that much needed morale boost, when things seemed to be looking gray, and there is no light at the end of the tunnel. I am also indebted to her for expanding my knowledge, enhancing my research skills and helping me to publish a paper.

I extend my thanks and gratitude to Dr. María Jesús Sobrido Gomez who provided me with the patient Data and all the required medical knowledge related to the Spinocerebellar Ataxia type 36 (SCA36) disease and the Scale for the Assessment and Rating of Ataxia (SARA). I also thank Dr. Manuel Arias and Dr. Ángel Sesar for participating in the validation process to test the validity of the developed tool.

This Ph.D. thesis was supported by the National Institute of Health Carlos III [grant no. FIS2012-PI12/00373: OntoNeurophen], FEDER for national and European funding; PEACE II, Erasmus Mundus Lot 2 Project [grant no. 2013-2443/001-001-EMA2]. Special thanks go to the European Commission and Erasmus Mundus program.

Furthermore, I express my deepest gratitude to Prof. Hatem Alamy, the Chairman of the Modern University for Business and Science (M.U.B.S), and to Dr. Bassem Kaissi for giving me the

opportunity to complete my doctoral study in Spain and have such an incredible international experience here.

On the other hand, I am grateful to Dr. Sami Ashour and each member of his family. Thank you for your invaluable support that encouraged me to overcome all the obstacles.

I would like to thank Dr. Nadim Anani, Dr. Rong Chen and Dr. Iago Corbal and for their great support with Guideline Definition Language (GDL). I also thank Dr. Ian McNicoll and Mrs. Hildegard Franke for their significant assistance with openEHR archetype modeling.

I would also like to thank all my colleagues from my lab at the Departamento de Electrónica e Computación and Knowledge Engineer Applied to Medicine (KEAM) at the Universidade de Santiago de Compostela, with whom I enjoyed so much and shared very special moments of my life.

My love, my soulmate, my wife, Noujoud Baroudi, thank you because you lived every moment of my Ph.D. process with me. Thank you for being there through all the obstacles, vicissitudes, success and failures. Her presence in my life and support have been my greatest strengths. I love you until the end.

I thank my entire family, Ayman, Rana, Samer, Afaf, Darine, Tarek, Nadine, Sherine and Nabil for encouraging me. Last but not least, I also thank my niece Sophia for making me happy with her innocent smile.

This acknowledgment cannot be complete without thanking my parents, for making me who I am. I wish my father would have lived to see the achievement of his son, but I know my mother is extremely proud of me. I am grateful to both of them, eternally.

RESUMO

No dominio clínico, as anomalías fenotípicas defínense como alteracións da morfoloxía, fisioloxía ou a conducta. En enfermidades xenéticas raras, a representación computacional das anomalías fenotípicas é crucial para mellorar a interpretación das probas xenéticas. A diferenza da tecnoloxía xenómica, a súa recollida e análise non se realiza seguindo un proceso estandarizado. As escalas clínicas representan un recurso importante para a recollida estandarizada de datos, especialmente en neuroloxía. Así mesmo, os arquetipos clínicos facilitan a estandarización computacional dos datos. Non obstante, as descricións fenotípicas xorden da interpretación clínica dos datos recollidos. Por iso, a súa representación estandarizada require recursos para explotar o razoamento nos arquetipos clínicos, o que supón un desafío hoxe en día.

O obxectivo principal desta tese de doutoramento foi facilitar a integración da semántica necesaria para interpretar automaticamente as coleccións de datos clínicos estandarizados. Para abordar o obxectivo, combinamos arquetipos clínicos, guías clínicas e ontoloxías para desenvolver un prototipo electrónico para a Escala de Avaliación e Valoración de Ataxia (SARA). Comezamos o proceso extraendo unha versión reducida da 'Human Phenotype Ontology' e utilizándoa como columna vertebral para normalizar o contido da SARA a través de arquetipos clínicos. O coñecemento necesario para explotar o razoamento modelouse como unidades de procesamento de información separadas e conectadas entre elas a través dos arquetipos definidos. Seguindo esta aproximación, implementamos un prototipo que se validou usando datos de 28 individuos anónimos afectados por a 'Ataxia da Costa da Morte' (SCA36). Os nosos resultados revelan un grao substancial de acordo entre o prototipo e os expertos humanos, confirmando que a combinación de arquetipos, ontoloxías e guías clínicas é unha boa solución para automatizar a extracción do coñecemento fenotípico relevante a partir dos datos de escalas de validación clínica.

PALABRAS CHAVE

Escalas clínicas, GDL, Ontoloxías, Arquetipos clínicos, SARA, Ataxia espinocerebelosa

RESUMEN

En el dominio clínico, las anomalías fenotípicas se definen como alteraciones de la morfología, la fisiología o la conducta. En enfermedades genéticas raras, la representación computacional de las anomalías fenotípicas es crucial para mejorar la interpretación de las pruebas genéticas. A diferencia de la tecnología genómica, su recopilación y análisis no se realiza siguiendo un proceso estandarizado. Las escalas clínicas representan un recurso importante para la adquisición estandarizada de los datos, especialmente en neurología. Asimismo, los arquetipos clínicos facilitan su estandarización computacional. Sin embargo, las descripciones fenotípicas proceden de la interpretación clínica de los datos recopilados. Por lo tanto, su representación computacional requiere herramientas que permitan explotar el razonamiento en arquetipos clínicos, lo cual es un reto hoy en día.

El objetivo principal de esta tesis doctoral ha sido facilitar la integración de la semántica requerida para interpretar automáticamente las colecciones de datos clínicos estandarizados. Para alcanzar el objetivo, combinamos arquetipos clínicos, guías clínicas y ontologías para desarrollar un prototipo electrónico para la escala de evaluación y clasificación de la ataxia (SARA). Comenzamos el proceso extrayendo una versión reducida de la 'Human Phenotype Ontology' y utilizándola como columna vertebral para normalizar el contenido de SARA a través de arquetipos clínicos. El conocimiento requerido para explotar el razonamiento se modeló como unidades separadas de procesamiento de información interconectadas a través de los arquetipos definidos. Siguiendo este enfoque, implementamos un prototipo que se validó utilizando datos de 28 sujetos anónimos afectados por la 'Ataxia da Costa da Morte' (SCA36). Nuestros resultados revelan un grado sustancial de acuerdo entre el prototipo y los expertos humanos, confirmando que la combinación de arquetipos, ontologías y guías clínicas es una buena solución para automatizar la extracción de

conocimiento fenotípico relevante a partir de las puntuaciones simples de escalas de valoración clínica en neurología.

PALABRAS CLAVE

Escalas de valoración clínica, GDL, Ontologías, Arquetipos clínico, SARA, Ataxia espinocerebelosa

SUMMARY

In the clinical domain, phenotypic abnormalities are defined as alterations in normal morphology, physiology, or behavior. In rare genetic diseases, computational representation of phenotypic abnormalities is crucial to improve the interpretation of the genetic tests. Unlike genomic technology, collecting and analyzing phenotype data is not usually conducted following a standardized process. Rating scales represent an important resource for standardized data collection, especially in neurology. Representing rating scales using clinical information archetypes promotes computational data standardization. However, phenotypic descriptions arise from clinical interpretation of the collected data. Hence, their computational representation requires facilities for exploiting reasoning on clinical archetypes, which is a challenge nowadays.

The main objective of this doctoral thesis was to facilitate the integration of the semantics required to automatically interpret collections of standardized clinical data. In order to address the objective, we combined the best performances from clinical archetypes, guidelines and ontologies for developing an electronic prototype for the Scale of the Assessment and Rating of Ataxia (SARA), broadly used in neurology. A scaled-down version of the Human Phenotype Ontology was automatically extracted and used as backbone to normalize the content of the SARA through clinical archetypes. The knowledge required to exploit reasoning on the SARA data was modeled as separate information-processing units interconnected via the defined archetypes. Based on this approach, we implemented a prototype named SARA Management System, to be used for both the assessment of cerebellar syndrome and the production of a clinical synopsis. For validation purposes, we used recorded SARA data from 28 anonymous subjects affected by SCA36. Our results reveal a substantial degree of agreement between the results achieved by the prototype and human experts, confirming that the combination of archetypes, ontologies and guidelines is a good solution to automate

the extraction of relevant phenotypic knowledge from plain scores of rating scales.

KEYWORDS

Rating scales, GDL, Human Phenotype Ontology, Clinical Archetypes, SARA, Spinocerebellar ataxia

RESUMEN AMPLIADO

A partir de la evidencia de que los pacientes con un mismo diagnóstico pueden presentar diferentes manifestaciones clínicas y, por tanto, reaccionar de forma distinta a la misma intervención, la *medicina personalizada* reconoce que cada paciente es único y por tanto debe ser tratado de forma individualizada. A partir de los noventa, con el impulso de la genómica y otras ciencias ómicas, se reconoce la importancia de *estratificar a los pacientes*, es decir, de clasificarlos en grupos similares biológicamente, con el objetivo de conseguir la respuesta óptima a las intervenciones planificadas en cada uno de los subgrupos. Adicionalmente, diversos estudios ya han demostrado que la identificación de estos subgrupos requiere analizar los datos ómicos junto con descripciones computacionales de calidad del fenotipo del paciente. En dominios clínicos, una anomalía fenotípica es una divergencia de la morfología, la fisiología o el comportamiento normal del paciente. Por tanto, el éxito en la estratificación de pacientes también dependerá, en gran medida, de los recursos computacionales disponibles para adquirir y representar el fenotipo de los pacientes, y para integrarlos adecuadamente con la información ómica y de imagen médica.

Las ontologías, como artefactos informáticos del campo de la Inteligencia Artificial, facilitan la organización y armonización de la información compleja y heterogénea, proporcionando facilidades de consulta e inferencia lógica sobre los datos almacenados. En los últimos años, una de las ontologías que ha experimentado el avance más importante en su uso para estudiar el diagnóstico clínico en enfermedades con base genética es la Human Phenotype Ontology (HPO). A la vez, diferentes consorcios internacionales han estado desarrollado modelos de datos que promueven la estandarización en la adquisición de los datos de pacientes, tales como ISO 13606, HL7 CDA, NINDS CDE e Intermountain Healthcare. El uso de dichos modelos es crucial para comparar resultados entre diferentes estudios,

integrar información entre diferentes aplicaciones, e implementar sistemas de ayuda a la decisión. Los esfuerzos de estos consorcios han dado lugar a especificaciones formales y computables del contenido clínico, que se conocen como *arquetipos clínicos*. Dichas especificaciones permiten representar, de forma consensuada, cualquier estructura de datos de la historia clínica del paciente, incluyendo tanto las definiciones (en forma de restricciones sobre las estructuras), como las interrelaciones entre dichas estructuras.

Mientras que los arquetipos clínicos estandarizan la captura de los datos clínicos del paciente, las ontologías de fenotipos estandarizan su significado e interpretación. Hay que tener en cuenta que las descripciones fenotípicas de los pacientes (que aparecen, por ejemplo, en los informes clínicos textuales) están en un nivel de abstracción más elevado que los datos de paciente recopilados a través de cuestionarios o pruebas clínicas, lo que provoca *impedance mismatch*. Una posible forma de solucionar el desfase entre la estandarización de los datos clínicos y la de fenotipos es utilizar las facilidades del razonamiento basado en ontologías sobre los datos recopilados con arquetipos. Sin embargo, a día de hoy, esta opción es todavía un reto. Aunque las especificaciones de los arquetipos clínicos proporcionan formas de expresar alineamientos (mappings) de los ítems del arquetipo a los conceptos de las ontologías, no existen recursos que faciliten el razonamiento basado en las ontologías alineadas. Hasta el momento, se han propuesto varias alternativas que abarcan la conversión de arquetipos al lenguaje de ontologías OWL-DL (Ontology Web Language-Description Language) o la definición de alineamientos intensivos en conocimiento desde las fuentes de datos a los arquetipos clínicos. Sin embargo, estas propuestas siguen sin proporcionar una tecnología sencilla que facilite el razonamiento. Por otra parte, el alineamiento de los datos basados en arquetipos clínicos con las ontologías no es una tarea trivial, y prueba de ello es que la mayoría de los arquetipos públicos no contienen dichos alineamientos. Siguiendo la aproximación estándar de desarrollo de arquetipos clínicos, el alineamiento ontológico se suele realizar en las últimas etapas de modelado. Ello conlleva un esfuerzo extra, en parte debido al gran tamaño de las ontologías. Además, el diseño de arquetipos clínicos de

forma separada de las ontologías puede conllevar discrepancias muy elevadas en el significado de los ítems clínicos.

Las escalas clínicas representan un recurso importante para la recopilación de datos estandarizados. Si bien las escalas clínicas se usan en todas las disciplinas médicas, son especialmente relevantes en especialidades que manejan variables fenotípicas complejas, como la neurología. Su uso incrementa la calidad de los datos, al reducir la subjetividad en las descripciones fenotípicas, y simplifica el diseño de los protocolos de recogida de datos en los estudios clínicos. Generalmente, las escalas clínicas valoran una o varias dimensiones clínicas mediante un conjunto de ítems y proporcionan una puntuación global. La hipótesis de partida de esta tesis doctoral es que reducir todo el contenido de la información recopilada a través de una escala de valoración a un único número (puntuación total) puede conllevar a la pérdida de información clínica relevante. El objetivo de esta tesis doctoral es demostrar que es posible realizar interpretaciones clínicas de forma automática sobre los datos recopilados por las escalas clínicas, de la misma manera que un experto clínico lo hace. Dichas interpretaciones automáticas pueden facilitar la evaluación médica, proporcionar ayuda para la escritura de informes de pacientes y la decisión médica. Para alcanzar el objetivo propuesto hemos desarrollado una aproximación novedosa orientada a modelar e implementar escalas clínicas electrónicas en el dominio de la neurología. Dicha aproximación busca representar computacionalmente tanto el contenido como la interpretación clínica de los datos recopilados. Para ello, se hace uso de los estándares de registros electrónicos de pacientes y de las tecnologías web semánticas. La principal innovación de nuestro trabajo ha sido el desarrollo de una aplicación que va más allá de una simple calculadora, con la incorporación del conocimiento clínico requerido para interpretar la información recopilada y generar automáticamente los correspondientes informes de pacientes. Los beneficios de nuestra solución innovadora son la provisión de estandarización clínica no sólo durante la recogida de los datos sino también durante la interpretación clínica de los hallazgos de pacientes, así como la producción de facilidades para la generación automática de informes, que liberan al médico de dicha tarea.

En este trabajo, optamos por abordar la Escala para la Evaluación y Clasificación de la Ataxia (SARA), un instrumento bien validado para evaluar la presencia y la gravedad de la ataxia cerebelosa. Esta escala tiene un uso muy extendido y ha sido aplicada por nuestro grupo para la evaluación de la ataxia espinocerebelosa tipo 36 (SCA 36). Para facilitar el alineamiento ontológico y evitar grandes discrepancias semánticas entre los arquetipos clínicos y las ontologías, se ha propuesto un método novedoso basado en la suposición de que el diseño de arquetipos debería ser soportado por ontologías. Por otro lado, la interpretación clínica de los datos recopilados por una escala de valoración requiere diferentes tipos de información para su automatización: datos para registrar (es decir, el contenido de la escala clínica), conocimiento sobre el significado de los términos en la escala (es decir, conocimiento terminológico), conocimiento de procedimientos para comprender el significado de los puntajes (que se pueden expresar fácilmente mediante guías clínicas) y conocimiento ontológico para deducir las anomalías fenotípicas de los pacientes. Elegimos utilizar una combinación de lenguajes de guías clínicas, arquetipos clínicos y ontologías para abordar los desafíos del modelado de la escala clínica. Las preguntas de investigación abordadas en este trabajo son: I) ¿La combinación de GDL (Guideline Definition Language), arquetipos clínicos y ontologías es adecuada para la descripción e interpretación de los datos colectados vía la escala SARA?, y II) ¿Es posible lograr la integración de estas herramientas computacionales para modelar e interpretar eficientemente la información clínica proporcionada por la escala clínica?

Nuestro enfoque de modelado se basa en cuatro pasos principales: I) Creación de una versión reducida del HPO, mediante la extracción de los módulos de ontología relevantes para la escala SARA, II) anotación de las descripciones de texto libre de la escala clínica con los módulos de ontología, III) desarrollo de dos tipos de arquetipos clínicos (observación y evaluación), y IV) definición de unidades de procesamiento de información para expresar el sistema de apoyo a la interpretación clínica. El modelado de la escala SARA involucró un nivel de datos - representación de los ítems de SARA - y un nivel de conocimiento - referido a la estrategia para calcular el puntaje total y la

interpretación del fenotipo. Los arquetipos se usaron para modelar el nivel de datos, mientras que GDL y OWL (Ontology Web Language) se usaron para modelar el nivel de conocimiento. Esta representación tenía restricciones, ya que los modelos openEHR son compatibles con GDL, pero no dan mucho soporte para OWL y el razonamiento relacionado. Para cerrar el gap entre los arquetipos clínicos y la ontología, se definieron alineamientos que facilitaron la traducción de las instancias de arquetipo al conjunto de datos OWL.

Para la extracción del módulo de la ontología relevante a la escala, comenzamos revisando y recopilando documentos de texto que describían la escala. Luego anotamos las fuentes extraídas con los términos de la ontología HPO, utilizando el OBO Annotator, un sistema de anotación de conceptos fenotípicos desarrollado en el grupo. A continuación, mapeamos los ítems de la escala con las anomalías fenotípicas estándar del paciente descritos en su propia escala. Para el diseño del arquetipo, hicimos una diferencia entre los datos recopilados por la escala y que se calculan directamente utilizando la información proporcionada por la escala (ítems, dimensiones clínicas, puntajes individuales y puntaje total), y los datos que no están directamente disponibles en el escala. Teniendo en cuenta esta distinción, decidimos utilizar dos tipos de arquetipos clínicos: *Observación* para normalizar los datos proporcionados por la escala, y *Evaluación*, para estandarizar las interpretaciones sobre los datos descritos en la escala. El primero permite representar el contenido de la escala, mientras que el segundo registra las interpretaciones clínicas derivadas de la escala. Los dos tipos de arquetipos clínicos se desarrollaron utilizando el editor de arquetipos proporcionado por OpenEHR. Para modelar el arquetipo de observación, después de agregar los metadatos, estructuramos el contenido de la escala de acuerdo con este arquetipo, es decir, mediante una estructura de árbol con elementos para las tres dimensiones clínicas (el conjunto de elementos, la puntuación total y la fecha de las observaciones) y los valores para las puntuaciones individuales. Luego, definimos los elementos con tipos de datos adecuados, descripciones, comentarios, detalles, ocurrencias, restricciones y valores posibles. Asignamos cada elemento a la clase de ontología correspondiente de la versión reducida del HPO, utilizando las anotaciones obtenidas. Finalmente, estructuramos y organizamos los ítems de acuerdo con la

ontología. El arquetipo de evaluación se modeló para registrar tres tipos de interpretaciones: I) Interpretaciones deducidas directamente de elementos individuales o de la puntuación global, II) Interpretaciones deducidas directamente de valores específicos para un elemento individual, y III) Interpretaciones sobre las anomalías del paciente asociadas con las dimensiones clínicas de la escala. Con respecto a las unidades de procesamiento de información (o inferencias) que interpretan automáticamente los datos de la escala se distinguieron:

- Las funciones de cálculo, que expresan las operaciones matemáticas para calcular las puntuaciones de los ítems individuales y la puntuación global de la escala.

- La valoración del grado de deterioro que gradúa la intensidad de los elementos expresados en el arquetipo de observación, que generalmente está dado por niveles de corte.

- La evaluación del síndrome cerebeloso que está basada en la puntuación global de la escala, y se calcula aplicando varias reglas heurísticas propuestas por el experto en neurología.

- La generación de la sinopsis clínica que describe las características fenotípicas (signos) que acompañan al síndrome cerebeloso.

Para la validación, utilizamos registros de datos de 28 sujetos anónimos con Ataxia Espinocerebelosa Tipo 36 (SCA36). Todos los pacientes fueron examinados siguiendo la SARA. La evaluación se llevó a cabo en tres pasos:

I) Cálculo e interpretación de la puntuación total: completamos los datos de cada paciente para obtener la interpretación. El sistema dedujo automáticamente 1) la intensidad de cada ítem en la escala, 2) la gravedad del síndrome cerebeloso, 3) la gravedad de la ataxia troncal y

4) la gravedad de la ataxia apendicular en los lados derecho e izquierdo.

II) Interpretación por dos neurólogos independientes: se envió el mismo conjunto de datos a dos neurólogos, pero se agregaron las puntuaciones totales calculadas para cada paciente. Los neurólogos utilizaron su experiencia en ataxia para determinar la gravedad del síndrome cerebeloso a partir de la puntuación total y la gravedad de la ataxia troncal y apendicular, si está presente, de las puntuaciones individuales.

III) Comparación de resultados entre el sistema y los expertos humanos: para validar el sistema, se realizaron los siguientes pasos: 1) creamos una hoja usando SPSS, 2) importamos las interpretaciones del sistema y de los neurólogos a SPSS, y 3) realizamos la prueba Weighted Kappa 12 veces para medir la fuerza del acuerdo entre el sistema y cada neurólogo, y entre los dos neurólogos.

Para demostrar la funcionalidad de nuestro enfoque, desarrollamos un prototipo llamado "SMS" (Sistema de gestión SARA), que permite la gestión de los datos del paciente del síndrome cerebeloso. Usamos JAVA como lenguaje de programación y MySQL como un sistema de administración de bases de datos para almacenar toda la información necesaria. La arquitectura de "SMS" está estructurada en tres capas: "Persistencia", "Operación" e "Interfaz". La capa de persistencia utiliza un sistema de gestión de base de datos para almacenar toda la información requerida (los arquetipos clínicos, la ontología, los datos del paciente recogidos usando arquetipos clínicos e inferidos por el sistema, e información adicional). Para poblar la base de datos con la información del arquetipo, se usó el analizador ADL (Archetype Description Language) para generar un árbol de dependencias de términos y luego crear un archivo XML con las instancias de los arquetipos. La capa de operación incluyó todas las unidades de procesamiento de información que son responsables de ejecutar las funciones de cálculo, la evaluación del grado de deterioro, la evaluación del síndrome cerebeloso y la sinopsis clínica. La versión actual de GDL no proporciona una API de Java ni ningún mecanismo para manipular las reglas definidas con el editor. Por lo tanto, la única solución era volver a escribir las reglas en un motor de reglas (como Drools o Clips) o implementar directamente en Java (decidimos esta

segunda solución). Sin embargo, la inferencia de fenotipos se ha podido implementar razonablemente utilizando la API de OWL y los alineamientos entre los ítems de los arquetipos y la ontología OWL. Los alineamientos son útiles para crear automáticamente individuos OWL de la clase definida a partir de las instancias de arquetipo almacenadas en la base de datos.

La capa de interfaz consiste en varios formularios de entrada y salida. La forma principal de la herramienta contiene dos menús: el menú de ontología y el menú SARA. El primer menú permite a los usuarios modificar la estructura de los módulos de ontología HPO. Se pueden agregar, actualizar y eliminar términos en los módulos de ontología. El formulario proporciona la posibilidad de generar automáticamente un gráfico jerárquico que muestra todas las clases disponibles y sus relaciones jerárquicas. El menú SARA consta de dos submenús: 1) Observación, donde los neurólogos pueden seleccionar pacientes e ingresar los valores de los elementos evaluados definidos en la escala SARA y modelados en el arquetipo de observación, y 2) Evaluación, que proporciona tres características principales:

- Una tabla que muestra todos los códigos de anomalías fenotípicas que tiene un paciente. El conjunto de estas anomalías se deduce automáticamente.
- Un informe textual que resume el estado de un paciente según la recopilación de datos. El informe describe la gravedad del síndrome cerebeloso y una breve sinopsis de las anomalías fenotípicas.
- Un gráfico de evaluación que visualiza todas las anormalidades del paciente

Para facilitar el intercambio semántico interoperable de datos, se pueden generar tanto los datos XML recopilados por SARA (usando los esquemas que cumplen con el arquetipo de observación) como los datos inferidos por la aplicación (usando los esquemas que cumplen con el arquetipo de evaluación). La validación del sistema se basó en los valores de Kappa ponderado obtenidos de las 12 pruebas realizadas. Se obtuvieron unos valores Kappa en el rango entre 0.65 y 0.93.

En la actualidad, los trastornos atáxicos todavía no tienen una terapia farmacológica exitosa, y los pacientes sufren la inevitable progresión de la enfermedad degenerativa. El objetivo de las escalas clínicas es facilitar la comprensión de la historia natural de los trastornos atáxicos y evaluar adecuadamente la eficacia de los fármacos en los ensayos clínicos. En la presente tesis doctoral, nos centramos en proporcionar soporte automático para la interpretación clínica de los datos recopilados usando la escala SARA. El trabajo contribuye a una mejor comprensión de cómo los arquetipos clínicos, las guías clínicas y las ontologías se pueden combinar para modelar e implementar una escala de valoración en el dominio de la neurología. Hay varias contribuciones en esta investigación. Una contribución es un enfoque basado en ontologías para modelar los dos arquetipos clínicos propuestos, lo que reduce el esfuerzo necesario para crear alineamientos y evita grandes discrepancias semánticas entre los arquetipos modelados y los módulos de ontología. Otra contribución es la separación clara y explícita entre los componentes estándar de la escala relacionados con el contenido (es decir, ítems, dimensiones clínicas y puntajes), que han sido modelados usando un arquetipo de observación, y las interpretaciones clínicas de estos componentes, que han sido normalizadas por un arquetipo de evaluación para uso local. Finalmente, una contribución clave es la identificación clara de todos los diferentes tipos de conocimiento requeridos para interpretar los datos recopilados por la escala y su modelado como unidades de procesamiento de información que se comunican entre sí a través de los dos arquetipos definidos, proporcionando un mecanismo simple de combinación ontología y razonamiento basado en reglas.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CONTENTS

1 INTRODUCTION

Nowadays, the scientific community in the field of clinical genetics and genomics is paying increasing attention to phenotype information. In the clinical domain, phenotypic abnormalities are defined as alterations in normal morphology (structural abnormalities such as *cerebellar atrophy*), physiology (functional abnormalities such as *incoordination of movement*), or behavior (such as *difficulty in social interactions*) [1]. Acquiring a better understanding of the full variety of phenotypic abnormalities associated with rare genetic diseases is crucial to improve the interpretation the genetic tests, and the translation of genomic information into clinical practice [2]. Unlike genomic pipelines, the collection and analysis of phenotype data in the routine clinical setting is not usually conducted following a standardized process. In clinical research, the evaluation of the patients' phenotypic features ranges from determining the set of data to be gathered to deciding the most appropriate computational representation. In general, this is a difficult, laborious and time-consuming task [3]. Phenotype annotation has a huge potential to automatically extract data from large amounts of existing patient records or controlled trials. Recently, substantial progress has been achieved in encoding phenotypes using the Human Phenotype Ontology (HPO) [4]. This ontology supplies a standardized core of human phenotypic abnormalities and the relationships between them. It is accessible online and contains over 12,000 classes and 16,000 hierarchical relationships [5].

Electronic rating scales represent an important resource for standardized data collection, often providing primary and secondary outcome measures. While rating scales are used in all medical disciplines, they are especially relevant in specialties with a richness of complex phenotypic variables, such as neurology [6]. Clinical scales can measure the so-called latent variables, i.e., those that cannot be directly measured but can only be assessed indirectly through their

manifestations. Examples of latent variables (or clinical dimensions) in neurological diseases include the quality and intensity of a tremor, the degree of gait imbalance or cognitive performance. These latent variables are assessed through a set of clinical questions (named statements or items) [7]. Each statement may have multiple ordered response options, for which an ordinal number (score) is assigned. The total score for the global clinical dimension is usually obtained adding up all individual scores for each statement. Well-known examples are the Mini-Mental State Examination (MMSE) [8], a 30-point survey used to measure cognitive impairment, or the Glasgow Coma Scale (GCS) [9], which is used to assess coma and impairment of consciousness. Using these instruments entails many advantages: improved data quality by reducing subjectivity during measurement, simplified design of data collection, and data harmonization across different clinical studies. Hence, computational implementation of rating scales offers a major chance for data quality improvement and harmonization across different clinical studies.

Additionally, electronic rating scales are an important resource to support automated inference of patient phenotype from the data collection. Usually, rating scales grade several clinical dimensions, each of them assessed by different items. For instance, in addition to the movement disorder (i.e., disease state), the Unified Parkinson's Disease Rating Scale (UPDRS) [10] assesses other clinical sub-dimensions (such as mental state, complications of treatment and activities of daily life) via 42 questions providing a total score that grades the progression of the disease. However, reducing all content of a rating scale to a unique number (score) inevitably causes the loss of some phenotype information implicitly collected by the scale. For example, in patients with the same total UPDRS score this number could be due to different clinical dimensions, therefore actually be quite different clinically. A more precise inference of the patient phenotype from the sub-scores would facilitate the automated codification of the clinical abnormalities for further analysis. Additionally, it would decrease subjectivity during the score interpretation and facilitate medical evaluation, report writing, and clinical decision-making. In this work, we chose to address the Scale

for the Assessment and Rating of Ataxia (SARA)[1] [11], a well-validated instrument to evaluate the presence and severity of cerebellar ataxia [12]. This scale is broadly used and it has been applied by our group in a research on the spinocerebellar ataxia type 36 (SCA 36).

Formal description of rating scales using systematic clinical information models promotes computational data standardization comparison of results across studies [13], integration of information from different sources and medical records, and implementation of decision support systems. Different international projects and consortia have been developing standardized data models for clinical research and electronic health records, such as ISO 13606 [14], HL7 CDA [15], openEHR [16], NINDS CDE [3] and Intermountain Healthcare [17]. The commonality among these approaches is that they are focused on computable and formal specifications of clinical content in the form of information models known as clinical models or archetypes. These clinical models supply standardized data structures to represent the clinical statements included into rating scales. Additionally, mechanisms to link the clinical statements to classes of some standard terminology or ontology are provided. Hence, representing rating scales using clinical archetypes/models and terminologies/ontologies aims to get both clinical and computational harmonization of data collections.

While both clinical archetypes and ontologies seek to structure the patient information, according to the needs of clinical research, however their perspectives are often dissimilar. In archetypes, the clinical statements that must be entered at the same time are aggregated together. Archetypes model the information to mirror patient records. For example, the items *paraparesis* and *facial palsy* were recorded together into the archetype *Stroke Scale Neurological Assessment* [18], which is available on the Clinical Knowledge Manager (CKM) [19] provided by the OpenEHR Foundation. Ontologies, on the other side, aim at representing the meaning of those clinical statements. Classes in the Human Phenotype Ontology [20] are arranged in a hierarchical structure of phenotypic abnormalities. For example, both *paraparesis* and *facial palsy* are represented as abnormalities of the nervous system

[1] http://www.ataxia-study-group.net/html/about/ataxiascales/sara/SARA.pdf

in the HPO. However, the former is represented as an abnormality of the physiology, whereas the second one as an abnormality of the morphology. This ontological distinction cannot be reflected into the clinical archetype and however it is valuable to interpret the patient status. Thus, integrating ontologies with clinical archetypes would not only provide a static knowledge store, but also a dynamic resource to automatically infer patient phenotype and standardize data collection.

Specifically, Braun et al. [21] developed four clinical archetypes (Timed 25-Foot Walk, Nine Hole Peg Test, Paced Auditory Serial Addition Test, and MSFC Score) [22-25] to represent a rating scale for the assessment of Multiple Sclerosis (MS) patients consisting of three neurological tests. They applied a standard archetype development approach, consisting of: analyzing the clinical domain and requirements, identifying the archetype contents and their organization from different sources (literature, record forms, etc.), selecting the archetype type, structuring the content according to the archetype type, and filling the parts of the archetype with the content. With this approach, terminology/ontology mapping is carried out during the later steps of archetype building, when the model is almost complete. At this stage, the effort required to create the mappings between archetype terms and ontology entities is substantial [26], due in part to the large size of the ontologies [27]. Furthermore, designing clinical archetypes separately from ontologies may lead to major discrepancies in the meaning of clinical statements. As a result, ontology mappings are not common in the openly accessible archetypes of the repositories. Nevertheless, mapping clinical archetypes and ontologies/terminologies is key to get semantic interoperability among different data sources. With the aim of facilitating the ontology mapping and preventing large semantic discrepancies between clinical archetypes and ontologies, we propose to reorganize the classical methodology. At the heart of our methodology is the assumption that the archetype design should be supported by ontologies in those clinical situations where it is expected that archetype contents be logically organized.

Furthermore, automated assessment and evaluation upon the information represented by clinical archetypes is still an open research issue. Clinical archetypes/models aim to record standardized

definitions of the clinical data in electronic medical records [28]. In the case of rating scales, this structure matches to its *content* (i.e., clinical dimensions, items and scores). Additionally, clinical archetypes/models offer the possibility to normalize clinical data by mapping them to formal ontologies. However, exploiting reasoning on this clinical knowledge is still limited and is another challenge. The Guideline Definition Language (GDL) [29] is a formal language recently authorized by openEHR for expressing decision support logic by a rule-based declarative strategy. Anani et al. [30] used GDL to implement knowledge on contraindications for using thrombolytic treatment in patients suffering acute stroke, and Lin et al. [31] to implement ten electronic clinical practice guidelines in the chronic kidney disease. As Anani et al. [30] have emphasized, GDL provides a rule authoring language aimed to represent declarative knowledge that can be shareable and standardized, as it is supported by OpenEHR clinical models. However, GDL does not yet provide much support for ontologies and related reasoning. Another alternative is to transform clinical archetypes/models into OWL-DL (Ontology Web Language – Description Language) [32-34]. Following this approach, ontology reasoners, such as Pellet[2], Hermit [3] or Fact++[4], can be used to both check the OWL-based archetype consistency [35], and use the ontology to draw any inferences on data collection. Additionally, having archetypes, ontologies and knowledge descriptive (inference rules) under the same syntactic structure provides support for interoperability of rule-based mechanisms [28, 33, 36]. However, having two separate, independent versions of the same standard model, one of them in the language of the model itself (ADL-Archetype Definition Language) [37] and the second one in OWL format, makes maintenance more difficult. Furthermore, procedural knowledge as the sum (counting or any complex mathematical calculation) of the scores in rating scales cannot be simply represented in OWL. An interesting alternative proposed by Mugzach et al. [36] to perform a particular counting (named *k-of-N* counting by the authors) in OWL was to develop a plug-in meeting the specific requirements. However, different calculation functions would then require implementing specific plug-ins. Other

[2] https://www.w3.org/2001/sw/wiki/Pellet
[3] http://www.hermit-reasoner.com/
[4] http://owl.cs.manchester.ac.uk/tools/fact/

researchers defined knowledge-intensive mappings from the data sources to openEHR archetypes [38, 39]. They distinguished between data-level and knowledge-level processing tasks. The former included calculation functions specified in the mappings and directly run on archetype data. The latter covered classification tasks defined using OWL classes with sufficient conditions. An integrated Personal Health Record is an alternative option proposed to simplify data integration and clinical decision-making [40].

1.1 RESEARCH OBJECTIVE

The main goal of our work was to develop an electronic rating scale in the clinical domain of the neurology representing both the content and the interpretation of the SARA, using the Electronic Health Record (EHR) standards and taking advantage of semantic web technologies to automatically interpret the phenotype from data collected by a rating scale.

To that goal, specific objectives are:

- To computationally represent the knowledge covered by the rating scales.
- To define the phenotypes in a computationally accessible way.
- To map the patients' clinical data gathered from the rating scales to phenotypes.
- To efficiently and computationally model the clinical information content provided by the rating scales using EHR standards.
- To computationally define the interpretation of the rating scales using EHR standards.

1.2 RESEARCH QUESTIONS

1) Is the combination of GDL, openEHR clinical archetypes and ontologies suitable for the description of all knowledge and content covered by the SARA?

2) Is it possible to achieve integration of these computational tools to efficiently model the clinical information provided by the rating scale?

1.3 RESEARCH CONTRIBUTION

The contributions of this thesis are:

- In order to facilitate ontology mapping and prevent large semantic discrepancies between clinical archetypes and ontologies. The classical methodology proposed by Braun et al. [21] was enhanced and we developed a novel method based on the assumption that archetype design should be supported by ontologies in those clinical situations where the archetype contents are logically organized.
- The current HPO does not cover all the details of needed neurodegenerative phenotypes. New ontology modules relevant to the SARA were developed using OWL.
- We demonstrated how the openEHR clinical archetypes Observation and Evaluation could be used to model the content of a rating scale and to record the clinical interpretations and phenotypic abnormalities.
- GDL and OWL were effectively used to express all the required knowledge to understand the meaning and the scores of terms in the scale, and to deduce patient phenotypes.
- We chose to use a combination of GDL, openEHR clinical archetypes and ontologies to address the challenges of modeling the rating scale.

1.4 STRUCTURE OF THE THESIS

The rest of the thesis is structured as follows:

- Chapter 2 describes the background, providing an overview of the main components of rating scales, as well as the models, languages and ontologies used in this work.

- Chapter 3 describes the methodology proposed in this work to model electronic rating scales, with the specific example of an application to the SARA.
- Chapter 4 presents the results of implementing and validating our method.
- Chapter 5 highlights the implications of this work.
- Finally, conclusions, limitations and future work are provided in Chapter 6.

2 BACKGROUND

In this chapter, we present an overview about the components, tools and technologies used to develop this research work. The first part of this review is focused on the clinical domain, including information on rating scales and their components, the specific scale for the assessment and rating of ataxia (SARA) and the rare syndrome named spinocerebellar ataxia type 36 (also known as *Ataxia da Costa da Morte*). It also covers the technologies used in this thesis: the clinical data models or archetypes, OWL ontologies and the Human Phenotype Ontology, as well as the available openEHR formal language to implement computerized clinical decision support system.

2.1 CLINICAL RATING OR ASSESSMENT SCALES

Most rating scales used in neurology are ordinal scales. They provide a set of items needed to quantify the severity of motor, sensitive, sensory, cognitive function or quality of life, whereby the rater has to assign a value, usually numeric, to the graded items. Thus, rating scales rank patients in degrees of disability according to certain external criteria. Some of them assay only one clinical attribute or item (single-item scale), such as the Modified Rankin Scale (MRS) [41]; whereas others consist of several items (multiple-item scale). In some cases, all items in the scale assess the same dimension (e.g., motor deficit), whereas in other cases, the scale consists of several multiple-item sub-scales, like UPDRS [10]. Usually, rating scales combine values of individual patient traits (items scores) into a total score, which measures the variable computed by the set of items. For example, the MMSE is a questionnaire with a total score of 30 that collects items related to different traits: orientation to time (5-score) and to place (5-score), attention and calculation (5-score), language (2-score), etc. In the MMSE [8], the total score is derived by summation of all items, while in other cases the scoring may involve complex calculations. Different components can be distinguished in a rating scale (Fig. 2.1).

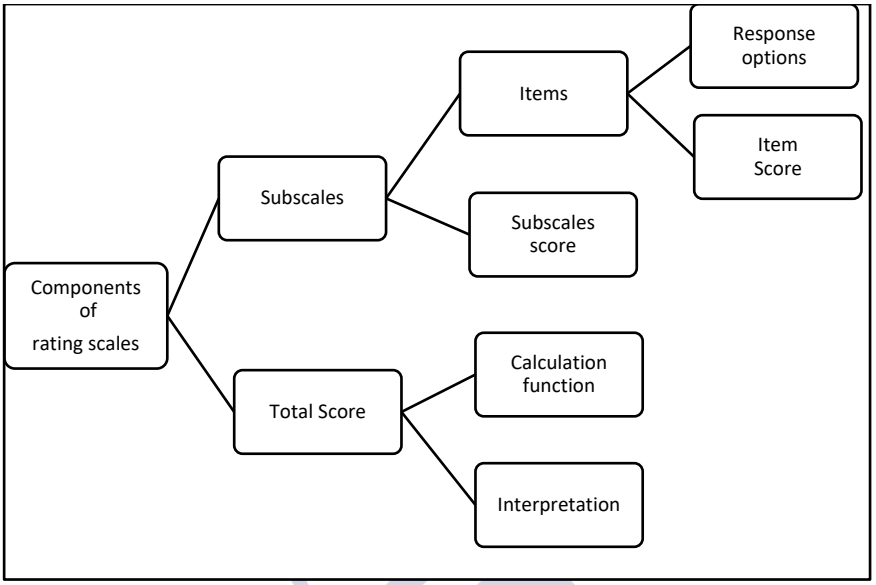

Fig. 2.1 Components of rating scales

2.1.1 Components of Rating Scales

The following components can be distinguished in clinical rating scales: items of the scale, response options and item scores, subscales, total score, calculation function, and interpretation of the score.

2.1.1.1 Items

Items of the scale are the different questions assessing a specific clinical dimension into the rating scale. For example, the UPDRS a scale that assesses disability due to Parkinson's disease, contains 42 different questions grouped into four clinical dimensions: I) Cognitive, behavioral and mood (4 items), II) Activities of daily living (13 items), III) Motor performance (14 items), and IV) Complications of treatment (11 items).

2.1.1.2 Response options and scores

Response options and scores are the possible values that raters can assign to items to quantify them. These values could belong to either

ordinal level, interval or ratio scales. For instance, the MRS ranges from 0 (No symptoms) to 6 (Death) to quantify the disability.

2.1.1.3 Subscales

Subscales are the clusters of items, which together measure a particular clinical dimension. For example, in the MMSE, the question 'What is the date?', including questions on year, season, month, date and day of the week, let the rater assess the clinical dimension 'orientation to time'.

2.1.1.4 Total Score

Total score is the value assessing the global clinical dimension measured by the rating scale. It is calculated once the different items have been evaluated.

2.1.1.5 Calculation Function

Calculation function is the set of mathematical operations performed on the item scores to calculate the total score. The sum of the item scores is the most usual approach for calculating the total score. However, alternative procedures, such as mean score or standardization to a reference population, are also frequent.

2.1.1.6 Interpretation of the total score

Interpretation of the total score is the explanation of the measurement result, which is usually left open to the rater. In some cases, a simple standard procedure is attached to the rating scale. For example, in the MMSE, there are four criteria to qualify the degree of impairment: '25-30 = questionably significant', '20-25 = mild', '10-20 = moderate', and '0-10 = Severe'.

2.2 CLINICAL ARCHETYPES

OpenEHR developed a two-level approach to make a separation between the semantic of information and knowledge into two levels:

the openEHR Information Model or openEHR Reference Model (RM) and Archetype Model (AM). The RM constitutes the base information model for openEHR system. It supports data types and data structures. The invariant semantics of the EHR are defined in the Information Model. On the other hand, the Archetype Model is the design specification for the archetypes. [42].

Clinical archetypes, which are themselves instances of an archetype model, enable clinical statements to be recorded as nested hierarchies of domain concepts [43]. For instance, the clinical statement of the physical condition of a newborn infant may need gathering clinical dimensions as skin color, heart rate, respiratory effort, etc. Sometimes, two or more archetypes are required to completely gather a clinical dimension [44]. For example, recording a clinical statement about diabetes may involve data fragments from three archetypes: blood pressure, glucose, and drug medication. Archetypes specify specifications of a data structure including optionality and multiplicity, relevant mappings to natural language and terminology systems, and data value constraints. They support interoperability and can be re-used across many types of healthcare. Archetypes facilitate the involvement of medical domain experts and computer scientists in the collaboration of standardized clinical content specifications and design for electronic health records.

One important feature of archetypes is their ability to translate the clinical data, to more than one natural language. For example, the clinical data in the Barthel index archetype [45] has been translated to Dutch besides the English language [44]. The intended aim of clinical archetypes is to empower clinicians to define contents, semantics, and user interfaces of systems independently from the Information model.

2.2.1 Archetypes Categories

There are four main categories of archetypes: Composition, Section, Entry, and Cluster. Each of them is used for different parts of the clinical recording and workflow processes. Each has particular attributes that support the capturing and re-use of clinical information [43].

2.2.1.1 Composition Class

All information recorded within the electronic health record will be contained within a Composition. A Composition class represents a container class in the openEHR reference model. Compositions are similar to clinical documents or events. Examples of Compositions are: Discharge Summary and Prescription.

2.2.1.2 Section Class

Sections are intended to organize the content. They are contained within Compositions and do not carry any semantic meaning. The headings that we may find on a blank piece of paper can be considered as sections. Most of the detailed clinical contents are available inside Entry and Cluster classes which themselves are contained in Sections. Examples of Sections are: Physical Examination and Vital Signs.

2.2.1.3 Entry Class

An Entry is a single 'semantic unit' of clinical information. It is used to usefully group information that can be re-used in many different settings with the same meaning. It has four concrete subtypes: observations, evaluations, instructions and actions. The selection decision among these subtypes is based clinical problem-solving process as shown in Fig. 2.2.

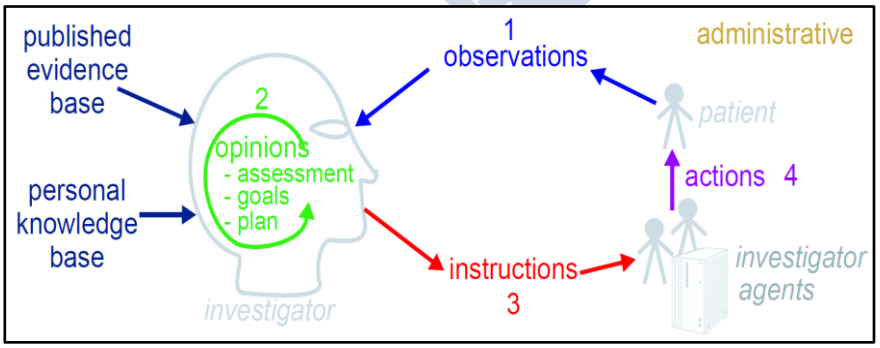

Fig. 2.2 Relationship of information types to the investigation process (Taken from [46])

2.2.1.3.1 Observation

The intended purpose of *Observations* is to capture the raw or uninterpreted information. Observations include measurement results, examination findings, symptoms and test results that can be reported by a patient. Examples of *Observations* are: Weight, Height, Electrocardiography (ECG), and Laboratory result. All *Observation* classes consist of four parts:

- The Data part has the core information (e.g. in the case of rating scales, the total score or Best Eye Response of the GCS).
- The State part contains information that is required for safe clinical interpretation of the core information (e.g. deafness and blindness can be considered as reasons for inability to record the Best Eye Response assessment of the GCS).
- The Protocol part holds information on how the information is collected or measured (e.g. many attempts and reasons for more than one attempt).
- The History part records information about the timing of the observation and the information width.

2.2.1.3.2 Evaluation

Evaluations are used to record clinical findings and to interpret information collected in Observations. They are *meta-observations* – opinions, assessment, goals and plan which arise within the clinician's mind. *Evaluations* consist of three parts: the Data part, State and protocol, and they have no history part (2.2.1.3.1 above). Examples of Evaluations are: Risk Assessment, Problem/Diagnosis, and Adverse reaction.

2.2.1.3.3 *Instructions*

Instructions contain statements that specify the Actions that should be performed in the future. *Instructions* may include different kind of interventions such as clinical orders for care.

2.2.1.3.4 Actions

Actions contain statements that describe what was actually done. They are used to record clinical activities like administering the clinical orders in the above *Instructions* (2.2.1.3.3). *Actions* complement *Instructions* as they can record the subsequent state of the Instructions, such as 'scheduled', 'completed', or 'cancelled'.

2.2.1.4 Cluster Class

Clusters represent reusable archetypes that are used within any *Entry* or other *Cluster*. For example, consider an *observation* archetype to model a 'Medical History' containing a *Symptom Cluster* to capture data about a presenting complaint of headache. This cluster can, in turn, contain other symptom clusters to capture headache-associated symptom details (such as vomiting or photophobia).

2.2.2 Archetype Definition Language

The specification of the archetypes is expressed in the Archetype Definition Language (ADL), which supplies the syntax for constraints on any domain entity. Fig. 2.3 displays an excerpt of the ADL file related to the openEHR archetype GCS [47]. An ADL archetype has four main sections: header, description, definition and ontology. The header and description sections introduce the archetype and contain metadata such as purpose, use or keywords for searching. The structures and constraints associated with clinical concepts are expressed in the definition section. For example, the GCS scale collects three items, which are nested into a hierarchy (*ITEM_TREE*). Each item, such as 'Best eye response', is modeled using an *ELEMENT*, which can be graded using four different values: '1-None', '2-To pressure', '3-To sound' and '4-Spontaneous'. The ontology section provides descriptions for each term defined in the definition section. For example, the description of the 'best eye response' will be available in the ontology section. It also allows a single clinical data node or fragment available in the archetype to be bound to more than one external terminology. This property is known as term mapping.

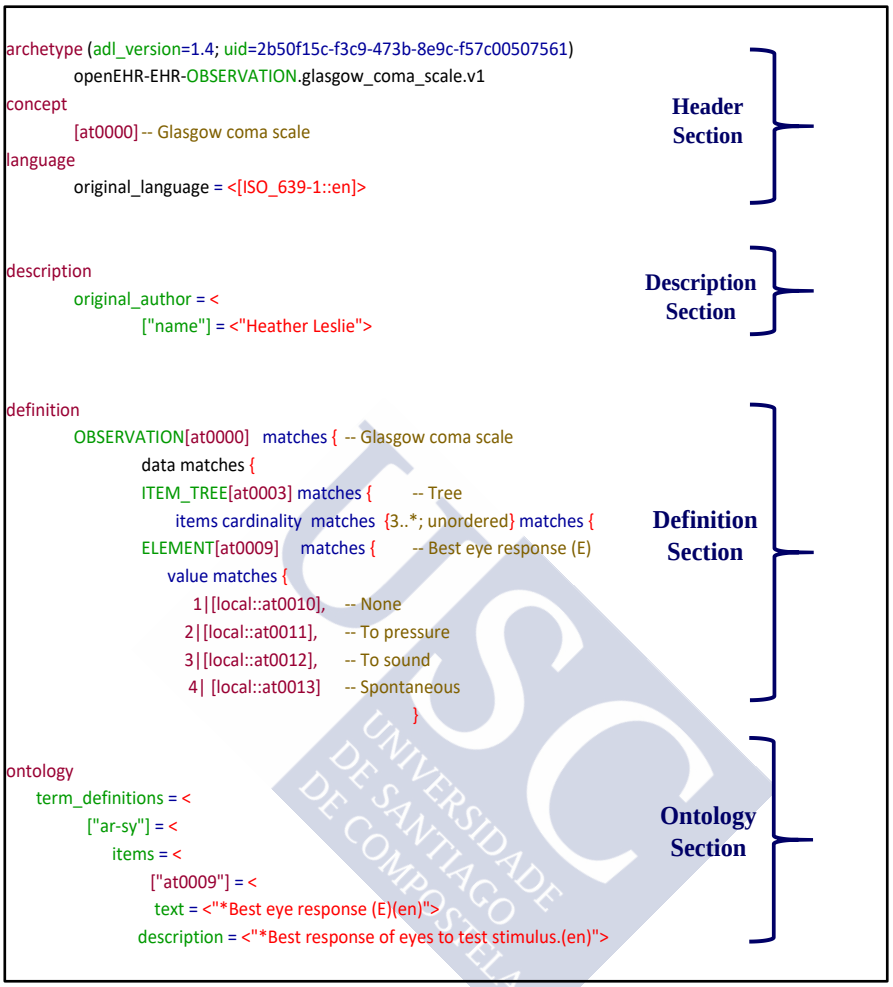

Fig. 2.3. Extract of the archetype GCS. It displays the structure of an ADL archetype and the main four parts: Header, Description, Definition and Ontology

There are some specific tools and repositories to deal with archetypes, such as the Archetype Editor[5] and the Clinical Knowledge Manager (CKM). In [48], a detailed list of openEHR tools, frameworks and platforms is provided.

5 http://www.openehr.org/downloads/archetypeeditor/home

2.3 CURRENT MODELING OF RATING SCALES AS CLINICAL ARCHETYPES

Overall, clinical archetypes in repositories, such as Clinical Knowledge Manager, model assessment scales as a set of clinical concepts required to record each scale item, plus the set of restrictions related to the different scores the items can take. Additionally, an extra element, often called *Total Score*, is added to model the overall score. Typically, this element (total score) includes a brief description of the calculation procedure. For example, Fig. 2.4 displays the description to calculate the total score element of the Glasgow Coma Scale and how this element is modeled in the GCS.

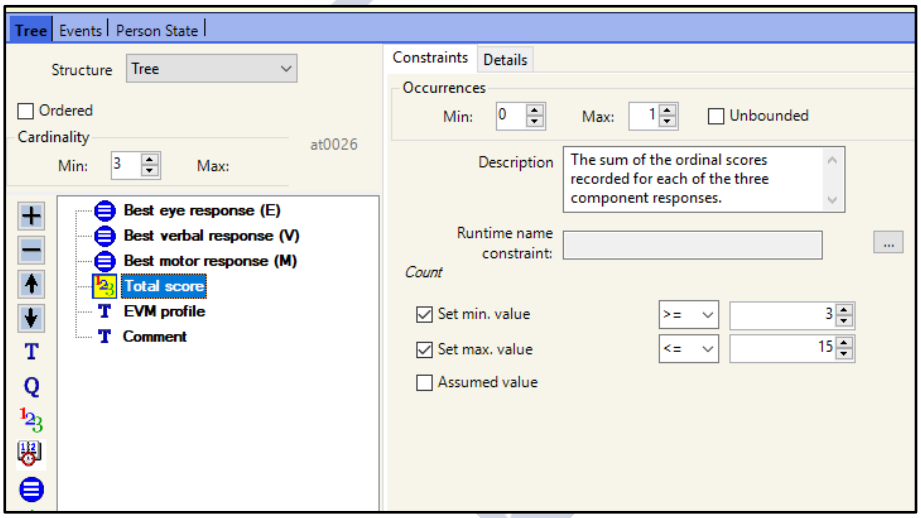

Fig. 2.4 Modeling the overall score of Glasgow Coma Scale archetype.

Additionally, other extra information is modeled according to the type of rating scales. For example, if they record data over time, this information is usually represented using point events provided by ADL, the duration of the observation, the state of the patient during data collection and the procedures used to gather the information. For example, the APGAR score is the first test given to a newborn to evaluate its physical condition.[49]. The APGAR test is given to a baby more than once. Therefore, the data must be recorded over time.

Fig. 2.5 displays the APGAR Observation archetype [50] that allows to record data after one minute, 2 minute, etc.

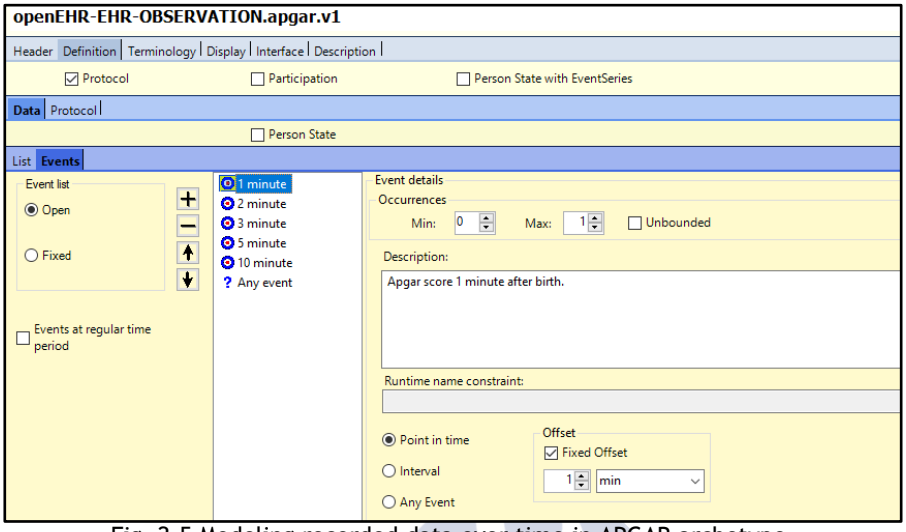

Fig. 2.5 Modeling recorded data over time in APGAR archetype.

2.4 HUMAN PHENOTYPE ONTOLOGY

Clinical archetypes use term mapping to standard terminologies/ontologies with the aim of normalizing the clinical data used in the model definition. Additionally, the use of a standard ontology provides the capability to automatically infer patient clinical phenotypes from data collected using the rating scale. The Human Phenotype Ontology (HPO) [20] delivers a structured and standardized vocabulary for phenotypic abnormalities encountered in human hereditary and other diseases. It is accessible at [5] and as of November 2017, it contains 13165 terms (classes) with 16794 is_a (is the same as subclassOf) relationships among those classes. Each class in the HPO describes an individual phenotypic abnormality. The is_a relationship describes the subclass-superclass relationships between the HPO classes. For example, *dysarthria is_a neurological speech impairment*. The *is_a* relationship is transitive, implying that if spastic *dysarthria is_a dysarthria*, which *is_a neurological speech impairment*, then *spastic dysarthria* is also a *neurological speech impairment.* As represented in OBO [51] format, each class can have up to 16 attributes

(*id, name, alternative ids, definition, synonym, references, is_a*, etc.). Fig. 2.6 displays the HPO class *Dysarthria* and its attributes.

```
[Term]
id: HP:0001260
name: Dysarthria
alt_id: HP:0002327
def: "Dysarthric speech is a general description referring to a
neurological speech disorder characterized by poor articulation.
Depending on the involved neurological structures, dysarthria may be
further classified as spastic, flaccid, ataxic, hyperkinetic and
hypokinetic, or mixed." [HPO:curators]
synonym: "Difficulty articulating speech" EXACT layperson []
synonym: "Dysarthric speech" EXACT []
xref: MSH:D004401
xref: SNOMEDCT_US:8011004
xref: UMLS:C0013362
is_a: HP:0002167 ! Neurological speech impairment
```

Fig. 2.6 The HPO class Dysarthria and its attributes.

2.4.1 The sub-ontologies of the HPO

The HPO has five sub-ontologies; Clinical Modifier, Mortality/Aging, Mode of Inheritance, Frequency, and Phenotypic Abnormality (Fig. 2.7)

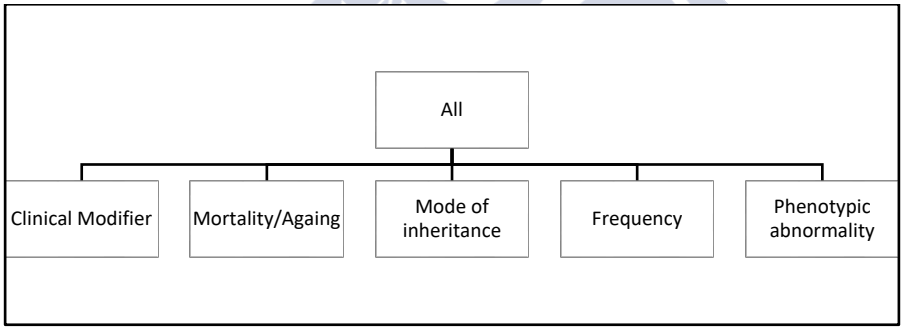

Fig. 2.7 The sub-ontologies of the Human Phenotype Ontology

2.4.1.1 Clinical Modifier

The Clinical Modifier sub-ontology contains classes that describe typical modifiers of clinical symptoms. For example Severity, the *Pace of progression*, the *Phenotypic variability* or the *Onset*. It comprises terms such as *Profound, Severe, Moderate, Mild, Profound, Childhood*

onset, *Variable progression rate*, or *Variable expressivity*. Fig. 2.8 shows the subclasses of the class Severity which is a subclass of the Clinical modifier class.

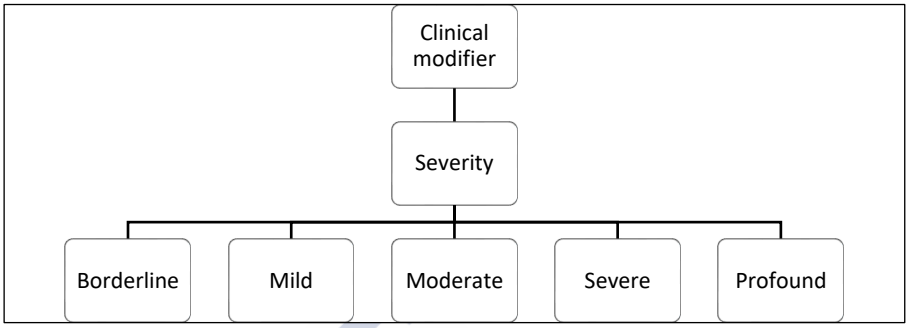

Fig. 2.8 The subclasses included into the class Severity

2.4.1.2 Mortality/Aging

This sub-ontology describes *Time of death* and includes classes such as *Death in early adulthood*, *Death in adolescence* or *Sudden death*.

2.4.1.3 Mode of Inheritance

This relatively small sub-ontology is intended to describe the mode of inheritance and contains terms such as *Autosomal dominant inheritance*, *Gonosomal inheritance*, *Multifactorial inheritance*, etc.

2.4.1.4 Frequency

This sub-ontology defines the frequency with that patients do show a particular clinical feature. It comprises terms such as *Very frequent, Very rare, Excluded, etc.*

2.4.1.5 Phenotypic abnormality

This is the core sub-ontology of the HPO and includes definitions of clinical abnormalities. It contains classes such as Abnormality of

blood and blood-forming tissues, Abnormality of the nervous system, Abnormality of the ear, etc. And it contains terms such as Intestinal carcinoid, Small intestine carcinoid, etc.

2.4.2 Terms Attributes

The majority terms of the HPO belong to the Phenotypic Abnormality sub-ontology. Each class has a unique ID such as *HP:0002503* and a name such as Spinocerebellar tract degeneration. Table 2.1 displays the name and description of each attribute that each class can have [51].

Table 2.1 The available attributes for each HPO class and their descriptions.

Attribute	Definition
id	The unique id of the current class. Cardinality: exactly one.
is_anonymous	To indicate if the current class has an anonymous id. Cardinality: zero or one.
name	The name of the current class. Each class may have only zero or one name defined. Cardinality: zero or one.
alt_id	Defines an alternate id for this class. Cardinality: any
def	The definition of the current class. Cardinality: zero or one
comment	A comment for this class. Cardinality: zero or one.
subset	It indicates a class subset to which this class belongs. Cardinality: any.
synonym	It gives a synonym for this class. Cardinality: any.
xref	It describes an analagous class in another vocabulary. It points to external disease databases such as Unified Medical Language System (UMLS)[6] and Medical Subject Headings (MeSH)[7]. Cardinality: any.
property_value	It binds a property to a value in this instance. Cardinality: any.
is_a	It describes a subclassing relationship. Cardinality: any.
created_by	Name of the creator of the class. Cardinality: zero or one.
creation_date	The creation date of the class. Cardinality: zero or one.
is_obsolete	It indicates whether the current class is obsolete. The allowable values are "true" and "false". Cardinality: zero or one.
replaced_by	It specifies a class which replaces an obsolete class. Cardinality: any.
consider	It determines a class which is an appropriate substitute for an obsolete class. Cardinality: any.

[6] https://www.nlm.nih.gov/research/umls/index.html
[7] https://www.nlm.nih.gov/mesh/

2.5 GUIDELINE DEFINITION LANGUAGE

The Guideline Definition Language (GDL)[8] is oriented to formally represent clinical procedural knowledge for computerized clinical decision support systems, using the format of knowledge rules [52]. GDL is designed to be natural language – and reference terminology - agnostic by leveraging the design of openEHR Archetype Model and openEHR Reference Model. GDL represents clear-cut clinical knowledge for singe-decision making. The importance of the GDL is:

- It allows expressing the rules of CDS using archetypes both as input and output for the rule execution.
- It is natural language-agnostic and support multiple language translations without changing the definitions of the rules
- It is reference terminology-agnostic so various terminologies can be used.
- It converts the CDS rules to main-stream general-purpose rule languages for execution
- It facilitates the reusability of the CDS rules in different clinical contexts.
- It allows grouping a set of related CDS rules in order to support complex decision making.
- It is technology independent.

2.5.1 Components of GDL

GDL has four main parts: Header, Definition, Rule and Ontology.

2.5.1.1 GDL Header

The *Header* introduces the GDL and contains metadata such as authors, keywords and information about the purpose, etc. Fig. 2.9 displays the *Header section of "CHA2DS2VASc"* GDL guide document and it illustrates the main parts in this section. *CHA2DS2-*

[8] http://www.openehr.org/releases/CDS/latest/docs/GDL/GDL.html#_guideline_definition_language_gdl

VASc is a clinical instrument for stroke risk stratification in atrial fibrillation [53].

```
(GUIDE) <
gdl_version = <"0.1">
id = <"CHA2DS2VASc_Score_calculation.v1">
concept = <"gt0001">
language = (LANGUAGE) <
                original_language = <[ISO_639-1::en]>  >
description = (RESOURCE_DESCRIPTION) <
details = <
   ["en"] = (RESOURCE_DESCRIPTION_ITEM) <
      keywords = <"atrial fibrillation", "stroke", "CHA2DS2-VASc">
      purpose = <"Calculates stroke risk for patients with atrial
      fibrillation, possibly better than the CHADS2 score.">
      use = <"Calculates stroke risk for patients with atrial fibrillation,
      possibly better than the CHADS2 score.">
                        >
                >
                original_author = <
                ["date"] = <"2012/12/03">
                ["email"] = <"rong.chen@cambio.se">
                ["name"] = <"Rong Chen">
                ["organisation"] = <"Cambio Healthcare Systems">
                >
      >
>
```

Fig. 2.9 Extract of the Header section in "CHA2DS2VASc.gdl".It illustrates the current version of the guideline, authorship information, keywords, purpose and use of the guideline. (Taken from [29])

2.5.1.2 GDL Definition

The *Definition* section contains all the elements used inside the guideline, the mappings to the archetypes and pre-conditions. Fig. 2.10 displays the *Definition section of "CHA2DS2VASc"* GDL guide document. It displays the archetype_binding within the

guide_definition section, which binds data elements from the archetypes to variables used by GDL rules. It also illustrates that there is a condition (pre-condition) must be met before the rules inside the guide can be executed. For example, this guideline will not be executed unless the patient has atrial fibrillation.

```
definition = (GUIDE_DEFINITION) <
    archetype_bindings = <
     ["gt0002"] = (ARCHETYPE_BINDING) <
     archetype_id = <"openEHR-EHR-EVALUATION.problem-diagnosis.v1">
     domain = <"EHR">
     elements = <
                              ["gt0015"] = (ELEMENT_BINDING) <
                              path = <"/data[at0001]/items[at0002.1]">
                          > > > >
     pre_conditions = <"$gt0015!=null",...>
>
```

Fig. 2.10 Extract of the Definition section in "CHA2DS2VASc.gdl". (Taken from [29])

2.5.1.3 GDL Rule

The *Rule* section contains the condition and action parts of rules. Each rule consists of two parts: the first part contains the conditions needed for the rule to execute and it starts with the keyword ‘When’, and the second one comprises the actions that will be carried out once the rule is activated and it starts with the keyword ‘Then’. Fig. 2.11 shows rules that inspect different diagnoses relevant to CHA2DS2-VASc score.

```
rules = <
["gt0027"] = (RULE) <
                when = <"$gt0016!=null",...>
                then = <"$gt0007=1|local::at0028|Present|",...>
                priority = <11>  >

["gt0028"] = (RULE) <
                when = <"$gt0017!=null",...>
                then = <"$gt0008=1|local::at0031|Present|",...>
                priority = <11>  >
        >
```

Fig. 2.11 Extract of the Rule section in "CHA2DS2VASc.gdl".(Taken from [29])

2.5.1.4 GDL Ontology

In the *Ontology* section, all the terms are bond to user interface labels and description of the terms in supported natural languages. In addition, terms are bound to external terminologies. Fig. 2.12 illustrates how the atrial fibrillation term is bound to a specific code in ICD10[9].

```
ontology = (GUIDE_ONTOLOGY) <
        term_bindings = <
          ["ICD10"] = (TERM_BINDING) <
                bindings = <
                ["gt0015"] = (BINDING) <
                codes = <[ICD10::I48],...>
                uri = <"">   >
                                > > >>
```

Fig. 2.12 Extract of the Ontology section in "CHA2DS2VASc.gdl".(Taken from [29])

The GDL editor[10] enables users to create and run GDL files, and it is a multiplatform desktop application.

[9] http://www.who.int/classifications/icd/ICD-10_2nd_ed_volume2.pdf
[10] https://sourceforge.net/projects/gdl-editor/

2.6 WEB ONTOLOGY LANGUAGE

Ontologies are used to capture and model knowledge about some domain of interest. The Web Ontology Language (OWL) [54] is the most recent development in standard ontology languages. The OWL is a description-logic-based language used to formalize the concepts and relationships of a domain in terms of individuals, properties, classes, and data values. It represents rich and complex knowledge about concepts of a domain and relations between concepts. It has a richer set of operators – e.g. negation, union, and intersection. OWL depends on a different logical model which facilitates the definition and the description of concepts. Therefore, complex concepts can be built up. Additionally, the logical model provides the use of reasoners. Reasoners like Hermit are used to check if definitions and statements in ontologies are mutually consistent and also to identify which concepts fit under which definitions. Correctly maintaining the hierarchy is a critical issue, especially when dealing with cases where there are concepts with at least two parents. Reasoners are very helpful in preserving this hierarchy [55] .

2.6.1 Components of OWL Ontologies

An OWL ontology consists of Individuals, Properties, and Classes [55].

2.6.1.1 Individuals

Individuals represent objects in the interested domain. They are also known as instances of classes. For example, Microsoft is an instance of a class called Company.

2.6.1.2 Properties

Properties are binary relations between individuals – i.e. They link two individual together. For example, the property hasOwner might link the individual Microsoft to Bill Gates. Properties can have inverses. For example, isOWnedBy is the inverse of hasOwner. They can also be either symmetric or transitive.

2.6.1.3 Classes

OWL classes are sets that comprise individuals. They are described using formal descriptions. Classes' descriptions specify precisely the requirements for memberships of the classes. For example, the class Person would contain all the individuals that are persons in the interested domain. Classes can be organized into a superclass-subclass hierarchy.

Fig. 2.13 shows a representation of two classes Company and Person which are represented as circles, a representation of four individuals Microsoft, Facebook, Bill Gates, and Mark Zuckerberg which are represented as diamonds, and a representation of a property isOwenBy which is represented as a curved line.

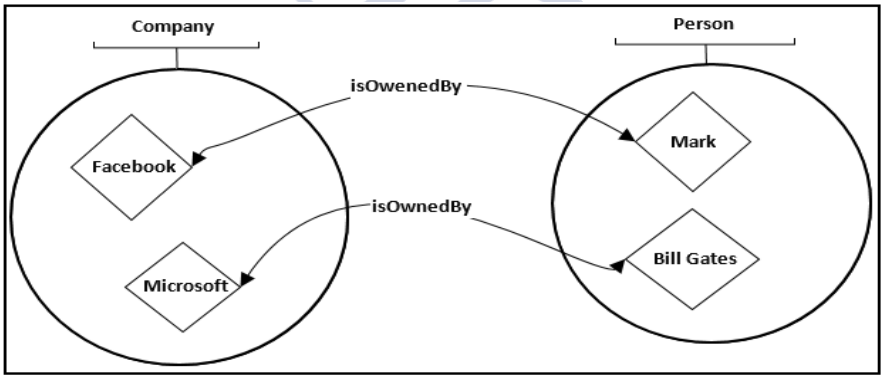

Fig. 2.13 Representation of Individuals, Properties and Classes. Individuals are represented by diamonds, properties by arrows and classes by circles.

2.7 THE SCALE FOR THE ASSESSMENT AND RATING OF ATAXIA (SARA)

The SARA assesses severity of cerebellar dysfunction through the evaluation of eight items reflecting motor performance (gait, stance, sitting, speech disturbance, finger-chase test, nose-finger test, fast alternating hand movements and heel-shin test) [56] (See Appendix A for the full details of the items). For the last four items, upper and lower

extremities are evaluated bilaterally, the mean values of both sides are calculated and added up to the scores of the first four items. The total score ranges from 0 (no ataxia) to 40 (most severe ataxia) (Fig. 2.14). For spinocerebellar ataxia, normative data are available with mean SARA score of 15.9 ± 8.5 (range = 1.5 to 40) and mean SARA score for controls of 0.4 ± 1.1 (range = 0 to 7.5). It is straightforward to apply SARA, as it takes requires no specific training, and has excellent internal consistency and inter-rater reliability. The SARA is used in clinical studies of cerebellar disorders for an accurate evaluation of the patient´s motor performance, both globally and at the individual items. It is also useful for quantitative comparison of patients, ataxia types, disease stages and response to treatment, among other applications. As for other medical scales, one of the challenges of the SARA is the need to derive a qualitative description and patient classification with diagnostic implications from numerical scores. An automated system to solve this translation would greatly facilitate the use of the SARA – and, by extension, other scoring systems- by clinicians on both research and clinical routine settings.

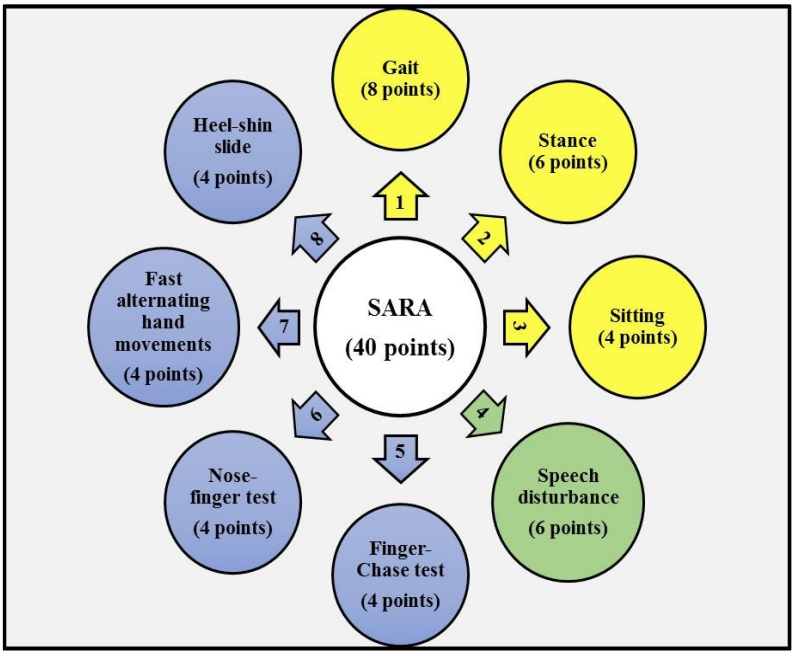

Fig. 2.14 The eight items included in the SARA scale for cerebellar ataxia.

2.8 SPINOCEREBELLAR ATAXIA TYPE 36 (SCA36)

The spinocerebellar ataxias (SCAs) are a heterogeneous group of dominantly inherited neurodegenerative disorders, caused by diverse mutation that affect the cerebellum and its connections (Fig. 2.15) [57, 58]. The spinocerebellar ataxia type 36 (SCA36 or Ataxia da Costa da Morte) is a SCA subtype identified in large families from the Northwestern coast of Galicia [59]. It is a slowly progressive neurodegenerative disorder characterized by adult-onset gait ataxia, eye movement abnormalities, tongue fasciculations, and variable upper motor neuron signs. People with this condition initially experience problems with coordination and balance (ataxia), usually in the fifth or sixth decades of life.

The phenotype of SCA36 is characterized by the following findings [60]:

- Midline cerebellar ataxia of late onset (usually between ages 40 and 60 years) and slow progression.
- Dysarthria and appendicular ataxia generally following the gait imbalance.
- Slowly progressive sensorineural hearing loss (SNHL) with onset usually a few years after the cerebellar manifestations
- Tongue atrophy and fasciculations, additional signs of motor neuron degeneration in some cases [59, 61]
- Other clinical features variably present: gaze-evoked nystagmus, eyelid ptosis, decreased vibration sense, and cognitive impairment

- On brain MRI: atrophy of the superior vermis in initial stages, global cerebellar atrophy in intermediate stages, and olivopontocerebellar atrophy in advanced stages.

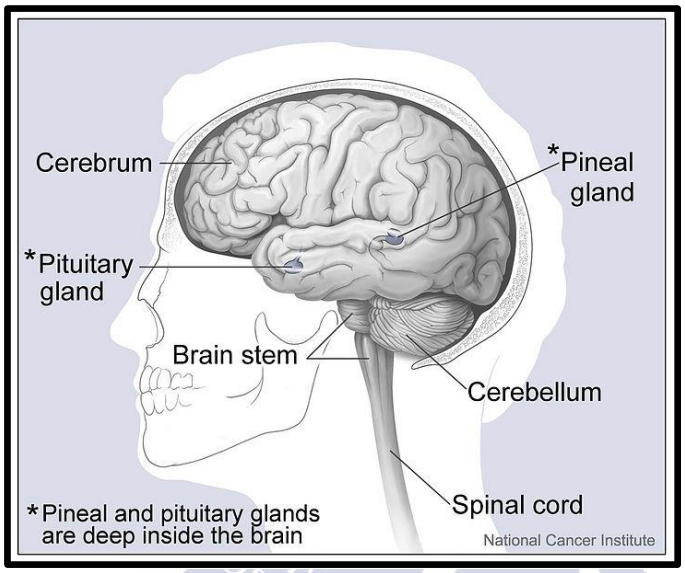

Fig. 2.15 Cerebrum and Cerebellum.(Taken from [62])

3 METHODS

The main contribution of this PhD work was to provide an approach to develop electronic rating scale in the clinical domain of the neurology, covering not only the content but also the clinical interpretation of the collected data. The main innovation was to build an application beyond a simple calculator that incorporated the expert knowledge required to clinically interpret the collected information. The benefits of this solution were to provide clinical standardization not only during the collection of data but also during the interpretation, and facilities to automatically generate reports, relieving the physician from doing this task. With the aim of implementing a solution that was easily interoperable with the clinical setting, we built it on Electronic Health Record (EHR) standards and on semantic web technologies.

With the aim of facilitating the ontology mapping and preventing large semantic discrepancies between clinical archetypes and ontologies, we propose to reorganize the classical methodology of clinical archetype modeling. At the heart of our methodology is the assumption that the archetype design should be supported by ontologies in those clinical situations where it is expected that archetype contents be logically organized On the other hand, the interpretation of the SARA required different types of information: data to be recorded (i.e., the content of the rating scale), knowledge about the meaning of terms in the scale (i.e., terminological knowledge), procedural knowledge to understand the meaning of scores (that can be easily expressed by guidelines), and ontological knowledge to deduce patient phenotypes. We chose to use a combination of GDL, openEHR clinical archetypes and ontologies to address the challenges of modeling the rating scale. The research questions addressed in this work, as we already stated in the introduction of this PhD work, are: (1) Is the combination of GDL, openEHR clinical archetypes and ontologies suitable for the description of all knowledge covered by the SARA? and (2) Is it

possible to achieve a reasonable integration of these technologies to efficiently model the rating scale?

In order to emphasize the benefits of our approach, we will start describing a use case of the SARA which highlights how the total score of a rating scale is not enough for interpreting the clinical status of a patient, and the physician usually interprets the patient data with background information about the scale. Then, the chapter focus on the main steps of the modeling approach. Section 3.2 describes the main activities proposed in our approach to model rating scales. Next, Sections 3.2.1 to 3.2.4 detail each proposed activity.

3.1 A USE CASE DESCRIPTION FOR THE SARA SCALE

Let´s consider patients 1, 2 and 3, with the same total SARA score of 20 (Table 3.1). Just based on the total score, the three patients would be considered to be in similar clinical stage. However, their functional situation is notably different. While patient 1 scores very high for midline ataxia (which is concluded from the high values of the three first items) and can barely walk or sit unaided, patients 2 and 3 have compromised speech (item number 4) and limb coordination (from the values of the last four items). In turn, patients 2 and 3 – with similar sitting, standing and walking performance – have different degree of speech impairment of speech (mild in patient 2, while verbal communication is impossible for patient 3). The total – and even just partial scores for limbs – also do not help differentiate the actual phenotype of patients 2 and 3, who have very different performance with their limbs (appendicular ataxia derived from the last four items). While patient 2 has significantly impaired motor coordination on both sides, patient 3 has a more asymmetrical cerebellar syndrome, with severe impairment on his left side, but only very mild involvement of his right side, which may be of enormous relevance to his functional ability if the patient is right-handed.

Table 3.1 Example scenario for the SARA scale. Table shows three hypothetical patients with the same value for the total SARA score. Items are represented by numbers: 1-Gait, 2-Stance, 3-Sitting, 4-Speech disturbance, 5-Finger-chase test 6-Nose-finger test, 7-Fast alternating hand movements and 8-Heel-shin test. Upper and lower extremities are evaluated bilaterally. R represents Right and L, left. Mean R-L is the mean value of both sides

SARA Items	Patient		
	1	2	3
1	6	2	2
2	6	2	2
3	3	1	1
4	1	3	6
5-R	0	3	1
5-L	0	4	4
Mean R-L	0	3.5	2.5
6-R	1	3	1
6-L	1	4	4
Mean R-L	1	3.5	2.5
7-R	1	3	1
7-L	1	3	4
Mean R-L	1	3	2.5
8-R	2	2	0
8-L	2	2	3
Mean R-L	2	2	1.5
Total Score	20	20	20

3.2 THE PROPOSED MODELING APPROACH

This chapter covers the key technical aspects of our proposal for modeling the SARA, which requires deciding the most appropriate computational representation for 1) the set of clinical statements included into the rating scale, 2) the calculation strategy to compute the total score grading the global clinical dimension, and 3) the underlying clinical knowledge of the rating scale required to properly interpreting

the results of the scale. Our modeling approach is based on four main steps (Fig. 3.1):

1) Building of a reduced version of the HPO, through extraction of the ontology modules relevant to the SARA
2) Annotation of the free-text descriptions of the rating scale with the ontology modules
3) Development of two clinical archetypes (Observation and Evaluation)
4) Definition of the information-processing units to express the clinical interpretation support system.

Each step involves several activities, described below and summarized in (Fig. 3.2).

The modeling of SARA involved a data level - representation of the SARA items- and a knowledge level- referred to the strategy to compute the total score and interpretation of phenotype. Archetypes were used to model the data level, while GDL and OWL were used to model the knowledge level. This representation had restrictions, since openEHR models support GDL, but do not give much support for OWL and related reasoning. To bring the gap between the clinical archetypes and the ontology, mappings were defined. These mappings facilitated the translation of the archetype instances to the OWL dataset.

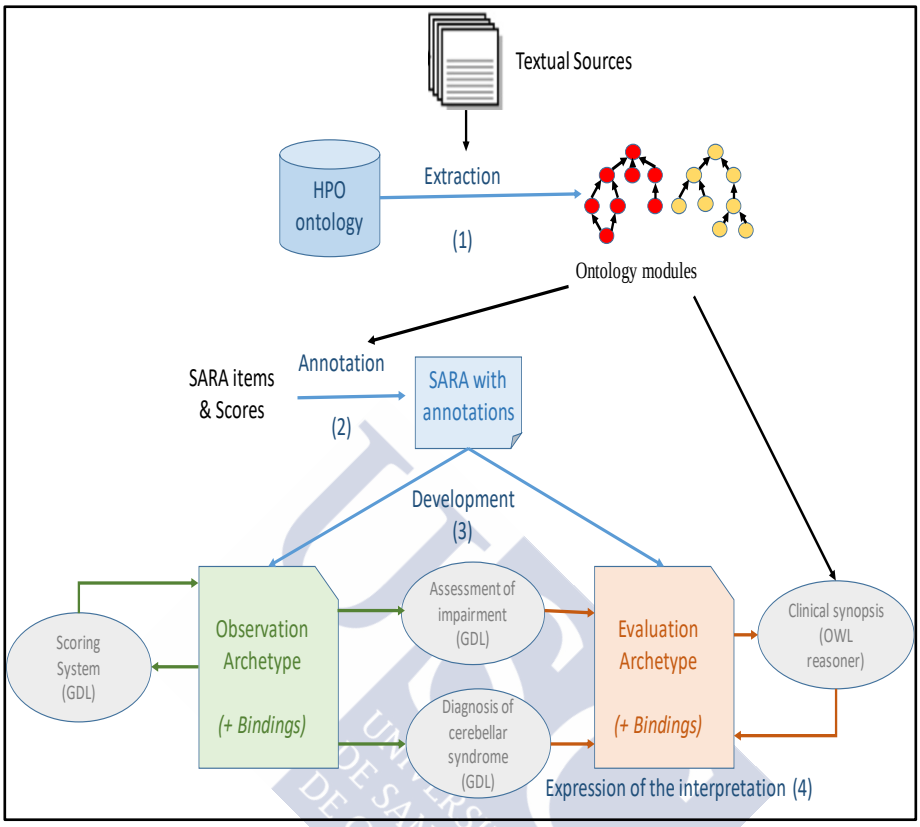

Fig. 3.1 The modeling approach. It includes: 1) Extraction of a reduced version of the HPO; 2) Annotation of the free-text descriptions of SARA items and scores; 3) Development of two archetypes (Observation and Evaluation); 4) Definition of the information-processing units in order to express the clinical interpretation of the SARA.

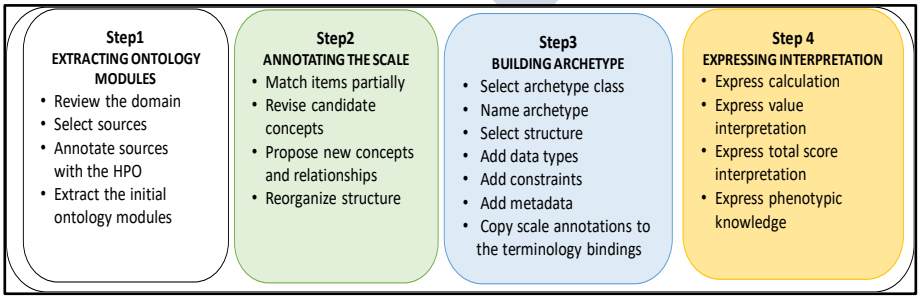

Fig. 3.2 Summary of the main proposed activities for the modeling of electronic rating scales.They cover: extracting ontology modules, annotating the scale, building archetypes, and expressing interpretation.

3.2.1 Extracting the HPO ontology modules relevant to the SARA

We reviewed and collected free-text sources describing the SARA, its component tests and application. For example, the following text describes the functional subdivisions of the cerebellum in order to explain the rationale of a coordination exam: "*The cerebellum has 3 functional subdivisions, …. The first is the vestibulocerebellum. … Dysfunction of this system results in nystagmus, truncal instability (titubation), and truncal ataxia ... The following tests of the neuro exam can be divided according to which system of the cerebellum is being examined….".* We then annotated the sources to HPO terms, using the OBO Annotator [63], a phenotype concept recognition system (Fig. 3.3). In total, 12 HPO classes were annotated, which provided the set of seed terms required to extracting a self-contained portion of the HPO. Such reduced version of the HPO covered all the classes relevant to the SARA.

In general, self-contained portions of an ontology are referred as ontology modules/segments [64], or slims in the context of the Gene Ontology. While these modules are subgroups of a base ontology, in our case the HPO, they are also equally valid on their own [65], but simpler and more manageable than the complete ontology.

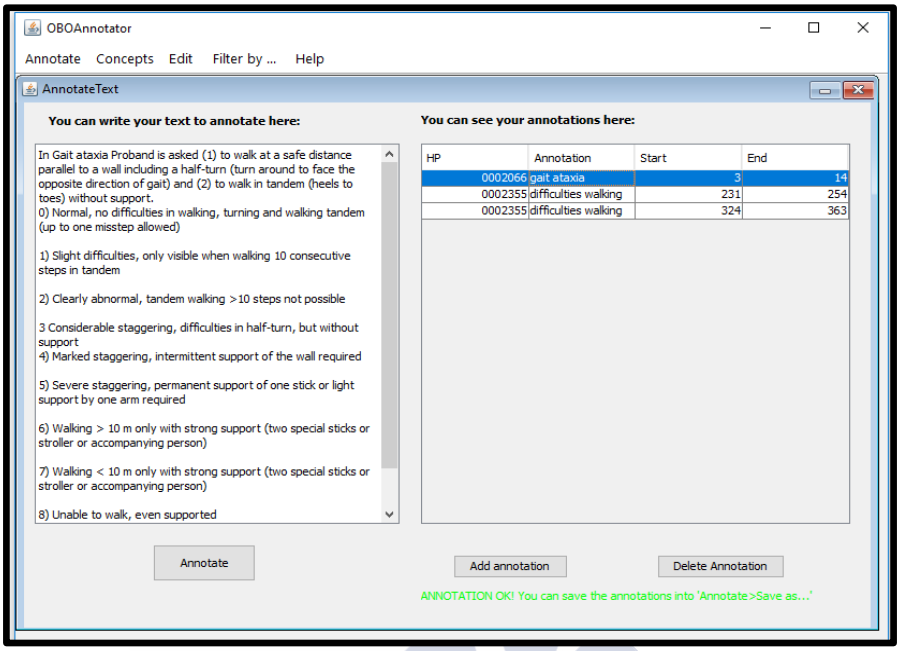

Fig. 3.3 OBO Annotator interface to annotate a text to HPO classes.The text related to Gait Ataxia is annotated to HP:0002066 (Gait Ataxia) and HP:0002355 (Difficulties Walking). Annotating the rating scale SARA

3.2.2 Annotating the rating scale SARA

The goal of this stage is to map the SARA items with standard patient phenotypes described into its own scale. Firstly, items were partially matched to ontology modules class names. Partial match happens when the item name is embedded inside some class name. An ontology engineering of our research group (Maria Taboada) programmed a specific method to run partial matches. Thus, one or more candidate classes for each item were obtained. Then the candidate mappings were revised by a neurologist (Maria Sobrido), who selected the most appropriated classes and proposed a minimal extension and reorganization of the the reduced version of the HPO, with the aid of the complete HPO, in order to cover all the details of the needed neurodegenerative phenotypes, but keeping it as close to the original HPO as possible. Six new classes, with the new relationships, were added to the ontology modules to precisely annotate four SARA items (stance, sitting, finger chase and heel-shin slide). Fig. 3.4 shows the

new classes and relationships, and the mappings between items in the SARA and classses in the ontology modules. The golden color is used to highlight the new classes and is_a relationships. On the other hand, each score in the SARA is accompanied by a textual description, describing a level of severity. In order to annotate the scores, we decided to reuse the general HPO classes describing the different levels of severity: borderline, mild, moderate, severe and profound. We introduced new subclasses to the eight HPO classes that were used to annotate the SARA items. The subclasses were defined based on the severity levels of their superclasses. For example, *moderate_dysarthria* was defined as a subclass of *dysarthria* with a moderate severity (Fig. 3.5).

Adittionally, two scores were identified by the neurologist as two HPO phenotypes. The first one is the score 8 for the Gait item, which was bound to *abasia* and the second is the score 6 for the Speech Disturbance item, which was bound to *anarthria*. In addition, the neurologist added a new class (named *astasia*) to map the score 6 for the stance item.

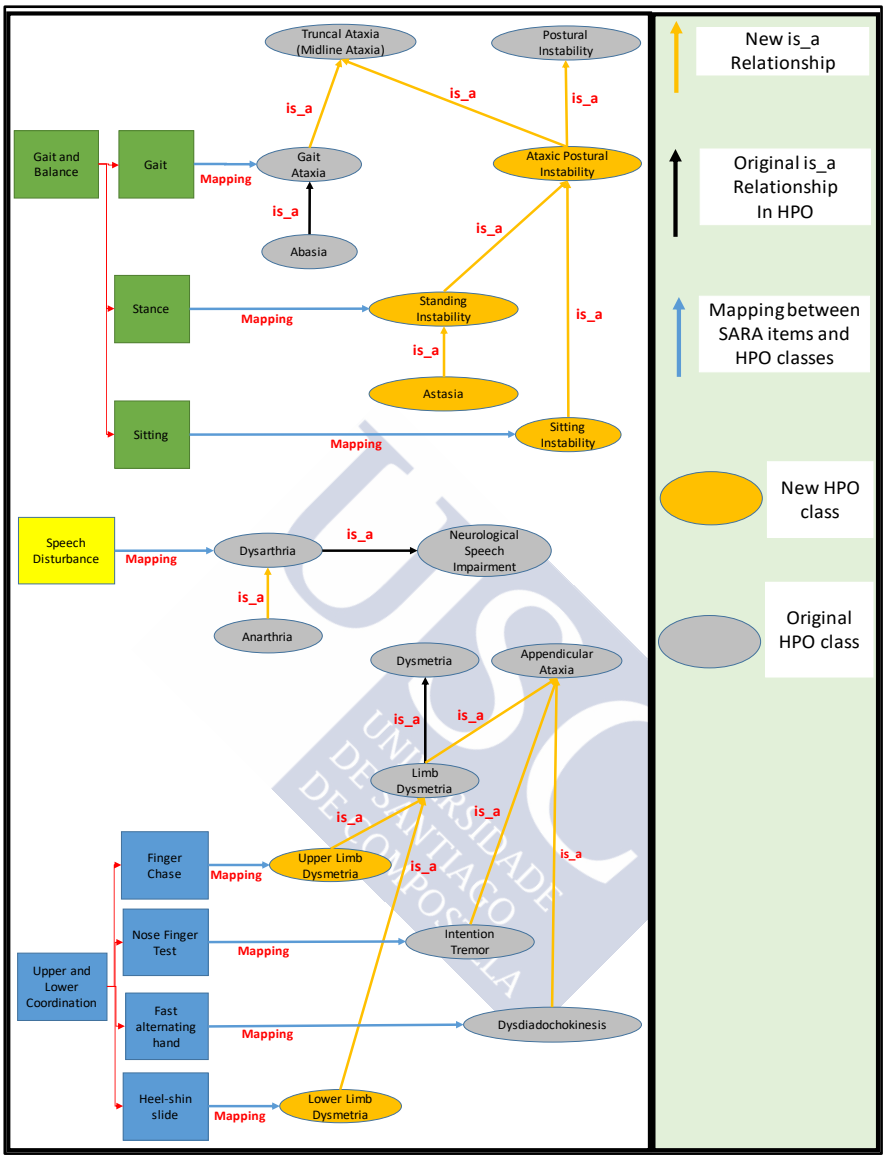

Fig. 3.4 Excerpt from the set of mappings between the SARA and the ontology.Squares represent the SARA items. Gray and golden ellipses, respectively, are the original HPO classes and classes added to the ontology modules. Blue arrows are mappings between SARA items and HPO classes. Black and golden arrows, respectively, represent the original and the additional is_a relationships. Note that subsumptions in Electronic health record (Open Biological Ontologies) are represented using the is_a relationship, whereas in OWL using the subclassOf constructor.

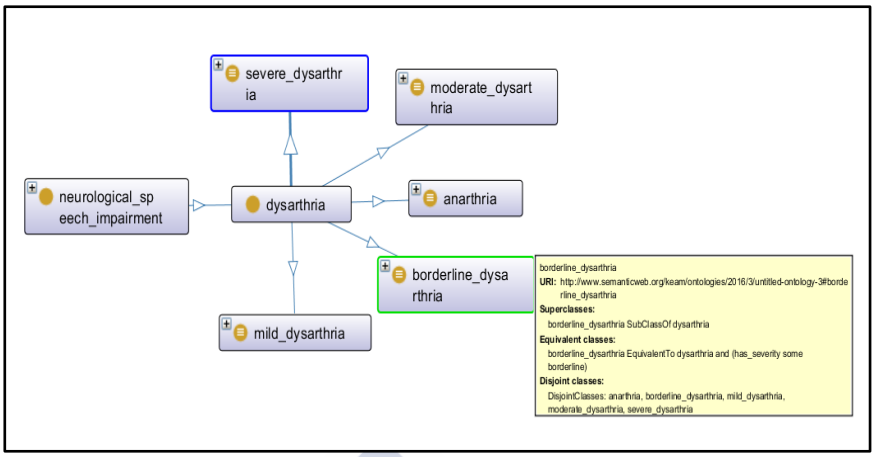

Fig. 3.5. The five subclasses of dysarthria class.

Once the ontology modules was modelled and the annotations were created, three main superclasses were identified taking the eight scale items into account: i) truncal ataxia (midline ataxia), which subsumed the classes *gait_ataxia* and ataxic_postural_instability (which subsumed *standing_instability* and *sitting*_instability); ii) dysarthria, annotating the fourth item; and iii) Appendicular Ataxia (limb ataxia which subsumed *limb dysmetria*, *intention tremor* and *dysdiadochokinesis* (Fig. 3.6). Thus, we structured and organized the rating scale items in accordance with the ontology, by inserting three *CLUSTER* nodes in a hierarchical structure: *i)* gait and balance, which is linked to truncal ataxia*; ii) speech disturbance*, which is relating to dysarthria*;* and iii) *upper and lower limb coordination*, which is associated with Appendicular Ataxia (limb ataxia). This new structure does not alter in any way the scale, as it continues to be based on an 8-item performance. It simply arranges the items. This new organization is shown on the upper part of the Fig. 3.7 A. The new nodes can be viewed as clinical dimensions, but without any assigned score. Additionally, each item in the third node or clinical dimension was split into left part and right part, in order to capture the item on each side, as set out in the original rating scale. Once again, we organized them in a hierarchical structure. Finally, the SARA-specific, reduced version of the HPO was translated to Protégé[11], and its properties were manually

[11] http://protege.stanford.edu/products.php#desktop-protege

modeled. To check the consistency, the HermiT reasoner was used.

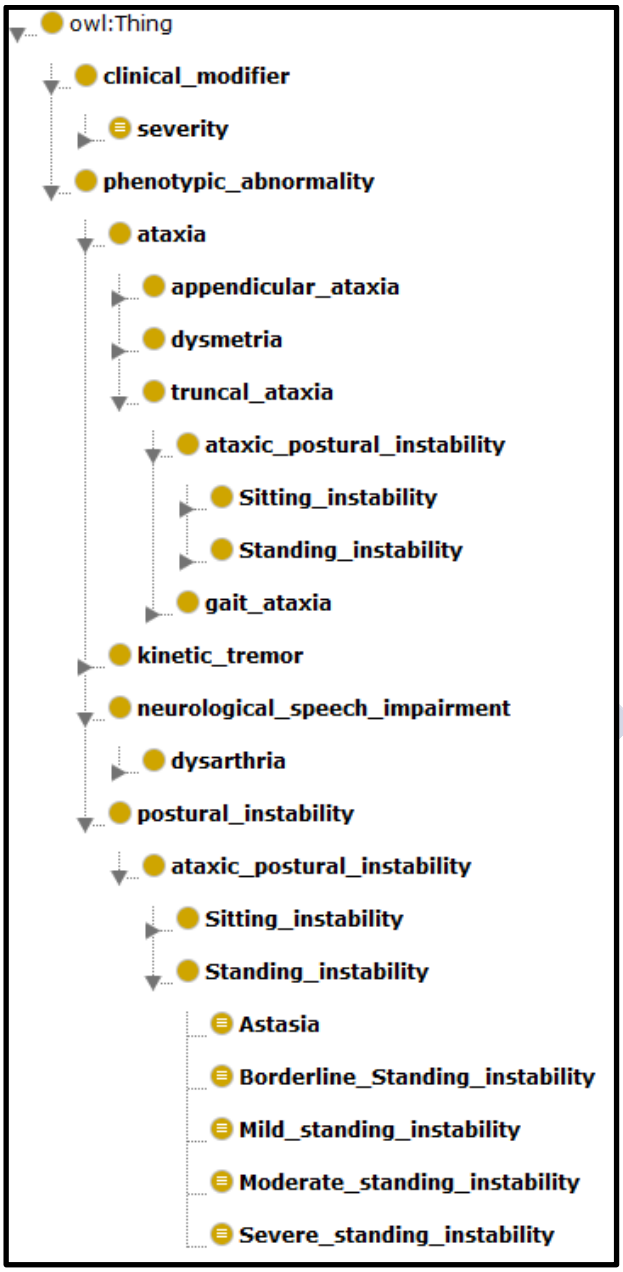

Fig. 3.6. Excerpt from the domain ontology.
It shows some classes of the reduced version of the HPO.

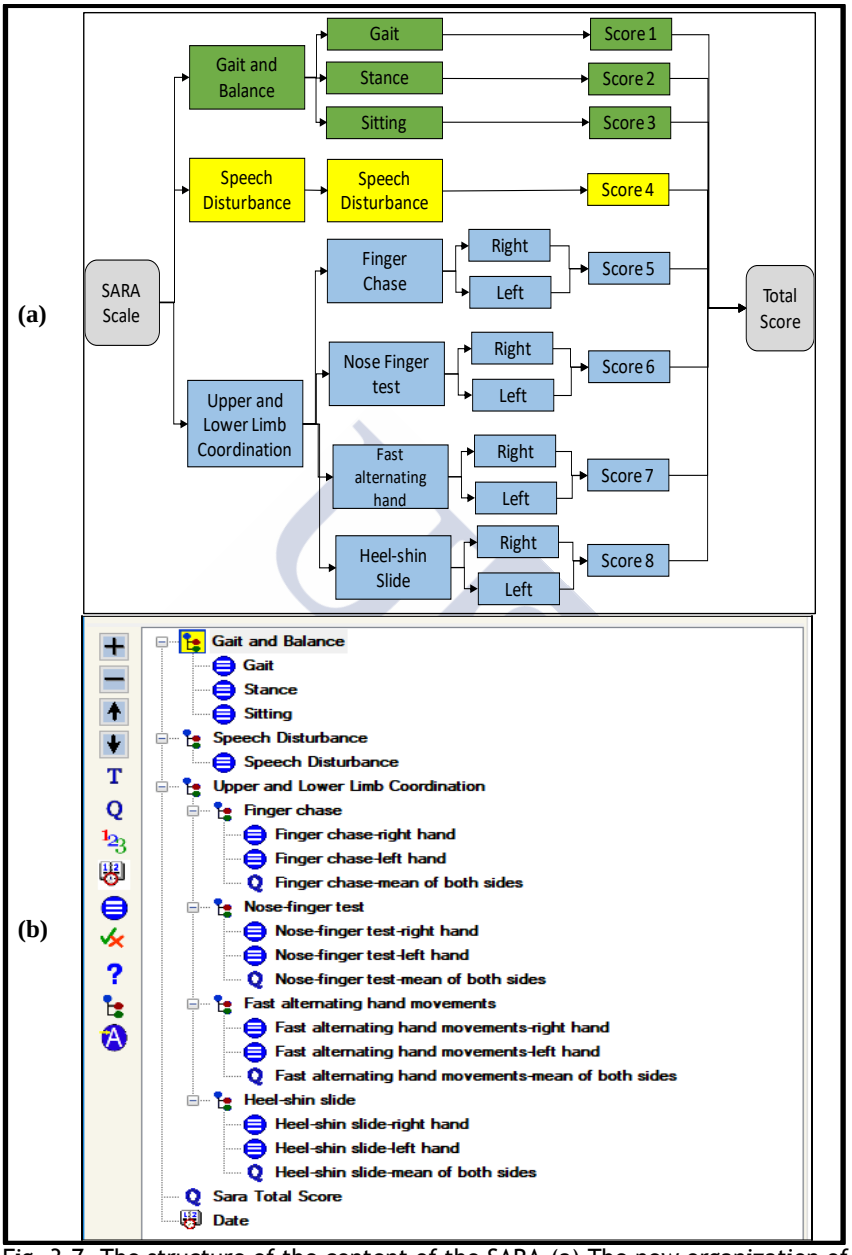

Fig. 3.7. The structure of the content of the SARA.(a) The new organization of the SARA, with three main patient's phenotype components: gait and balance, speech disturbance, and upper and lower limb coordination; (b) The clinical observation archetype developed for the SARA.

3.2.3 Developing the archetypes for the SARA

For the archetype design, we made a difference between data to be gathered by the scale or directly calculated using the information provided by the scale (items, clinical dimensions, individual scores and the total score), and data that are not directly available in the scale. Taking account of this distinction, we decided to use two types of clinical archetypes: *Observation* for normalizing data provided by the scale, and *Evaluation*, for data not directly available in the scale. The first one fits to capture the scale content, whereas the second one records clinically interpreted findings, such as the phenotypic abnormalities derived from the rating scale. The two clinical archetypes were developed using the archetype editor[12] provided by OpenEHR.

3.2.3.1 Modeling the SARA Observation archetype

To model the observation archetype, after adding the metadata (e.g. purpose, keywords, definition, author, etc.), we structured the content of the SARA according to this archetype, i.e., by means of a tree structure with *elements* for the three clinical dimensions (the set of items, the total score and the date of the observations), and *values* for the individual scores. Then, we defined the *elements* with proper data types, descriptions, comments, details, occurrences, constraints and possible values. We mapped each *element* to the corresponding ontology class of the reduced version of the HPO, using the achieved annotations (See Section 3.2.2). Finally, we structured and organized the items in accordance with the ontology, by inserting three *CLUSTER* nodes in a hierarchical structure: *i) gait and balance*, which was linked to *truncal ataxia; ii) speech disturbance*, which was linked to *dysarthria;* and iii) *upper and lower limb coordination*, which was associated with *appendicular ataxia*. This new structure did not alter the SARA, as it continued to be based on the same 8-items, but only arranged these items in a specific way (Fig. 3.7 A). The new nodes can be viewed as clinical dimensions, but without any assigned score. Fig. 3.7 B shows that the observation archetype meets the rating scale structure. As the archetype editor did not provide HPO in the list of

[12] http://www.openehr.org/downloads/archetypeeditor/home

terminologies, we added the terminology to the file “Terminology.xml” used by the archetype editor. Fig. 3.8 displays an excerpt of the modified “Terminology.xml” file including the human phenotype ontology.

```
<TerminologyIdentifiers VSAB="HHC96" SourceName="Home Health Care Classification, 1996" Authority="UMLS2003AA" />

<TerminologyIdentifiers VSAB="HL7_1998-2002" SourceName="Health Level Seven Vocabulary, 1998-2002" Authority="UMLS2003AA" />

<TerminologyIdentifiers VSAB="HPO2015_04_20" SourceName="Human Phenotype Ontology, 2015_04_20" Authority="UMLS2015AB" />

<TerminologyIdentifiers VSAB="HLREL_1998" SourceName="ICPC2E-ICD10 relationships from Dr. Henk Lamberts, 1998" Authority="UMLS2003AA" />
```

Fig. 3.8 Excerpt of the modified "Terminology.xml" including HPO.

3.2.3.2 Modeling the SARA Evaluation archetype

The Evaluation archetype (Fig. 3.9) is intended to record the clinical interpretations from the patient observations (collected following the Observation archetype). We distinguished three types of interpretations:

- Interpretations directly inferred from individual items or total score (i.e., from the corresponding element in the observation archetype). For example, the presence of gait ataxia and the severity degree from the values for the first item.
- Interpretations directly inferred from specific values for an individual item. For example, the presence of abasia from the value 8 (unable to walk, even supported) for the item gait.
- Interpretations elucidating the patient abnormalities associated with the clinical dimensions of the scale. For example, the presence of midline ataxia and the severity degree from the values for the first three items (gait, stance, and sitting).

We modeled this archetype following the same structure of the SARA Observation archetype, by means of a tree with three main clusters, one for each main clinical dimension: gait & balance, speech disturbance, and upper & lower limb coordination. Each cluster consists of the interpretations directly or indirectly derived from the corresponding elements in the observation archetype. Additionally, every element involves different levels of severity, so the "Choice" data type was selected with Text/ Internal codes as constraints. Finally, we mapped the elements to ontology classes.

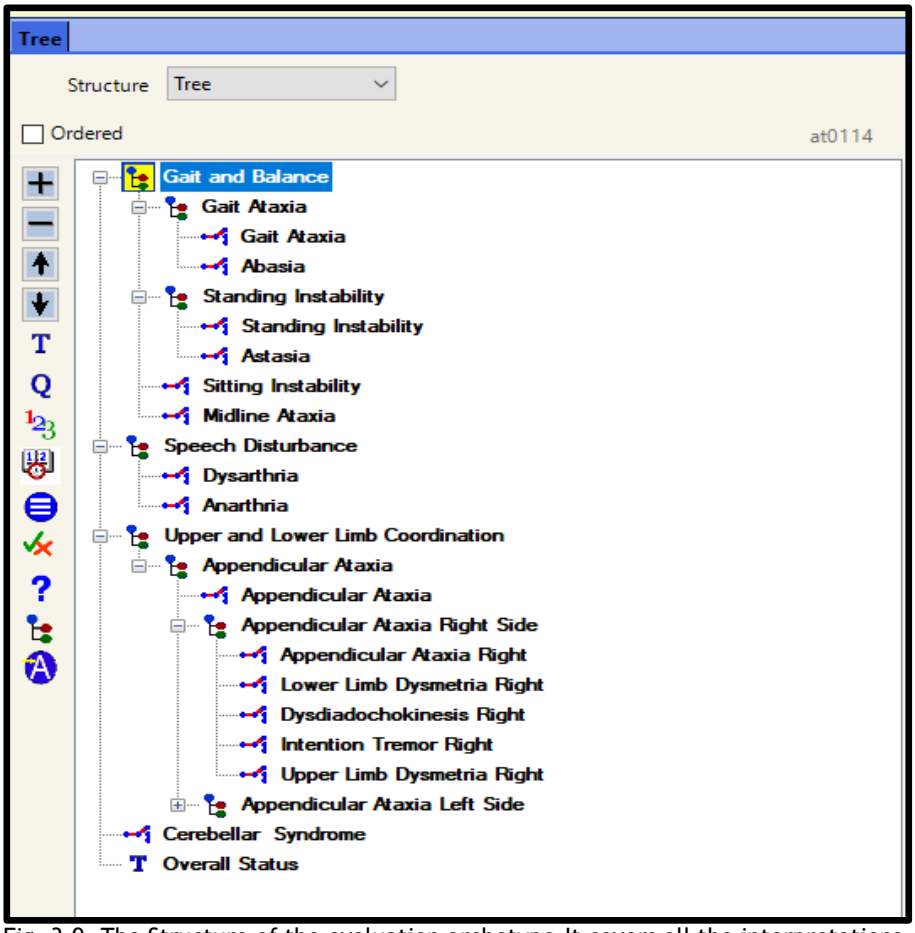

Fig. 3.9. The Structure of the evaluation archetype.It covers all the interpretations that can be derived from the SARA.

3.2.4 Modeling the SARA interpretation with GDL

Firstly, we analyzed the different information-processing units (or inferences) required to automatically interpret the data obtained through the application of SARA. In Table 3.2, these inferences are shown with the knowledge format, the input, the output and the representation model. We need point out that the SARA supplies standardized calculation functions, whereas it does not provide any interpretation knowledge. Hence, with the support of the ontology and acquiring the experience of the neurologist, we were able to express the interpretation of the SARA, although this is not standardized. Appendix B illustrates all the implemented GDL rules

Table 3.2 Information-processing units. Each unit is defined by means of the used knowledge format, the inputs, outputs and the representation language.

Information-processing units	Knowledge format	Input	Output	Representation language
Calculation functions	Mathematical expressions	Observation archetype items	Observation archetype items	GDL
Assessment of the degree of impairment	Cut-off level based severity rules	Observation archetype item	Evaluation archetype items	GDL
Assessment of cerebellar syndrome	Heuristic rules	Observation archetype items	Evaluation archetype items	GDL
Clinical synopsis	Reduced version of the HPO	Evaluation archetype items	Evaluation archetype items	OWL

3.2.4.1 Calculation functions

The calculation functions expressed the mathematical operations to compute both the arithmetic means of the item scores and the total score. The input and output of these functions were elements defined in the observation archetype. In total, five GDL rules were modeled to

cover these functions. Fig. 3.10 displays the rule that calculates the mean of finger chase.

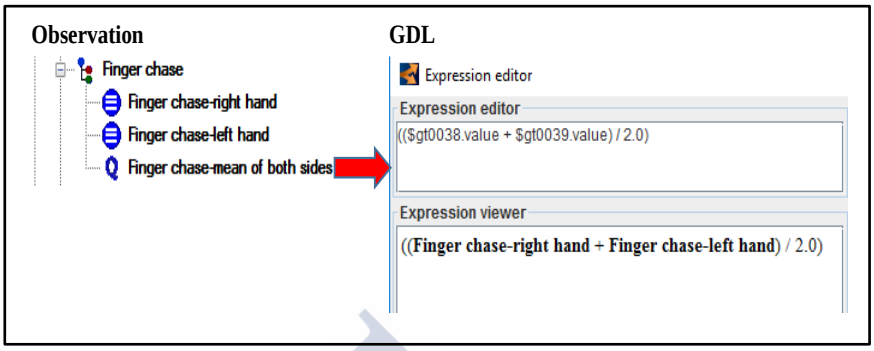

Fig. 3.10. Example of a calculation function. It is expressed by a GDL rule. The function calculates the mean of an "element" defined in the observation archetype (specifically, the element finger chase). The output of the rule refers to another element of the same observation archetype.

3.2.4.2 Assessment of the degree of impairment

The assessment of the degree of impairment grades the severity of the elements expressed into the observation archetype (Table 3.3), which is usually given by cut-off levels. Translating this cut-off levels to GDL required to develop sixty-one rules. For example, if the score of the Sitting item is 4, then the Sitting Instability is registered as 'severe' (Fig. 3.11).

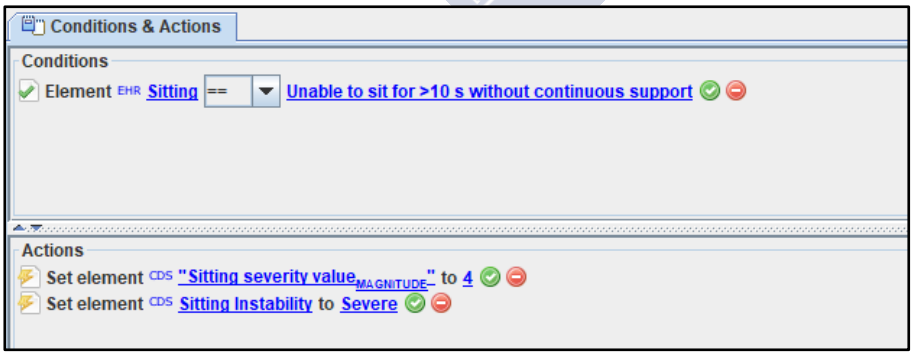

Fig. 3.11. GDL rule to assess the degree of sitting instability. The rule links the score 4 for the "sitting" item to the "severe sitting instability" class.

Table 3.3 SARA items severity levels.

SARA Items	Response Options								
	0	1	2	3	4	5	6	7	8
Gait	Normal	Borderline	Mild	Moderate	Moderate	Severe	Severe	Severe	Profound
Stance	Normal	Borderline	Mild	Mild	Moderate	Severe	Profound		
Sitting	Normal	Borderline	Mild	Moderate	Severe				
Speech disturbance	Normal	Borderline	Mild	Moderate	Moderate	Severe	Profound		
Finger Chase	Normal	Mild	Moderate	Severe	Severe				
Nose-Finger test	Normal	Mild	Moderate	Severe	Severe				
Fast alternating hand movements	Normal	Mild	Moderate	Severe	Severe				
Heel-shin slide	Normal	Mild	Moderate	Severe	Severe				

3.2.4.3 Assessment of cerebellar syndrome

The assessment of the cerebellar syndrome was based on the total score. The neurologist proposed several heuristic rules, which were modeled using GDL. Fig. 3.12 shows the rule inferring the absence of cerebellar syndrome. This knowledge was aligned with the results in [11], where the mean SARA score for controls was 0.4 ± 11.

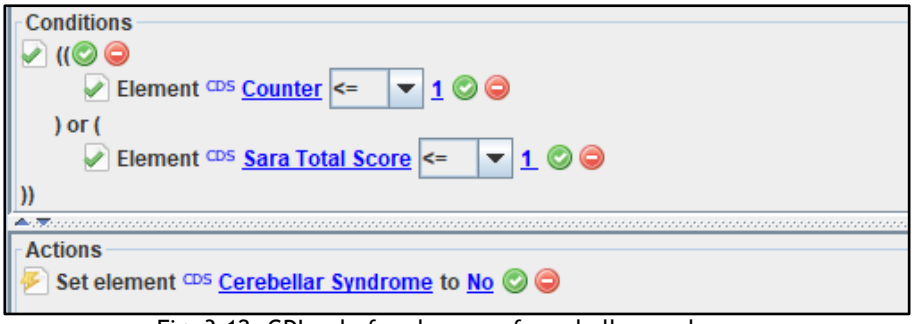

Fig. 3.12. GDL rule for absence of cerebellar syndrome

3.2.4.4 Rules interpreting the total score

1) If the total score is less than or equal to one, or the number of SARA items with values different than zero is less than or equal to one, then a patient does not suffer from a cerebellar syndrome.

2) If the total score is greater than one and less than three, then there is no a significant cerebellar syndrome.

3) If the total score is greater than or equal to three and less than or equal to eight, then there is a mild cerebellar syndrome.

4) If the total score is greater than eight and less than or equal to fifteen, then there is a moderate cerebellar syndrome.

5) If the total score is greater than fifteen, then there is severe a cerebellar syndrome.

3.2.4.5 Clinical synopsis

Finally, clinical synopsis outlines the phenotypic features (signs) accompanying the cerebellar syndrome. Examples of these signs are midline ataxia (truncal ataxia) or appendicular ataxia in Fig. 3.9. The first one affects the proximal musculature and it can be inferred by the ontology reasoner from the presence of the first three items gathered by the SARA (gait, stance and/or sitting) (Fig. 3.6). If one item among the first three items with a value greater than zero and the total score is greater than or equal to three, then midline ataxia exists, and its severity is determined by the highest severity of the three items. The second one affects movements of the extremities and it can be inferred from the last four items in the SARA (Finger chase, Nose-finger test, Fast alternating hand movements, and Heel-shin slide). If one item among the last four items with a value greater than zero and the total score is greater than or equal to three, then appendicular ataxia exists, and its severity is determined by the highest severity of the four items.

Clinical synopsis is inferred from the ontology modules partially based on the HPO. It requires using elements and values from the Evaluation archetype (such as, gait ataxia, sitting and standing instability) to infer other elements and values from the same archetype (such as, midline ataxia) using the OWL ontology. During this modeling phase, we only could simulate the ontology reasoning using the Protégé tool. From the values inferred by the GDL for the elements of the evaluation archetype, and taking into account the mappings of this archetype, we manually entered the individuals of the linked OWL classes and run the reasoner.

4 RESULTS

With the aim of testing the appropriateness of the methods presented in the previous chapter, we implemented a prototype of electronic rating scale for the SARA, called SARA Management System. The prototype can be used both for the assessment of cerebellar syndrome and for the production of a clinical synopsis. Additionally, we validated the approach in a real clinical setting, by using the recorded SARA data from 28 anonymous subjects affected by Spinocerebellar Ataxia Type 36 (SCA36). This chapter details the implemented prototype and, dataset and validation of the method and the results of the validation.

4.1 THE SARA MANAGEMENT SYSTEM

To demonstrate the functionality of our approach, we developed a framework entitled “SMS” (SARA Management System). SMS allows the management of patient data of cerebellar syndrome. We used the JAVA as a programming language, NetBeans as the integrated development environment, JAVA Swing as user interface toolkit, and MySQL as a database management system to store all the needed information.

4.1.1 SARA Management System Architecture

The architecture of “SMS” was structured in three layers: “Persistence”, “Operation” and “Interface”.

4.1.1.1 Persistence Layer

The persistence layer used MySQL to store the clinical archetypes, the patient input data, the data inferred by the system, and additional information. Fig. 4.1 shows the relational model of the database with three types of tables. The first type included the class,

subclass, scale, element and value tables. All of them modeled the attributes of the clinical archetypes (elements, values and mappings to ontology classes). To a greater extent, they were built based on the structure of the modeled archetypes. In order to populate the database with data from the archetypes, the ADL parser [13] provided by the OpenEHR Foundation was used to generate a dependency tree of terms and then create an XML file, which was aimed at producing archetype instances when needed. The second type of tables stored the patient data, and the last type (test and test_value tables) recorded the set of SARA tests.

4.1.1.2 Operation Layer

The operation layer included all information-processing units (Table 3.2) that are responsible to run Calculation Functions, Assessment of the Degree of Impairment. Assessment of Cerebellar Syndrome, and Clinical Synopsis. The current version of GDL does not provide a Java API or any mechanism for manipulating the rules defined using the editor. Hence, the only solution was to rewrite the rules in a rule engine (such as Drools[14] or Clips[15]) or to implement directly in Java (we decided this second solution). However deriving phenotype information was reasonably implemented using the OWL API [66] and the mappings linking archetype elements and values to OWL classes. The mappings were useful to automatically create OWL individuals from the archetype instances stored into the database.

[13] https://github.com/openEHR/java-libs/tree/master/adl-parser
[14] http://www.drools.org/
[15] http://www.clipsrules.net/

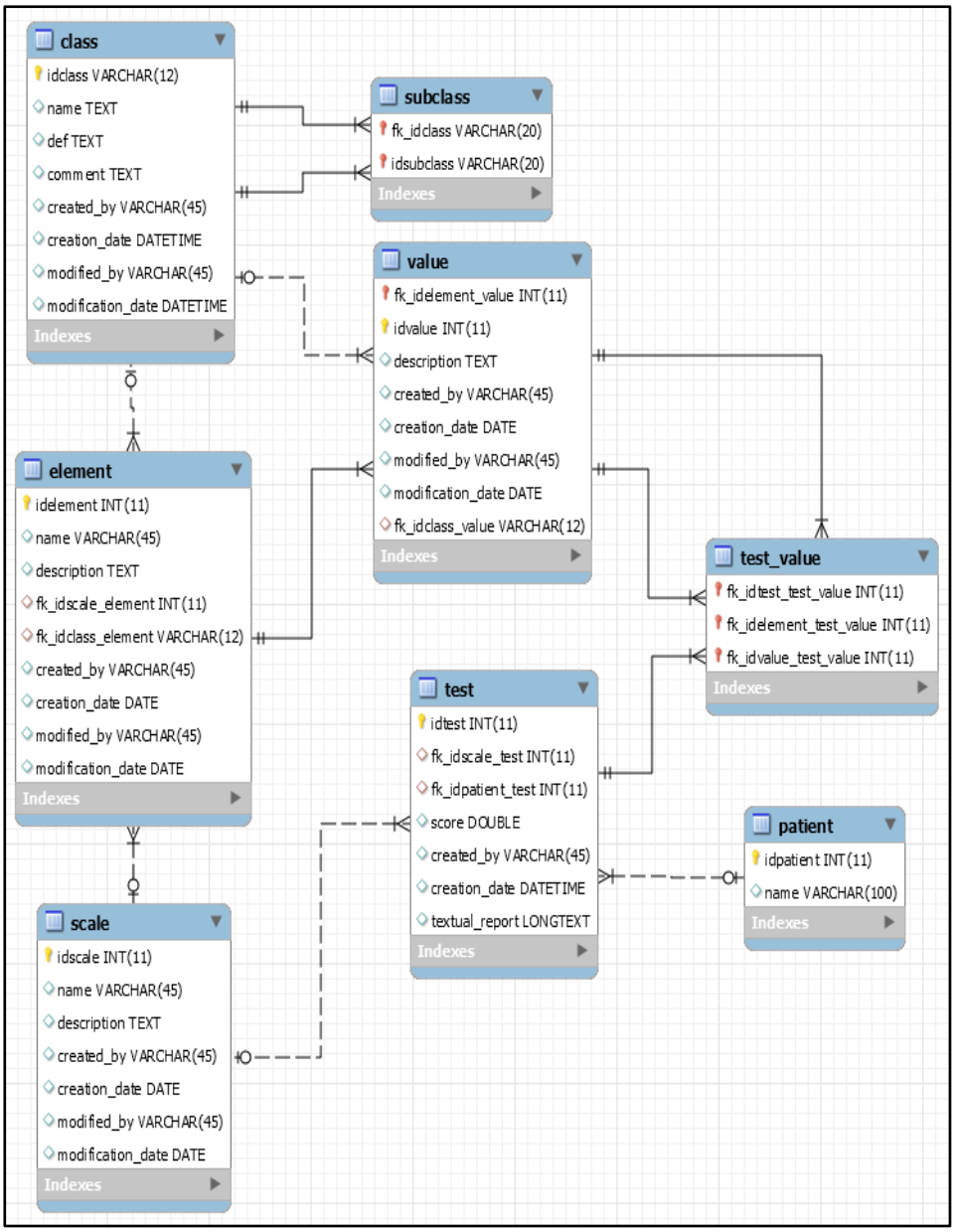

Fig. 4.1. Database Relational ModelThe table ‘patient’ recorded the patient identifier, the tables ‘test’ and ‘test_values’ stored the information about the tests covered by the SARA, and the rest of the tables recorded the information modeled in the clinical archetypes.

4.1.1.3 Interface Layer

On the other hand, the interface layer consisted of several input and output forms. The main form of the tool contained two menus: the ontology menu and the SARA menu. The first menu (Fig. 4.2) allowed users to modify the structure of the ontology, by adding, updating or deleting classes in the ontology . It also provided an option to automatically generate a hierarchical graph that displayed all the available classes and their is_a relationships. The SARA menu consisted of two sub-menus: 1) Observation, where neurologists could select patients and enter the 12 values of the assessed elements defined in the SARA scale and modeled in the observation archetype (Fig. 4.3); and 2) Evaluation, which provided three main features (Fig. 4.4):

- A table with all phenotypic abnormalities inferred from the collected patient data.

- A textual report summarizing the patient status. The report included the assessment of cerebellar syndrome and a brief synopsis of the phenotypic abnormalities accompanying the syndrome.

- An evaluation graph visualizing all the patient phenotypic abnormalities (Fig. 4.5). The green ellipses are the patient annotations based on the observation data and the yellow ones are the inferred abnormalities.

To facilitate semantic interoperable data exchange, the approach was designed to deliver the XML data collected by the SARA (using the schemas compliant with the observation archetype) and data inferred by the application (using the schemas compliant with the evaluation archetype). Additionally, the system provided the facility to send the SARA results and the patient report by e-mail. The doctor could attach it to the patient medical record.

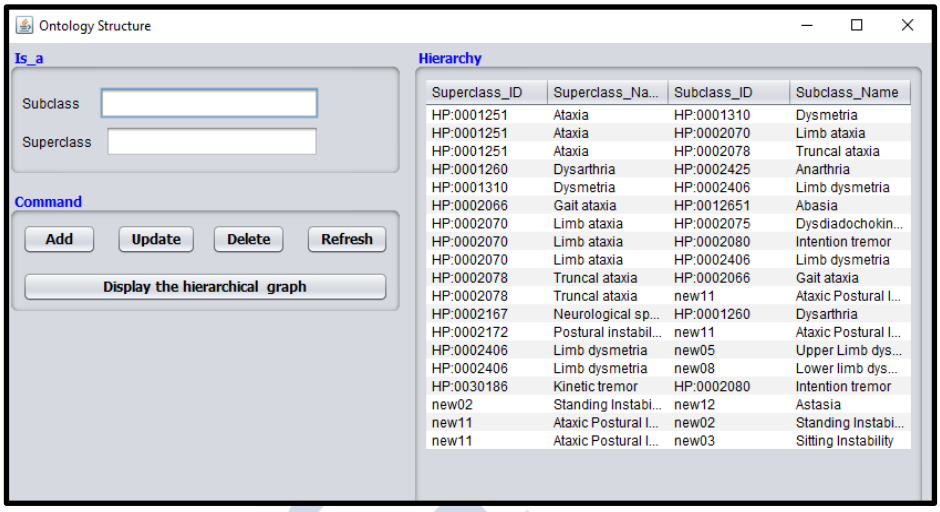

Fig. 4.2. Screenshot of ontology update. The form can be used for checking, remove and update classes of the reduced version of the HPO used by the SMS. All the classes can be viewed in both graphical and tabular formats.

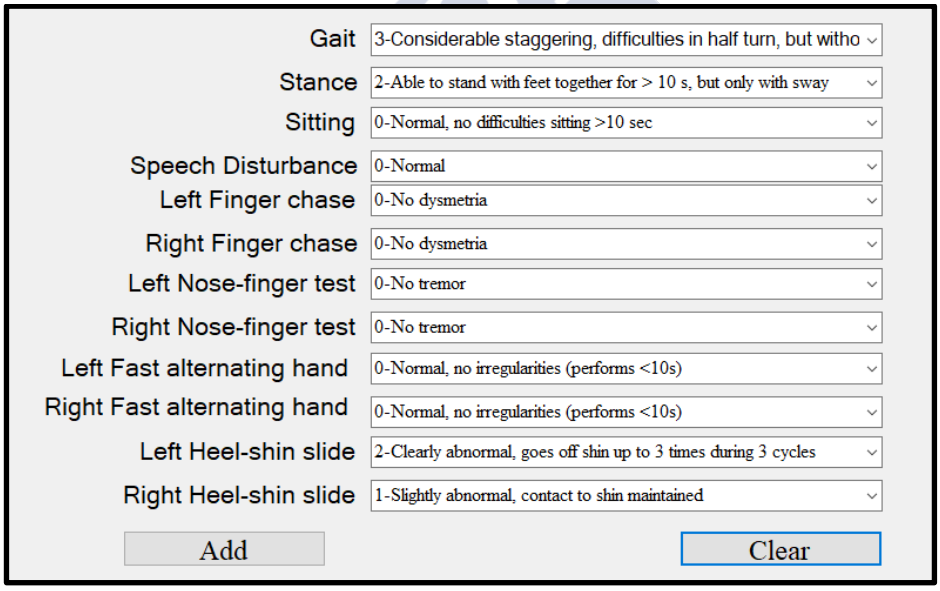

Fig. 4.3. Observation form. It allows neurologists to enter the values of the SARA scale items.

Patient: P9

Observation: 33-2017-05-01 20:49:23.0 Score=6.5

Phenotypic abnormalities codes			
Class id	Class name	Value	Severity
HP:002066	Gait Ataxia	3	Moderate
new02	Standing Instability	2	Mild
new08	Lower Limb Dysmetria	1.5	Mild

Display the Evaluation Graph

Patient status report

Date:2017-05-01 20:49:23.0
ID patient:9
The patient has a total SARA score of 6.5.
This corresponds to a MILD Cerebellar Syndrome, involving:
- Moderate truncal ataxia (Moderate gait impairment, Mild standing impairment, and Normal sitting instability).
- Appendicular ataxia:
  o Upper Limb Dysmetria: Normal bilaterally
  o Intention tremor: Normal bilaterally
  o Dysdiadochokinesis: Normal bilaterally
  o Lower Limb Dysmetria: Mild on the right side and Moderate on the left side

Fig. 4.4. Evaluation form.It displays an example of phenotypic abnormalities derived from the data of a patient (on the left side) and a textural report summarizing the status of this patient (on the right side).

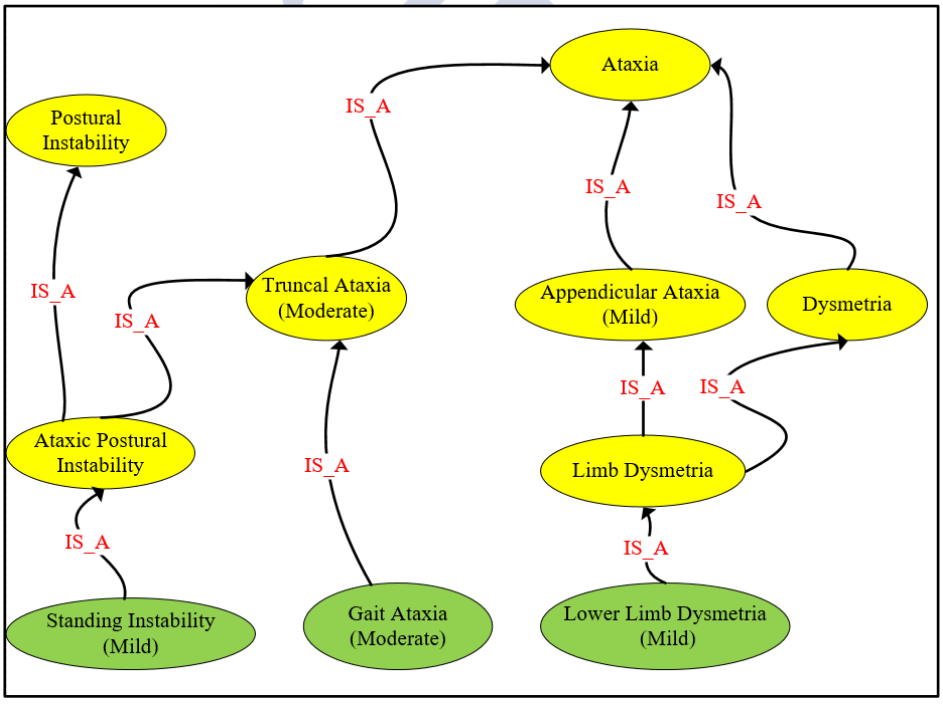

Fig. 4.5. Screenshot of a graphical summary. It displays a graph visualizing the set of phenotypic abnormalities inferred by the SMS. The green ellipses represent the lower classes in the hierarchy and the yellow ellipses, the superclasses.

4.2 DATASET AND VALIDATION OF THE METHOD

For validation purpose, we used data records from 28 anonymous subjects with Spinocerebellar Ataxia Type 36 (SCA36) [67]. All patients were examined following the SARA. Only the set of scores for each item collected by the SARA was taken into account during this evaluation. The institutional research ethics committee approved the recruitment and study protocol, and all participants gave their written informed consent. Two independent neurologists validated the feasibility of our approach. The evaluation was carried out in three steps:

1) Total score calculation and interpretation: We filled out the score data for each patient to get the interpretation. The system automatically inferred:

   - The severity for each item in the scale
   - The severity of cerebellar syndrome,
   - The severity of truncal ataxia
   - The severity of appendicular ataxia on the right and left sides.

2) Interpretation by two independent neurologists: The same data set was sent to two neurologists, but adding the calculated total scores for each patient. The neurologists used their expertise in ataxia to determine the severity of the cerebellar syndrome from the total score, and the severity of truncal and appendicular ataxia, if present, from the individual scores.

3) Comparison of results between the system and the human experts (neurologists): The validation process was carried out based on the results-oriented validation perspective [68]. This perspective is based on comparing the performance of the developed tool with an expected performance provided by human experts, in order to assess whether the tool produces the required output correctly. There are many methods of assessing inter-rater agreement. Specifically, Cohen's kappa [69] and Weighted kappa [70] have been widely applied in the

medical field [71]. Weighted Kappa is more compatible with ordinal scales [72-82], hence we decided to use it as the SARA is ordinal and the difference between severity levels is meaningful. To validate the system, the following steps were accomplished: 1) we created a sheet using SPSS[16], 2) we imported the interpretations of the system and the neurologists into SPSS, and 3) we ran Weighted Kappa test 12 times to measure the strength of agreement between the system and each neurologists, and between the two neurologists themselves.

4.3 VALIDATION OF THE SYSTEM

The validation of the system was based on the Weighted Kappa values obtained from the 12 tests conducted. Table 4.1 displays the results of measuring the strength of agreement between the system and the first neurologist, the system and the second neurologist and between the two neurologists themselves. Kappa values are between 0.65 and 0.93, as illustrated in Table 4.1.

Table 4.1 Agreement between automated and manual ratings (Weighted Kappa)

	System vs. 1st Neurologist Kappa value	System vs. 2nd Neurologist Kappa value	1st Neurologist vs. 2nd Neurologist Kappa value
Cerebellar syndrome	0.86	0.84	0.85
Midline Ataxia	0.80	0.84	0.86
Appendicular Ataxia (Right side)	0.71	0.80	0.86
Appendicular Ataxia (Left side)	0.62	0.78	0.84

[16] http://www.ibm.com/analytics/us/en/technology/spss/

5 DISCUSSION

In this doctoral thesis, a mixed method to support the development of the SARA has been presented. The method combined OpenEHR archetypes, guidelines, ontologies and reasoning. The innovation of our method rests on how these approaches were combined to get the full benefit of them. We distinguished between the modeling phase and the implementation phase. During the former, we addressed the calculation and assessment tasks by defining and executing GDL rules, and the clinical synopsis task by defining OWL classes and executing a reasoner. However, due to the lack of integration between these frameworks, we first ran the GDL framework, and then we manually entered the results in Protégé in order to infer the phenotypic abnormalities. During the implementation phase, we addressed the calculation and assessment tasks by rewriting the rules directly in Java, and the clinical synopsis task by integrating the OWL API into the system and using the mappings to create OWL individuals. We designed the approach as an archetype-based stand-alone application, providing a meaningful way for collecting and interpreting healthcare data. The application released the local EHR system of integrating the SARA, providing a standard way of delivering the collected and inferred data. Thus, the main role of this electronic rating scale was to collect the normalized data, execute the decision support logic and deliver both data and interpretations to the EHR system.

Turning to the research questions in this research, a few conclusions can be drawn. With respect to question (1) - Is the combination of GDL, openEHR clinical archetypes and ontologies suitable for the description of all knowledge covered by the SARA?) - , we can conclude that a combination of OpenEHR, GDL and OWL offers a suitable framework for the purpose of describing the data and knowledge levels of the SARA. OpenEHR provides a formal specification at the data level, whereas GDL and ontologies offer formal specifications of different types of knowledge for data

interpretation purpose. However, it should be emphasized that in our particular case, the knowledge level could be broken into separate information-processing units interconnected in a simple way through the two defined archetypes. However, the interpretation of a rating scale may require more complex control mechanisms, demanding more interoperability between GDL and OWL. Furthermore, the current version of GDL uses archetype data as input and output variables for all the rules, but it provides no facility to define auxiliary variables. This type of variables is sometimes necessary to model procedural knowledge, such as the counting of the scores in rating scales. At the moment, there are two solutions: 1) adding new elements to the archetype, or 2) defining an auxiliary archetype containing all the needed variables. The advantage of this second option is that leaves intact the original archetype. Based on the second solution, an auxiliary archetype was created in this work (Fig. 5.1).

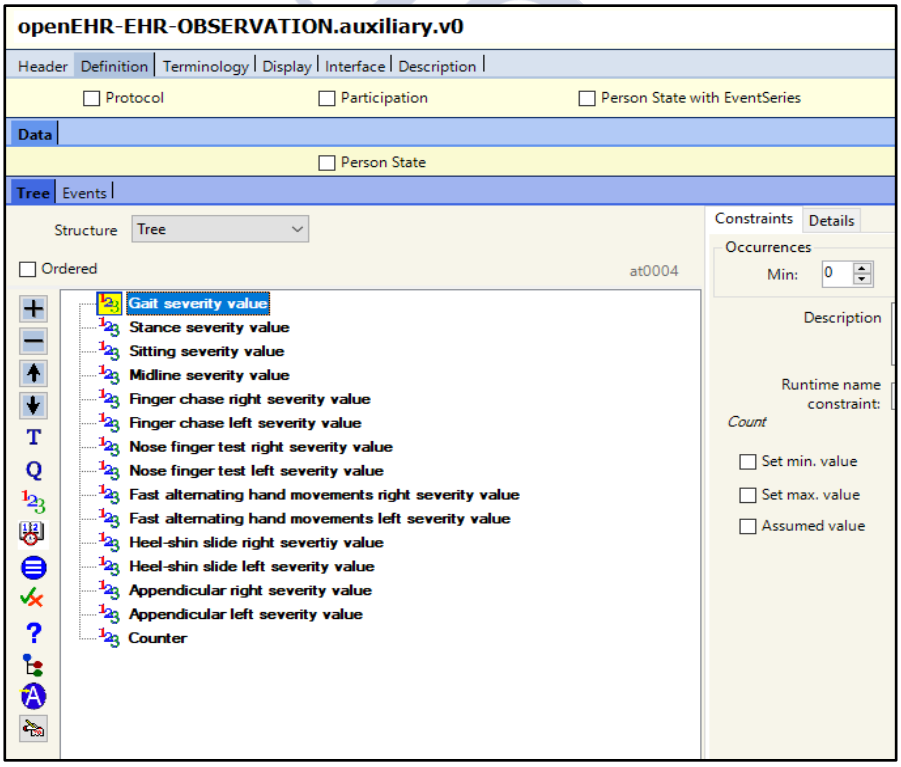

Fig. 5.1 Auxiliary archetype.

The new major version of ADL includes specifications for defining explicit rules of invariant assertions, i.e., expressions that should be satisfied by all instances of an archetype. These assertions cover some calculation functions over one or several items, and also definitions of mandatory items in the presence of specific values of other items. The definition of these rules provides the same functionality as some of the GDL rules defined in our system. However, the syntax of these specifications is not stable and it is still a need of tools that offer support for the automatic handling of invariants on archetype instances. Similarly, the new ADL specification covers a section for mapping to external terminologies, which has been improved with richer mappings. Specifically, it is possible to map post-coordinated archetype codes to ontology pre-coordinated classes. This facility can be very relevant when ontology mapping is carried out in the final stages of the modeling process.

With respect to question (2) – Is it possible to achieve a reasonable integration of these technologies to efficiently model the rating scale? , we showed that a full integration of these technologies to model the rating scale is not possible at the moment. In the modeling stage, the use of GDL facilitated the development and interconnection of most processing units, without resorting to external resources and encouraging knowledge sharing. We could verify and validate the SARA by testing use cases in the GDL editor. We should bear in mind that this tool automatically generates entry forms based on the defined archetypes (Fig. 5.2). The forms are used to collect data from the user, run the engine and display the outcomes (Fig. 5.3) . However, the editor does not supply any other facility for delivering the outcomes. For example, the generation of XML instances of the archetypes would be a remarkable advance to provide the option of combining the tool with other different inference engines, such as description logic reasoners. Regarding to ontology reasoning, testing use cases based on archetypes and OWL requires tools that automate the process of converting archetype instances to OWL individuals, run the reasoner on the OWL dataset, and deliver the outcomes as instances of the archetypes.

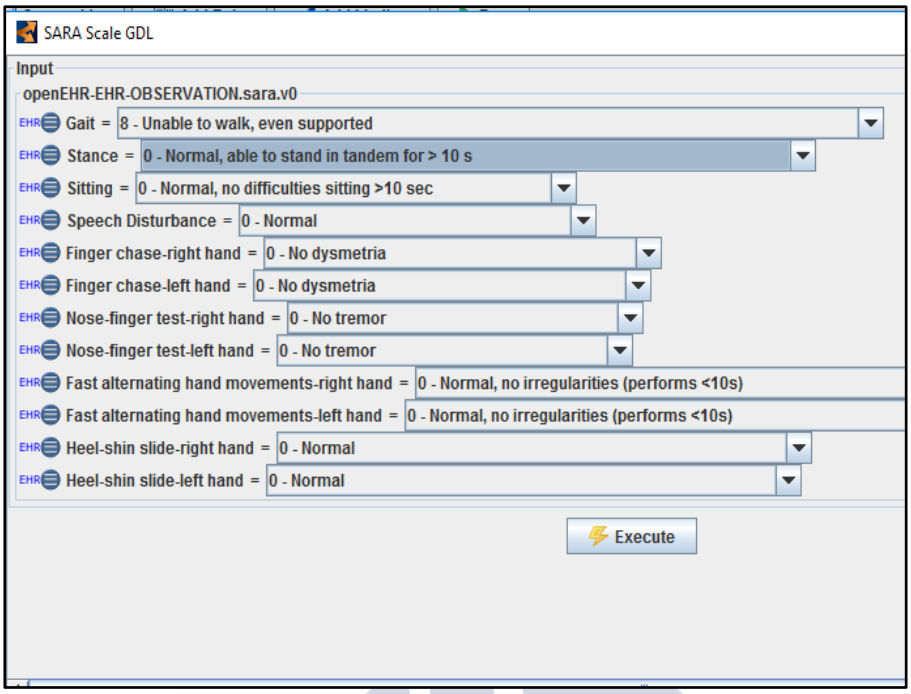

Fig. 5.2 Entry form generated by GDL editor.It displays the elements defined in the observation archetype and the response options of each element.

Fig. 5.3 The outcomes form generated by GDL.

In the implementation stage, the used version of GDL did not provide any utility to translate the modeled rules into some execution engine (e.g., drools or clips), as has been mentioned previously. Thus, this part of the implementation required substantial effort. In order to decrease the time devoted to implementation, we parsed the ADL archetypes and developed a database model based on the archetype structure. Regarding to ontology reasoning, the archetype mappings facilitated the translation of the archetype instances into OWL individuals. The suggested approach not only focuses on the syntactic structure of the SARA, but also on leveraging a reduced version of the HPO from the earliest stages of the modeling of archetypes. This ontology version was a valuable resource to facilitate

1) The syntactic structure of the rating scale.
2) The terminology mapping.
3) The automated interpretation of collected data.
4) The communication process among the information-processing units.

Regarding the first point, we organized the SARA items by means of a tree structure (Fig. 3.7 B), using the CLUSTERS class provided by OpenEHR. As mentioned above, this new organization preserved the 8-item performance of the original scale. It also differentiated the three main clinical dimensions of the SARA, although these were not assessed quantitatively. Following the OpenEHR documentation, the CLUSTERS class is provided to represent common domain patterns required in many clinical scenarios. The clinical dimensions identified into the SARA can be viewed as common domain patterns that provide a more accurate assessment of the patient's phenotype components, clarifying the interpretation of the results. However, the observation archetype [17] (Fig. 5.4) that was uploaded to the CKM, where is publicly accessible, follows the flat structure of the original rating scale. As the main goal of the CKM is to provide high-quality information models, the CKM consortium considered that a flat structure that complied with the original scale structure was more convenient. However, we think the approach presented here remains valid, as usually rating scales

[17] http://openehr.org/ckm/#showArchetype_1013.1.2661

grade several clinical dimensions [7] and the proposed structure using CLUSTERS classes allows the proper representation of these dimensions. On the other hand, the evaluation archetype was not uploaded to the CKM, as only those archetypes that are based on some documented international assessment or very generic requirement are accepted. Following the CKM recommendations, the SARA evaluation archetype is perfectly suitable for local use.

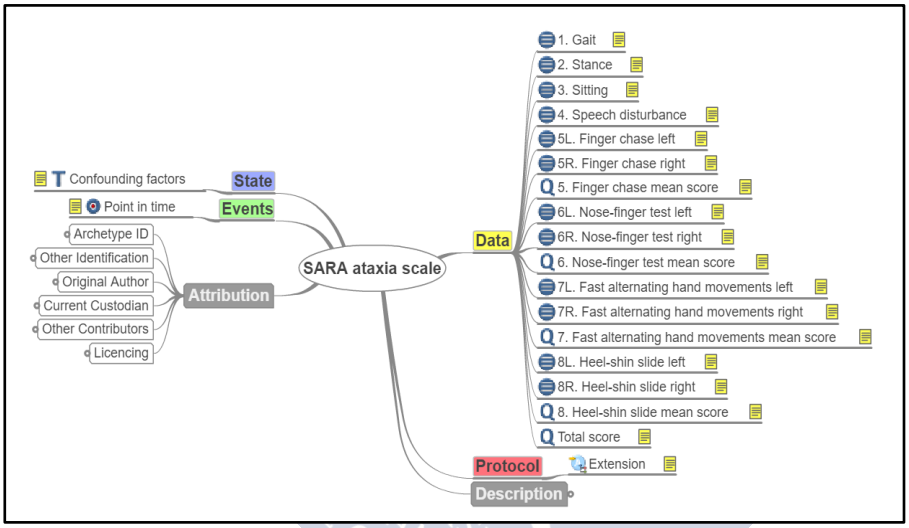

Fig. 5.4 The mind map representation of the uploaded SARA Observation archetype .It shows the flat structure of the original SARA scale.

Regarding the second point, mappings to standard vocabularies are uncommon in the clinical archetypes that are published in openly accessible repositories. In general, terminologies include a huge number of clinical terms; so manual mapping turns out to be unfeasible in practice. The extraction of the reduced version of the HPO provided us a means of performing terminology mapping in the earliest stages of archetype building. Just as for the clinical archetype, some parts (i.e., classes and relationships) of the reduced version of the HPO were reorganized to cover the SARA domain required for the ontology-driven modeling. This approach, known as ontology reuse, is an important design principle in ontologies [83, 84] that facilitates the development of specific applications.

Regarding the third point, the ontology version provided the knowledge required to infer patient phenotypic information from the data collection. For example, from the score 8 of the item gait, the system inferred that the patient had abasia, and so gait ataxia (Fig. 3.6). However, exploiting reasoning on both ADL and ontologies is not possible at the moment. In our approach, this reasoning was needed to interpret the presence of the phenotypic abnormalities associated to the clinical dimensions of the scale. As mentioned early, a critical success factor for exploiting reasoning is the availability of ontology-based reasoning tools that use data expressed in ADL format and with capabilities to fire GDL rules. Such an integrated editor would assist with the effort at the authoring level. On the other hand, following the approaches developed in [28, 33, 36], we will transform the clinical archetypes into OWL-DL and use the ontology and rule-based mechanisms provided by Protégé to draw interpretations on data collection, with the goal of comparing the results with the ones achieved the approach developed in this work.

With regard to the interpretation of the results of our application, our goal was to test whether the performance of the system reached limits considered as acceptable. The problem lies in defining appropriate limits. Landis and Koch [85] proposed the cut-off levels shown in Table 5.1 to interpret kappa statistic. Table 5.2 shows the interpretation of the results of Table 4.1, following the Landis & Koch criteria. The interpretations in Table 5.2 reflect a very high degree of agreement between the system and the two neurologists, confirming that the approach can be a good solution to develop electronic rating scales. Even so, these excellent results should also be viewed with much caution, as the validation was carried out only with 28 patient data, all of them affected by the same rare disease (SCA36).

Additionally, although the two neurologists who carried out the assessment were independent, they work in the same hospital and one of them is in the same research group as Maria Sobrido, the neurologist involved in the modeling process. It therefore has to be assumed that there exists consistency between the three neurologists. Therefore, in our future work, we will evaluate the application with a larger number of patient data that are affected by diverse cerebellar ataxias, and with the help of neurologists from different hospitals. If the results are still

highly satisfactory, we will develop a simple mobile application for the automatic transmission of the interpretation to the health information system.

Table 5.1. Kappa interpretation rules-Landis and Koch (1977)

Kappa Statistic	Strength of agreement
0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Table 5.2. Strength of agreement between automated and manual ratings. It follows the interpretation rules proposed by Landis and Koch.

	System vs 1st Neurologist	**System vs 2nd Neurologist**	**1st Neurologist vs 2nd Neurologist**
Cerebellar Syndrome	Almost Perfect	Almost Perfect	Almost Perfect
Truncal Ataxia (Midline Ataxia)	Substantial	Almost Perfect	Almost Perfect
Appendicular Ataxia (Right Side)	Substantial	Substantial	Almost Perfect
Appendicular Ataxia (Left Side)	Substantial	Substantial	Almost Perfect

Finally, although our approach was designed to implement a prototype for managing the SARA, it is rather generic and hence applicable to model other electronic rating scales, possibly in other

clinical domains. To take an example, the approach could be applied to the domain of the autism spectrum disorders, which exhibit complex phenotypes affecting variables that are difficult to measure. As a consequence, standardized scales are often used to collect a large amount of phenotypic data. Recently, a phenotype ontology has been developed to identify behavioral features of importance [86]. The availability of this ontology and also the mappings to the rating scales would facilitate the implementation of prototypes like the one presented here.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6 CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

1. Reducing all content of a rating scale to a unique number may lead to loss of useful clinical information about the dimensions implicitly collected by the scale. In this doctoral thesis, we developed a model to infer the full components of the patient's phenotype from the clinical dimensions represented by the rating scores. This model provides automated support for medical evaluation, report writing, and clinical decision-making. The proposed approach has been shown for the Scale for the Assessment and Rating of Ataxia (SARA), a well-validated instrument to evaluate the presence and severity of cerebellar ataxia.

2. Integrating electronic rating scales with the electronic health records and related systems requires formally describing these scales using standard clinical information models, such as openEHR. In this doctoral thesis, a novel combination of the best performances from OpenEHR clinical archetypes, guidelines and ontologies has been proposed to be able to reason on clinical archetypes. We showed for the specific field of ataxias, how clinical information models can be mapped to standard terminologies or ontologies, which provide the required meaning of their concepts.

3. The integration of phenotype ontologies with clinical archetypes provides not just a static knowledge store, but also a dynamic resource that allows automatic inference of a patient´s medical status (phenotype) from systematized collection of clinical data.

6.2 SPECIFIC CONTRIBUTIONS

This doctoral thesis work contributes to a better understanding of how clinical archetypes, guidelines and ontologies can be combined for modeling and implementing the SARA. There are several contributions in this research.

1. This research proposes an ontology-aware approach of clinical models, guidelines and terminologies to model electronic rating scales, where the ontology provides the backbone for normalizing the content of the scale through clinical archetypes.
2. The modeling approach distinctly clarifies the line of demarcation between the data level - representation of the scale items- and a knowledge level- referred to the strategy to compute the total score and the interpretation of patient phenotype. Archetypes facilitate the standard modeling of the data level, while GDL and OWL enable the standard modeling of the knowledge level.
3. The novel archetype development approach reduces the effort necessary for creating mappings, which is key to achieve semantic interoperability among different data sources. It also prevents large semantic discrepancies between the modeled archetypes and the ontology modules.
4. Additionally, a clear and explicit separation between the standard components of the scale related to the content (i.e., items, clinical dimensions and scores), and the clinical interpretations of these components are established.
5. Another key contribution was the clear identification of all different types of knowledge required to interpret the data collected by the scale.
6. The knowledge required to exploit reasoning on the scale data was modeled as separate information-processing units interconnected via the defined archetypes, providing a simple mechanism of combining ontology and rule-based reasoning.

7. A prototype named SARA Management System was developed to demonstrate the validity of the modeling approach. The prototype can be used for both the assessment of cerebellar syndrome and the production of a clinical synopsis.
8. The prototype was validated using recorded data from 28 anonymous subjects affected by Spinocerebellar Ataxia Type 36 (SCA36). The results reveal a substantial degree of agreement between the results achieved by the ontology-aware system and the human experts.

6.3 LIMITATIONS OF THE WORK

The innovation of our method rests on how clinical models, guidelines and terminologies were combined to get the full benefit of them. We have distinguished between the modeling phase and the implementation phase. During the former, we addressed the calculation and assessment tasks required by the scale by means of defining and executing GDL rules, and the clinical synopsis task by defining OWL classes and executing a reasoner. However, due to the lack of integration between GDL and OWL, we first ran the GDL framework, and then we manually entered the results in Protégé in order to infer the phenotypic abnormalities. This is clearly a limitation of the work, resulting from the current gaps in technology. Additionally, during the implementation phase, we addressed the calculation and assessment tasks by rewriting the rules directly in Java, and the clinical synopsis task by integrating the OWL API into the system and using the mappings to create OWL individuals. Once again, the inability of the current technology to automatically translate GDL rules to Drools or Clips rules to be integrated in a Java framework with the OWL API must be solved in the future work.

From the results achieved in this doctoral thesis, we have concluded that a combination of OpenEHR, GDL and OWL offers a suitable framework for the purpose of describing the data and knowledge levels of the SARA. However, it should be emphasized that in our particular case, the knowledge level could be broken into separate information-processing units interconnected in a simple way through the two defined archetypes (one for *observations* and another

for *evaluations*). However, the interpretation of a rating scale may require more complex control mechanisms, demanding more interoperability between GDL and OWL. Furthermore, the current version of GDL uses archetype data as input and output variables for all the rules, but it provides no facility to define auxiliary variables. This type of variables is sometimes necessary to model procedural knowledge, such as the counting of the scores in rating scales.

We showed that a full integration of the current technologies to model the rating scale is not possible at the moment. In the modeling stage, the use of GDL facilitated the development and interconnection of most processing units, without resorting to external resources and encouraging knowledge sharing. However, the current editor does not supply any facility for interoperability. For example, the generation of XML instances of the archetypes would be a remarkable advance to provide the option of combining the tool with other different inference engines, such as description logic reasoners.

Finally, the interpretation of the results of our prototype reflects a very high degree of agreement between the prototype and the experts, confirming that the approach can be a good solution to develop electronic rating scales. Even so, these excellent results should also be viewed with much caution, as the validation was carried out only with 28 patient data, all of them affected by the same rare disease (SCA36). Additionally, although the two neurologists who carried out the assessment were independent, they work in the same hospital and one of them is in the same research group as the neurologist involved in the modeling process. It therefore has to be assumed that there exists consistency between the three neurologists.

6.4 FUTURE WORK

Nowadays, the most ataxic disorders still have no successful pharmacological therapy, and patients suffer the unavoidable degenerative disease progression. The aim of well-validated rating scales is to understand better the natural history of ataxic disorders and evaluate properly drug efficacy in clinical trials. Rating scales facilitate clinical standardization of data collection, mainly in specialties with a richness of complex phenotypic variables, such as neurology.

However, the current electronic approaches are simple calculators with no integration with the electronic health records and related systems. In this doctoral thesis, a new solution to work towards this goal is provided. Exploiting reasoning on clinical archetypes represents a challenge

With the aim of achieving a full integration of the current technologies to model rating scales, we plan to evaluate the expressivity of the new major version of ADL. In particular, we plan to evaluate the specifications for defining explicit rules of invariant assertions. (i.e., expressions that should be satisfied by all instances of an archetype). If the definition of these rules provides the same functionality as GDL rules defined in our system, we will implement the facilities required to automatically execute these ADL rules and integrate with the OWL API. We will also evaluate Owlready2, a module for ontology-oriented programming in Python. We think that this module may provide the needed functionality for full integration.

Furthermore, in our future work, we will evaluate the SARA application with a larger number of patient data that are affected by diverse cerebellar ataxias, and with the help of neurologists from different hospitals. This new evaluation will provide us a stronger validation of our approach.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

REFERENCES

1. Robinson P N, Mungall C J, and Haendel M. *Capturing phenotypes for precision medicine.* Molecular Case Studies. 2015. **1**(1): p. a000372.
2. Baynam G, Walters M, Claes P, Kung S, LeSouef P, Dawkins H, Bellgard M, Girdea M, Brudno M, and Robinson P. *Phenotyping: targeting genotype's rich cousin for diagnosis.* Journal of paediatrics and child health. 2015. **51**(4): p. 381-386.
3. Grinnon S T, Miller K, Marler J R, Lu Y, Stout A, Odenkirchen J, and Kunitz S. *National institute of neurological disorders and stroke common data element project–approach and methods.* Clinical Trials. 2012. **9**(3): p. 322-329.
4. Köhler S, Vasilevsky N A, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello S M, Boerkoel C F, and Boycott K M. *The human phenotype ontology in 2017.* Nucleic acids research. 2017. **45**(D1): p. D865-D876.
5. Human Phenotype Ontology. 2017. http://human-phenotype-ontology.github.io. Accessed 13 April 2017.
6. Hobart J C, Cano S J, Zajicek J P, and Thompson A J. *Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations.* The Lancet Neurology. 2007. **6**(12): p. 1094-1105.
7. Martinez-Martin P. *Composite rating scales.* Journal of the Neurological Sciences. 2010. **289**(1): p. 7-11.
8. Pangman V C, Sloan J, and Guse L. *An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice.* Applied Nursing Research. 2000. **13**(4): p. 209-213.
9. Teasdale G, Maas A, Lecky F, Manley G, Stocchetti N, and Murray G. *The Glasgow Coma Scale at 40 years: standing the test of time.* The Lancet Neurology. 2014. **13**(8): p. 844-854.
10. Goetz C G, Tilley B C, Shaftman S R, Stebbins G T, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stern M B, and Dodel R. *Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS):*

Scale presentation and clinimetric testing results. Movement disorders. 2008. **23**(15): p. 2129-2170.
11. Schmitz-Hübsch T, Fimmers R, Rakowicz M, Rola R, Zdzienicka E, Fancellu R, Mariotti C, Linnemann C, Schöls L, and Timmann D. *Responsiveness of different rating instruments in spinocerebellar ataxia patients.* Neurology. 2010. **74**(8): p. 678-684.
12. Saute J A M, Donis K C, Serrano-Munuera C, Genis D, Ramirez L T, Mazzetti P, Pérez L V, Latorre P, Sequeiros J, and Matilla-Dueñas A. *Ataxia rating scales—psychometric profiles, natural history and their application in clinical trials.* The Cerebellum. 2012. **11**(2): p. 488-504.
13. Min H, Ohira R, Collins M A, Bondy J, Avis N E, Tchuvatkina O, Courtney P K, Moser R P, Shaikh A R, and Hesse B W. *Sharing behavioral data through a grid infrastructure using data standards.* Journal of the American Medical Informatics Association. 2014. **21**(4): p. 642-649.
14. Kalra D and Lloyd D. EN 13606 Health informatics - Electronic health record communication - Part 1: Reference model. European Committee for Standardisation. 2007. http://discovery.ucl.ac.uk/14026/. Accessed 15 January 2016.
15. HL7. HL7 Standards Product Brief - HL7 Implementation Guide for CDA® Release 2: IHE Health Story Consolidation, Release 1.1 - US Realm. 2014. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=258. Accessed 25 February 2016.
16. Beale T and Heard S. openEHR - Release 1.0.2. 2016. http://www.openehr.org/programs/specification/releases/1.0.2. Accessed 25 February 2016.
17. Oniki T A, Zhuo N, Beebe C E, Liu H, Coyle J F, Parker C G, Solbrig H R, Marchant K, Kaggal V C, and Chute C G. *Clinical element models in the SHARPn consortium.* Journal of the American Medical Informatics Association. 2015: p. ocv134.
18. openEHR-Foundation. Stroke Scale Neurological Assessment (NIHHS). http://www.openehr.org/ckm/#showArchetype_1013.1.2041. Accessed 25 February 2016.

19. openEHR-Foundation. Clinical Knowledge Manager. http://www.openehr.org/ckm. Accessed 25 February 2016.
20. Köhler S, Doelken S C, Mungall C J, Bauer S, Firth H V, Bailleul-Forestier I, Black G C, Brown D L, Brudno M, and Campbell J. *The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data.* Nucleic acids research. 2013: p. gkt1026.
21. Braun M, Brandt A U, Schulz S, and Boeker M. *Validating archetypes for the multiple sclerosis functional composite.* BMC medical informatics and decision making. 2014. **14**(1): p. 64.
22. Braun M. Timed 25-Foot Walk. http://www.openehr.org/ckm/#showArchetype_1013.1.1200. Accessed 1 March 2016.
23. Braun M. Nine Hole Peg Test. http://www.openehr.org/ckm/#showArchetype_1013.1.1202. Accessed 1 March 2017.
24. Braun M. Paced Auditory Serial Addition http://www.openehr.org/ckm/#showArchetype_1013.1.1296. Accessed 1 March 2016.
25. Braun M. MSFC Score. http://www.openehr.org/ckm/#showArchetype_1013.1.1368. Accessed 1 March 2016.
26. Meizoso M, Allones J L I, Hernández D M, and Iglesias M J T. *Semantic similarity-based alignment between clinical archetypes and SNOMED CT: an application to observations.* International journal of medical informatics. 2012. **81**(8): p. 566-578.
27. Berges I, Bermudez J, and Illarramendi A. *Binding SNOMED CT Terms to Archetype Elements.* Methods of information in medicine. 2015. **54**(1): p. 45-49.
28. Ellouze A S, Bouaziz R, and Ghorbel H. *Integrating semantic dimension into openEHR archetypes for the management of cerebral palsy electronic medical records.* Journal of Biomedical Informatics. 2016. **63**: p. 307-324.
29. Chen R and Corbal I. Guideline Definition Language (GDL). 2015. http://www.openehr.org/releases/CDS/latest/docs/GDL/GDL.

html#_guideline_definition_language_gdl. Accessed May 2016.
30. Anani N, Chen R, Moreira T P, and Koch S. *Retrospective checking of compliance with practice guidelines for acute stroke care: a novel experiment using openEHR's Guideline Definition Language.* BMC Medical Informatics and Decision Making. 2014. **14**(1): p. 39.
31. Lin C-H, Lo Y-C, Hung P-Y, and Liou D-M. *Building Chronic Kidney Disease Clinical Practice Guidelines Using the openEHR Guideline Definition Language.* Methods of Information in Medicine. 2016. **55**(6): p. 495-505.
32. Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis J T, and Maldonado J A. *A model-driven approach for representing clinical archetypes for Semantic Web environments.* Journal of biomedical informatics. 2009. **42**(1): p. 150-164.
33. Lezcano L, Sicilia M-A, and Rodríguez-Solano C. *Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules.* Journal of biomedical informatics. 2011. **44**(2): p. 343-353.
34. Tao C, Jiang G, Oniki T A, Freimuth R R, Zhu Q, Sharma D, Pathak J, Huff S M, and Chute C G. *A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data.* Journal of the American Medical Informatics Association. 2013. **20**(3): p. 554-562.
35. Menárguez-Tortosa M and Fernández-Breis J T. *OWL-based reasoning methods for validating archetypes.* Journal of biomedical informatics. 2013. **46**(2): p. 304-317.
36. Mugzach O, Peleg M, Bagley S C, Guter S J, Cook E H, and Altman R B. *An ontology for autism spectrum disorder (ASD) to infer ASD phenotypes from autism diagnostic interview-revised data.* Journal of biomedical informatics. 2015. **56**: p. 333-347.
37. Heard S and Beale T. Archetype Definition Language. . http://www.openehr.org/releases/1.0.1/architecture/am/adl2.pdf. Accessed 10 May 2016.
38. Marcos M, Maldonado J A, Martínez-Salvador B, Boscá D, and Robles M. *Interoperability of clinical decision-support systems and electronic health records using archetypes: a case study in*

clinical trial eligibility. Journal of biomedical informatics. 2013. **46**(4): p. 676-689.

39. Fernández-Breis J T, Maldonado J A, Marcos M, del Carmen Legaz-García M, Moner D, Torres-Sospedra J, Esteban-Gil A, Martínez-Salvador B, and Robles M. *Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts.* Journal of the American Medical Informatics Association. 2013. **20**(e2): p. e288-e296.
40. González-Ferrer A and Peleg M. *Understanding requirements of clinical data standards for developing interoperable knowledge-based DSS: a case study*. Computer Standards & Interfaces. 2015. **42**: p. 125-136.
41. Van Swieten J, Koudstaal P, Visser M, Schouten H, and Van Gijn J. *Interobserver agreement for the assessment of handicap in stroke patients.* Stroke. 1988. **19**(5): p. 604-607 DOI: https://doi.org/10.1161/01.STR.19.5.604.
42. Beale T. *Archetypes: Constraint-based domain models for future-proof information systems*. in *OOPSLA 2002 workshop on behavioural semantics*. 2002.
43. Leslie H. Introduction to Archetypes and Archetype classes. 2012. https://openehr.atlassian.net/wiki/display/healthmod/Introduction+to+Archetypes+and+Archetype+classes#suk=. Accessed 20 January 2017.
44. Qamar R, *Semantic mapping of clinical model data to biomedical terminologies to facilitate interoperability*. 2008, Citeseer.
45. Heard S. Barthel index. http://openehr.org/ckm/#showArchetype_1013.1.128. Accessed May 2016.
46. Beale T and Heard S. *openEHR architecture overview. openEHR Foundation.* London, UK. 2008.
47. Leslie H. Glasgow Coma Scale. http://openehr.org/ckm/#showArchetype_1013.1.137. Accessed 15 June 2016.
48. Demski H, Garde S, and Hildebrand C. *Open data models for smart health interconnected applications: the example of*

openEHR. BMC medical informatics and decision making. 2016. **16**(1): p. 137.
49. Pediatrics A A o. *The APGAR score*. Advances in Neonatal Care. 2006. **6**(4): p. 220-223.
50. Sam H. Apgar score http://www.openehr.org/ckm/#showArchetype_1013.1.172. Accessed 1 March 2016.
51. Mungall C and Ireland A. The OBO Flat File Format Guide, version 1.4. http://owlcollab.github.io/oboformat/doc/GO.format.obo-1_4.html. Accessed 15 December 2015.
52. Chen R and Corbal I. *Guideline Definition Language (GDL)*. Release 0.9. 2013: p. 1-23.
53. Coppens M, Eikelboom J W, Hart R G, Yusuf S, Lip G Y, Dorian P, Shestakovska O, and Connolly S J. *The CHA2DS2-VASc score identifies those patients with atrial fibrillation and a CHADS2 score of 1 who are unlikely to benefit from oral anticoagulant therapy*. European heart journal. 2012. **34**(3): p. 170-176.
54. OWL W C. OWL 2 Web Ontology Language Document Overview (Second Edition). 2017. http://www.w3.org/TR/owl2-overview/. Accessed.
55. Horridge M, *A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools Edition 1.3*. 2011.
56. Schmitz-Hübsch T, Du Montcel S T, Baliko L, Berciano J, Boesch S, Depondt C, Giunti P, Globas C, Infante J, and Kang J-S. *Scale for the assessment and rating of ataxia Development of a new clinical scale*. Neurology. 2006. **66**(11): p. 1717-1720.
57. Schöls L, Bauer P, Schmidt T, Schulte T, and Riess O. *Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis*. The Lancet Neurology. 2004. **3**(5): p. 291-304.
58. Sequeiros J, Martins S, and Silveira I. *Epidemiology and population genetics of degenerative ataxias*. Handbook of clinical neurology. 2011. **103**: p. 227-251.
59. García-Murias M, Quintáns B, Arias M, Seixas A I, Cacheiro P, Tarrío R, Pardo J, Millán M J, Arias-Rivas S, and Blanco-

Arias P. *'Costa da Morte'ataxia is spinocerebellar ataxia 36: clinical and genetic characterization.* Brain. 2012: p. aws069.
60. Arias M, Quintáns B, García-Murias M, and Sobrido M J. *Spinocerebellar ataxia type 36.* 2014.
61. Ikeda Y, Ohta Y, Kurata T, Shiro Y, Takao Y, and Abe K. *Acoustic impairment is a distinguishable clinical feature of Asidan/SCA36.* Journal of the neurological sciences. 2013. **324**(1): p. 109-112.
62. National Cancer Institute: Brain Tumors. 2017. https://www.cancer.gov/publications/patient-education/brain.pdf. Accessed November 2017.
63. Taboada M, Rodríguez H, Martínez D, Pardo M, and Sobrido M J. *Automated semantic annotation of rare disease cases: a case study.* Database. 2014. **2014**: p. bau045.
64. Seidenberg J and Rector A. *Web ontology segmentation: analysis, classification and use*. in *Proceedings of the 15th international conference on World Wide Web*. 2006. ACM.
65. Stuckenschmidt H, Parent C, and Spaccapietra S, *Modular ontologies: concepts, theories and techniques for knowledge modularization*. Vol. 5445. 2009: Springer.
66. Horridge M and Bechhofer S. *The owl api: A java api for owl ontologies.* Semantic Web. 2011. **2**(1): p. 11-21.
67. Aguiar P, Pardo J, Arias M, Quintáns B, Fernández-Prieto M, Martínez-Regueiro R, Pumar J M, Silva-Rodríguez J, Ruibal Á, and Sobrido M J. *PET and MRI detection of early and progressive neurodegeneration in spinocerebellar ataxia type 36.* Movement Disorders. 2016.
68. Lee S and O'Keefe R M. *Developing a strategy for expert system verification and validation.* IEEE Transactions on Systems, Man, and Cybernetics. 1994. **24**(4): p. 643-655.
69. Cohen J. *A coefficient of agreement for nominal scales.* Educational and psychological measurement. 1960. **20**(1): p. 37-46.
70. Cohen J. *Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.* Psychological bulletin. 1968. **70**(4): p. 213.

71. Viera A J and Garrett J M. *Understanding interobserver agreement: the kappa statistic.* Fam Med. 2005. **37**(5): p. 360-363.
72. Mielke Jr P W, Berry K J, and Johnston J E. *Unweighted and weighted kappa as measures of agreement for multiple judges.* International Journal of Management. 2009. **26**(2): p. 213.
73. Cicchetti D V. *Testing the normal approximation and minimal sample size requirements of weighted kappa when the number of categories is large.* Applied Psychological Measurement. 1981. **5**(1): p. 101-104.
74. Kramer M S and Feinstein A R. *Clinical biostatistics: LIV. The biostatistics of concordance.* Clinical Pharmacology & Therapeutics. 1981. **29**(1): p. 111-123.
75. Banerjee M, Capozzoli M, McSweeney L, and Sinha D. *Beyond kappa: A review of interrater agreement measures.* Canadian journal of statistics. 1999. **27**(1): p. 3-23.
76. Kingman A. *Beyond weighted kappa when evaluating examiner agreement for ordinal responses*. in *Journal of Dental Research*. 2002. INT AMER ASSOC DENTAL RESEARCHI ADR/AADR 1619 DUKE ST, ALEXANDRIA, VA 22314-3406 USA.
77. Ludbrook J. *Statistical techniques for comparing measurers and methods of measurement: a critical review.* Clinical and Experimental Pharmacology and Physiology. 2002. **29**(7): p. 527-536.
78. Perkins S M and Becker M P. *Assessing rater agreement using marginal association models.* Statistics in medicine. 2002. **21**(12): p. 1743-1760.
79. Fleiss J L, Levin B, and Paik M C. *Poisson regression.* Statistical Methods for Rates and Proportions, Third Edition. 2003: p. 340-372.
80. Kundel H L and Polansky M. *Measurement of observer agreement 1.* Radiology. 2003. **228**(2): p. 303-308.
81. Schuster C. *A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales.* Educational and Psychological Measurement. 2004. **64**(2): p. 243-253.

82. Berry K J, Johnston J E, and Mielke Jr P W. *Exact and resampling probability values for weighted kappa.* Psychological reports. 2005. **96**(2): p. 243-252.
83. Pathak J, Johnson T M, and Chute C G. *Survey of modular ontology techniques and their applications in the biomedical domain.* Integrated computer-aided engineering. 2009. **16**(3): p. 225-242.
84. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg L J, Eilbeck K, Ireland A, and Mungall C J. *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration.* Nature biotechnology. 2007. **25**(11): p. 1251-1255.
85. Landis J R and Koch G G. *The measurement of observer agreement for categorical data.* biometrics. 1977: p. 159-174.
86. McCray A T, Trevvett P, and Frost H R. *Modeling the autism spectrum disorder phenotype.* Neuroinformatics. 2014. **12**(2): p. 291-305.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

LIST OF FIGURES

LIST OF TABLES

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APPENDIX A: The eight items included in the SARA scale

Gait item	
A patient is asked (1) to walk at a safe distance parallel to a wall including a half-turn (turn around to face the opposite direction of gait) and (2) to walk in tandem (heels to toes) without support.	
Value	**Description**
0	Normal, no difficulties in walking, turning and walking tandem (up to one misstep allowed)
1	Slight difficulties, only visible when walking 10 consecutive steps in tandem
2	Clearly abnormal, tandem walking >10 steps not possible
3	Considerable staggering, difficulties in half-turn, but without support
4	Marked staggering, intermittent support of the wall required
5	Severe staggering, permanent support of one stick or light support by one arm required
6	Walking > 10 m only with strong support (two special sticks or stroller or accompanying person)
7	Walking < 10 m only with strong support (two special sticks or stroller or accompanying person)
8	Unable to walk, even supported

Stance item	
A patient is asked to stand (1) in natural position, (2) with feet together in parallel (big toes touching each other) and (3) in tandem (both feet on one line, no space between heel and toe). Proband does not wear shoes, eyes are open. For each condition, three trials are allowed. Best trial is rated.	
Value	**Description**
0	Normal, able to stand in tandem for > 10 s
1	Able to stand with feet together without sway, but not in tandem for > 10s
2	Able to stand with feet together for > 10 s, but only with sway
3	Able to stand for > 10 s without support in natural position, but not with feet together
4	Able to stand for >10 s in natural position only with intermittent support
5	Able to stand >10 s in natural position only with constant support of one arm
6	Unable to stand for >10 s even with constant support of one arm

Sitting item	
A patient is asked to sit on an examination bed without support of feet, eyes open and arms outstretched to the front.	
Value	**Description**
0	Normal, no difficulties sitting >10 sec
1	Slight difficulties, intermittent sway
2	Constant sway, but able to sit > 10 s without support
3	Able to sit for > 10 s only with intermittent support
4	Unable to sit for >10 s without continuous support

Speech disturbance item	
Speech is assessed during normal conversation	
Value	**Description**
0	Normal
1	Suggestion of speech disturbance
2	Impaired speech, but easy to understand
3	Occasional words difficult to understand
4	Many words difficult to understand
5	Only single words understandable
6	Speech unintelligible / Anarthria

Finger chase item	
A patient sits comfortably. If necessary, support of feet and trunk is allowed. Examiner sits in front of proband and performs 5 consecutive sudden and fast pointing movements in unpredictable directions in a frontal plane, at about 50 % of proband´s reach. Movements have an amplitude of 30 cm and a frequency of 1 movement every 2 s. Proband is asked to follow the movements with his index finger, as fast and precisely as possible. Average performance of last 3 movements is rated. The right and left sides are rated independently, then the mean of both sides is calculated.	
Value	**Description**
0	No dysmetria
1	Dysmetria, under/ overshooting target <5 cm
2	Dysmetria, under/ overshooting target < 15 cm
3	Dysmetria, under/ overshooting target > 15 cm
4	Unable to perform 5 pointing movements

Nose-finger test sara item	
A patient sits comfortably. If necessary, support of feet and trunk is allowed. Proband is asked to point repeatedly with his index finger from his nose to examiner's finger which is in front of the proband at about 90 % of proband's reach. Movements are performed at moderate speed. Average performance of movements is rated according to the amplitude of the kinetic tremor. The right and left sides are rated independently, then the mean of both sides is calculated.	
Value	**Description**
0	No tremor
1	Tremor with an amplitude < 2 cm
2	Tremor with an amplitude < 5 cm
3	Tremor with an amplitude > 5 cm
4	Unable to perform 5 pointing movements

Fast alternating hand movements item	
A patient sits comfortably. If necessary, support of feet and trunk is allowed. Proband is asked to perform 10 cycles of repetitive alternation of pro- and supinations of the hand on his/her thigh as fast and as precise as possible. Movement is demonstrated by examiner at a speed of approx. 10 cycles within 7s. Exact times for movement execution have to be taken. The right and left sides are rated independently, then the mean of both sides is calculated.	
Value	**Description**
0	Normal, no irregularities (performs <10s)
1	Slightly irregular (performs <10s)
2	Clearly irregular, single movements difficult to distinguish or relevant interruptions, but performs <10s
3	Very irregular, single movements difficult to distinguish or relevant interruptions, performs >10s
4	Unable to complete 10 cycles

Heel-shin slide item	
A patient lies on examination bed, without sight of his legs. Proband is asked to lift one leg, point with the heel to the opposite knee, slide down along the shin to the ankle, and lay the leg back on the examination bed. The task is performed 3 times. Slide-down movements should be performed within 1 s. If proband slides down without contact to shin in all three trials, rate 4. The right and left sides are rated independently, then the mean of both sides is calculated.	
Value	**Description**
0	Normal
1	Slightly abnormal, contact to shin maintained
2	Clearly abnormal, goes off shin up to 3 times during 3 cycles
3	Severely abnormal, goes off shin 4 or more times during 3 cycles
4	Unable to perform the task

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APPENDIX B. List of all rules modeled in the developed GDL document.

Rule Initialization
When
Element **Gait severity value** does not exist
Element **Stance severity value** does not exist
Element **Sitting severity value** does not exist
Element **Midline severity value** does not exist
Element **Finger chase right severity value** does not exist
Element **Finger chase left severity value** does not exist
Element **Nose finger test right severity value** does not exist
Element **Nose finger test left severity value** does not exist
Element **Fast alternating hand movements right severity value** does not exist
Element **Fast alternating hand movements left severity value** does not exist
Element **Heel-shin slide right severity value** does not exist
Element **Heel-shin slide left severity value** does not exist
Element **Appendicular right severity value** does not exist
Element **Appendicular left severity value** does not exist
Element **Counter** does not exist
Then
Set element **Gait severity value** to **0**
Set element **Stance severity value** to **0**
Set element **Sitting severity value** to **0**
Set element **Midline severity value** to **0**
Set element **Finger chase right severity value** to **0**
Set element **Finger chase left severity value** to **0**
Set element **Nose finger test right severity value** to **0**
Set element **Nose finger test left severity value** to **0**
Set element **Fast alternating hand movements right severity value** to **0**
Set element **Fast alternating hand movements left severity value** to **0**
Set element **Heel-shin slide right severity value** to **0**
Set element **Heel-shin slide left severity value** to **0**
Set element **Appendicular right severity value** to **0**

Set element **Appendicular left severity value** to **0** Set element **Counter** to **0**
Rule Counter check gait **When** Element **"Gait$_{VALUE}$"** is greater than 0 Element **Gait** does not exist **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Gait** to **Gait**
Rule Counter check stance **When** Element **Stance** does not exist Element **"Stance$_{VALUE}$"** is greater than 0 **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Stance** to **Stance**
Rule Counter check sitting **When** Element **"Sitting$_{VALUE}$"** is greater than 0 Element **Sitting** does not exist **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Sitting** to **Sitting**
Rule Counter check speech disturbance **When** Element **Speech Disturbance** does not exist Element **"Speech Disturbance$_{VALUE}$"** is greater than 0 **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Speech Disturbance** to **Speech Disturbance**
Rule Counter check finger chase right **When** Element **"Finger chase-right hand$_{VALUE}$"** is greater than 0 Element **Finger chase-right hand** does not exist **Then** Set element **Finger chase-right hand** to **Finger chase-right hand** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1)

Rule Counter check finger chase left **When** Element **Finger chase-left hand** does not exist Element **"Finger chase-left hand$_{VALUE}$"** is greater than 0 **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Finger chase-left hand** to **Finger chase-left hand**
Rule Counter Nose finger test right **When** Element **Nose-finger test-right hand** does not exist Element **"Nose-finger test-right hand$_{VALUE}$"** is greater than 0 **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Nose-finger test-right hand** to **Nose-finger test-right hand**
Rule Counter Nose finger test left **When** Element **Nose-finger test-left hand** does not exist Element **"Nose-finger test-left hand$_{VALUE}$"** is greater than 0 **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Nose-finger test-left hand** to **Nose-finger test-left hand**
Rule Counter check fast alternating hand movements right **When** Element **Fast alternating hand movements-right hand** does not exist Element **"Fast alternating hand movements-right hand$_{VALUE}$"** is greater than 0 **Then** Set element **Fast alternating hand movements-right hand** to **Fast alternating hand movements-right hand** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1)
Rule Counter check fast alternating hand movements left **When** Element **Fast alternating hand movements-left hand** does not exist Element **"Fast alternating hand movements-left hand$_{VALUE}$"** is

greater than 0 **Then** Set element **Fast alternating hand movements-left hand** to **Fast alternating hand movements-left hand** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1)
Rule Counter check heel shin slide right **When** Element **Heel-shin slide-right hand** does not exist Element **"Heel-shin slide-right hand$_{VALUE}$"** is greater than 0 **Then** Set element **Heel-shin slide-right hand** to **Heel-shin slide-right hand** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1)
Rule Counter check heel shin slide left **When** Element **"Heel-shin slide-left hand$_{VALUE}$"** is greater than 0 Element **Heel-shin slide-left hand** does not exist **Then** Set element **"Counter$_{MAGNITUDE}$"** to (**Counter** + 1) Set element **Heel-shin slide-left hand** to **Heel-shin slide-left hand**
Rule Calculate finger chase mean **When** **Then** Set element **"Finger chase-mean of both sides$_{MAGNITUDE}$"** to ((**Finger chase-right hand** + **Finger chase-left hand**) / 2.0) Set element **"Finger chase-mean of both sides$_{PRECISION}$"** to 1
Rule Calculate nose finger test mean **When** **Then** Set element **"Nose-finger test-mean of both sides$_{PRECISION}$"** to 1 Set element **"Nose-finger test-mean of both sides$_{MAGNITUDE}$"** to ((**Nose-finger test-right hand** + **Nose-finger test-left hand**) / 2.0)
Rule Calculate fast alternating hand movements mean **When** **Then** Set element **"Fast alternating hand movements-mean of both sides$_{PRECISION}$"** to 1

Set element **"Fast alternating hand movements-mean of both sides$_{MAGNITUDE}$"** to ((**Fast alternating hand movements-right hand** + **Fast alternating hand movements-left hand**) / 2.0)
Rule Calculate Heel-shin mean **When** **Then** Set element **"Heel-shin slide-mean of both sides$_{PRECISION}$"** to 1 Set element **"Heel-shin slide-mean of both sides$_{MAGNITUDE}$"** to ((**Heel-shin slide-right hand** + **Heel-shin slide-left hand**) / 2.0)
Rule Calculate total score **When** **Then** Set element **"Sara Total Score$_{MAGNITUDE}$"** to (((((((**Gait** + **Stance**) + **Sitting**) + **Speech Disturbance**) + **Finger chase-mean of both sides**) + **Nose-finger test-mean of both sides**) + **Fast alternating hand movements-mean of both sides**) + **Heel-shin slide-mean of both sides**) Set element **"Sara Total Score$_{PRECISION}$"** to 1
Rule Gait normal **When** Element **Gait** equals to **Normal, no difficulties in walking, turning and walking tandem (up to one misstep allowed)** **Then** Set element **Gait** to **Gait** Set element **Gait Ataxia** to **Normal**
Rule Gait Borderline **When** Element **Gait** equals to **Slight difficulties, only visible when walking 10 consecutive steps in tandem** **Then** Set element **"Gait severity value$_{MAGNITUDE}$"** to 1 Set element **Gait Ataxia** to **Borderline**
Rule Gait mild **When** Element **Gait** equals to **Clearly abnormal, tandem walking >10 steps not possible** **Then**

Set element **"Gait severity value$_{MAGNITUDE}$"** to 2 Set element **Gait Ataxia** to **Mild**
Rule Gait moderate **When** ((Element **Gait** equals to **Considerable staggering, difficulties in half-turn, but without support**) **or** (Element **Gait** equals to **Marked staggering, intermittent support of the wall required**)) **Then** Set element **"Gait severity value$_{MAGNITUDE}$"** to 3 Set element **Gait Ataxia** to **Moderate**
Rule Gait severe **When** ((Element **Gait** equals to **Severe staggering, permanent support of one stick or light support by one arm required**) **or** (((Element **Gait** equals to **Walking > 10 m only with strong support (two special sticks or stroller or accompanying person)**) **or** (Element **Gait** equals to **Walking < 10 m only with strong support (two special sticks or stroller or accompanying person)**)))) **Then** Set element **Gait Ataxia** to **Severe** Set element **"Gait severity value$_{MAGNITUDE}$"** to 4
Rule Gait profound **When** Element **Gait** equals to **Unable to walk, even supported** **Then** Set element **Gait Ataxia** to **profound** Set element **"Gait severity value$_{MAGNITUDE}$"** to 5

Rule Stance normal **When** Element **Stance** equals to **Normal, able to stand in tandem for > 10 s** **Then** Set element **Stance** to **Stance** Set element **Standing Instability** to **Normal**
Rule Stance borderline **When** Element **Stance** equals to **Able to stand with feet together without sway, but not in tandem for > 10s** **Then** Set element **"Stance severity value$_{MAGNITUDE}$"** to 1 Set element **Standing Instability** to **Borderline**
Rule Stance mild **When** ((Element **Stance** equals to **Able to stand with feet together for > 10 s, but only with sway**) **or** (Element **Stance** equals to **Able to stand for > 10 s without support in natural position, but not with feet together**)) **Then** Set element **"Stance severity value$_{MAGNITUDE}$"** to 2 Set element **Standing Instability** to **Mild**
Rule Stance moderate **When** Element **Stance** equals to **Able to stand for >10 s in natural position only with intermittent support** **Then** Set element **"Stance severity value$_{MAGNITUDE}$"** to 3 Set element **Standing Instability** to **Moderate**
Rule Stance severe **When** Element **Stance** equals to **Able to stand >10 s in natural position only with constant support of one arm** **Then**

Set element **"Stance severity value$_{MAGNITUDE}$"** to 4 Set element **Standing Instability** to **Severe**
Rule Stance profound **When** Element **Stance** equals to **Unable to stand for >10 s even with constant support of one arm** **Then** Set element **Standing Instability** to **profound** Set element **"Stance severity value$_{MAGNITUDE}$"** to 5
Rule Sitting normal **When** Element **Sitting** equals to **Normal, no difficulties sitting >10 sec** **Then** Set element **Sitting Instability** to **Normal** Set element **Sitting** to **Sitting**
Rule Sitting borderline **When** Element **Sitting** equals to **Slight difficulties, intermittent sway** **Then** Set element **"Sitting severity value$_{MAGNITUDE}$"** to 1 Set element **Sitting Instability** to **Borderline**
Rule Sitting mild **When** Element **Sitting** equals to **Constant sway, but able to sit > 10 s without support** **Then** Set element **"Sitting severity value$_{MAGNITUDE}$"** to 2 Set element **Sitting Instability** to **Mild**
Rule Sitting moderate **When** Element **Sitting** equals to **Able to sit for > 10 s only with intermittent support** **Then** Set element **"Sitting severity value$_{MAGNITUDE}$"** to 3 Set element **Sitting Instability** to **Moderate**
Rule Sitting severe **When**

Element **Sitting** equals to **Unable to sit for >10 s without continuous support** **Then** Set element **"Sitting severity value$_{MAGNITUDE}$"** to 4 Set element **Sitting Instability** to **Severe**
Rule speech disturbance normal **When** Element **Speech Disturbance** equals to **Normal** **Then** Set element **Speech Disturbance** to **Speech Disturbance** Set element **Dysarthria** to **Normal**
Rule speech disturbance borderline **When** Element **Speech Disturbance** equals to **Suggestion of speech disturbance** **Then** Set element **Dysarthria** to **Borderline**
Rule speech disturbance mild **When** Element **Speech Disturbance** equals to **Impaired speech, but easy to understand** **Then** Set element **Dysarthria** to **Mild**
Rule speech disturbance moderate **When** ((Element **Speech Disturbance** equals to **Occasional words difficult to understand**) **or** (Element **Speech Disturbance** equals to **Many words difficult to understand**)) **Then** Set element **Dysarthria** to **Moderate**
Rule speech disturbance severe **When** Element **Speech Disturbance** equals to **Only single words understandable**

Then
Set element **Dysarthria** to **Severe**

Rule speech disturbance profound
When
Element **Speech Disturbance** equals to **Speech unintelligible / anarthria**
Then
Set element **Dysarthria** to **Profound**

Rule Finger chase right normal
When
Element **Finger chase-right hand** equals to **No dysmetria**
Then
Set element **Upper Limb Dysmetria Right** to **Normal**
Set element **Finger chase-right hand** to **Finger chase-right hand**

Rule Finger chase right mild
When
Element **Finger chase-right hand** equals to **Dysmetria, under/ overshooting target <5 cm**
Then
Set element **"Finger chase right severity value$_{MAGNITUDE}$"** to 2
Set element **Upper Limb Dysmetria Right** to **Mild**

Rule Finger chase right moderate
When
Element **Finger chase-right hand** equals to **Dysmetria, under/ overshooting target < 15 cm**
Then
Set element **"Finger chase right severity value$_{MAGNITUDE}$"** to 3
Set element **Upper Limb Dysmetria Right** to **Moderate**

Rule Finger chase right severe
When
((
Element **Finger chase-right hand** equals to **Dysmetria, under/ overshooting target > 15 cm**
) **or** (
Element **Finger chase-right hand** equals to **Unable to perform 5 pointing movements**
))

Then
Set element **"Finger chase right severity value**$_{\text{MAGNITUDE}}$**"** to 4
Set element **Upper Limb Dysmetria Right** to **Severe**

Rule Finger chase left normal
When
Element **Finger chase-left hand** equals to **No dysmetria**
Then
Set element **Upper Limb Dysmetria Left** to **Normal**
Set element **Finger chase-left hand** to **Finger chase-left hand**

Rule Finger chase left mild
When
Element **Finger chase-left hand** equals to **Dysmetria, under/ overshooting target <5 cm**
Then
Set element **"Finger chase left severity value** $_{\text{MAGNITUDE}}$**"** to 2
Set element **Upper Limb Dysmetria Left** to **Mild**

Rule Finger chase left moderate
When
Element **Finger chase-left hand** equals to **Dysmetria, under/ overshooting target < 15 cm**
Then
Set element **"Finger chase left severity value** $_{\text{MAGNITUDE}}$**"** to 3
Set element **Upper Limb Dysmetria Left** to **Moderate**

Rule Finger chase left severe
When
((
Element **Finger chase-left hand** equals to **Dysmetria, under/ overshooting target > 15 cm**
) **or** (
Element **Finger chase-left hand** equals to **Unable to perform 5 pointing movements**
))
Then
Set element **"Finger chase left severity value** $_{\text{MAGNITUDE}}$**"** to 4
Set element **Upper Limb Dysmetria Left** to **Severe**

Rule nose finger test right normal
When
Element **Nose-finger test-right hand** equals to **No tremor**

Then Set element **Intention Tremor Right** to **Normal** Set element **Nose-finger test-right hand** to **Nose-finger test-right hand**
Rule nose finger test righ mild **When** Element **Nose-finger test-right hand** equals to **Tremor with an amplitude < 2 cm** **Then** Set element **"Nose finger test right severity value$_{MAGNITUDE}$"** to 2 Set element **Intention Tremor Right** to **Mild**
Rule nose finger test righ moderate **When** Element **Nose-finger test-right hand** equals to **Tremor with an amplitude < 5 cm** **Then** Set element **"Nose finger test right severity value$_{MAGNITUDE}$"** to 3 Set element **Intention Tremor Right** to **Moderate**
Rule nose finger test righ severe **When** ((Element **Nose-finger test-right hand** equals to **Tremor with an amplitude > 5 cm**) **or** (Element **Nose-finger test-right hand** equals to **Unable to perform 5 pointing movements**)) **Then** Set element **"Nose finger test right severity value$_{MAGNITUDE}$"** to 4 Set element **Intention Tremor Right** to **Severe**
Rule nose finger test left normal **When** Element **Nose-finger test-left hand** equals to **No tremor** **Then** Set element **Intention Tremor Left** to **Normal**

Set element **Nose-finger test-left hand** to **Nose-finger test-left hand**
Rule nose finger test left mild **When** Element **Nose-finger test-left hand** equals to **Tremor with an amplitude < 2 cm** **Then** Set element **"Nose finger test left severity value$_{MAGNITUDE}$"** to 2 Set element **Intention Tremor Left** to **Mild**
Rule nose finger test left moderate **When** Element **Nose-finger test-left hand** equals to **Tremor with an amplitude < 5 cm** **Then** Set element **"Nose finger test left severity value$_{MAGNITUDE}$"** to 3 Set element **Intention Tremor Left** to **Moderate**
Rule nose finger test left severe **When** ((Element **Nose-finger test-left hand** equals to **Tremor with an amplitude > 5 cm**) **or** (Element **Nose-finger test-left hand** equals to **Unable to perform 5 pointing movements**)) **Then** Set element **"Nose finger test left severity value$_{MAGNITUDE}$"** to 4 Set element **Intention Tremor Left** to **Severe**
Rule fast alternating hand movements right normal **When** Element **Fast alternating hand movements-right hand** equals to **Normal, no irregularities (performs <10s)** **Then** Set element **Fast alternating hand movements-right hand** to **Fast alternating hand movements-right hand** Set element **Dysdiadochokinesis Right** to **Normal**
Rule fast alternating hand movements right mild **When**

Element **Fast alternating hand movements-right hand** equals to **Slightly irregular (performs <10s)** **Then** Set element **"Fast alternating hand movements right severity value** MAGNITUDE**"** to 2 Set element **Dysdiadochokinesis Right** to **Mild**
Rule fast alternating hand movements right moderate **When** Element **Fast alternating hand movements-right hand** equals to **Clearly irregular, single movements difficult to distinguish or relevant interruptions, but performs <10s** **Then** Set element **Dysdiadochokinesis Right** to **Moderate** Set element **"Fast alternating hand movements right severity value** MAGNITUDE**"** to 3
Rule fast alternating hand movements right severe **When** ((Element **Fast alternating hand movements-right hand** equals to **Very irregular, single movements difficult to distinguish or relevant interruptions, performs >10s**) **or** (Element **Fast alternating hand movements-right hand** equals to **Unable to complete 10 cycles**)) **Then** Set element **"Fast alternating hand movements right severity value** MAGNITUDE**"** to 4 Set element **Dysdiadochokinesis Right** to **Severe**
Rule fast alternating hand movements left normal **When** Element **Fast alternating hand movements-left hand** equals to **Normal, no irregularities (performs <10s)** **Then** Set element **Dysdiadochokinesis Left** to **Normal** Set element **Fast alternating hand movements-left hand** to **Fast alternating hand movements-left hand**

Rule fast alternating hand movements left mild **When** Element **Fast alternating hand movements-left hand** equals to **Slightly irregular (performs <10s)** **Then** Set element **"Fast alternating hand movements left severity value$_{MAGNITUDE}$"** to 2 Set element **Dysdiadochokinesis Left** to **Mild**
Rule fast alternating hand movements left moderate **When** Element **Fast alternating hand movements-left hand** equals to **Clearly irregular, single movements difficult to distinguish or relevant interruptions, but performs <10s** **Then** Set element **"Fast alternating hand movements left severity value$_{MAGNITUDE}$"** to 3 Set element **Dysdiadochokinesis Left** to **Moderate**
Rule fast alternating hand movements left severe **When** ((Element **Fast alternating hand movements-left hand** equals to **Very irregular, single movements difficult to distinguish or relevant interruptions, performs >10s**) **or** (Element **Fast alternating hand movements-left hand** equals to **Unable to complete 10 cycles**)) **Then** Set element **"Fast alternating hand movements left severity value$_{MAGNITUDE}$"** to 4 Set element **Dysdiadochokinesis Left** to **Severe**
Rule heel shin slide right normal **When** Element **Heel-shin slide-right hand** equals to **Normal** **Then** Set element **Heel-shin slide-right hand** to **Nose-finger test-right hand** Set element **Lower Limb Dysmetria Right** to **Normal**

Rule heel shin slide right mild **When** Element **Heel-shin slide-right hand** equals to **Slightly abnormal, contact to shin maintained** **Then** Set element **"Heel-shin slide right severity value$_{MAGNITUDE}$"** to 2 Set element **Lower Limb Dysmetria Right** to **Mild**
Rule heel shin slide right moderate **When** Element **Heel-shin slide-right hand** equals to **Clearly abnormal, goes off shin up to 3 times during 3 cycles** **Then** Set element **"Heel-shin slide right severity value$_{MAGNITUDE}$"** to 3 Set element **Lower Limb Dysmetria Right** to **Moderate**
Rule heel shin slide right severe **When** ((Element **Heel-shin slide-right hand** equals to **Severely abnormal, goes off shin 4 or more times during 3 cycles**) **or** (Element **Heel-shin slide-right hand** equals to **Unable to perform the task**)) **Then** Set element **"Heel-shin slide right severity value$_{MAGNITUDE}$"** to 4 Set element **Lower Limb Dysmetria Right** to **Severe**
Rule heel shin slide left normal **When** Element **Heel-shin slide-left hand** equals to **Normal** **Then** Set element **Heel-shin slide-left hand** to **Heel-shin slide-left hand** Set element **Lower Limb Dysmetria Left** to **Normal**
Rule heel shin slide left mild **When** Element **Heel-shin slide-left hand** equals to **Slightly abnormal, contact to shin maintained** **Then**

Set element **"Heel-shin slide left severity value$_{MAGNITUDE}$"** to 2 Set element **Lower Limb Dysmetria Left** to **Mild**
Rule heel shin slide left moderate **When** Element **Heel-shin slide-left hand** equals to **Clearly abnormal, goes off shin up to 3 times during 3 cycles** **Then** Set element **"Heel-shin slide left severity value$_{MAGNITUDE}$"** to 3 Set element **Lower Limb Dysmetria Left** to **Moderate**
Rule heel shin slide left severe **When** ((Element **Heel-shin slide-left hand** equals to **Severely abnormal, goes off shin 4 or more times during 3 cycles**) **or** (Element **Heel-shin slide-left hand** equals to **Unable to perform the task**)) **Then** Set element **"Heel-shin slide left severity value$_{MAGNITUDE}$"** to 4 Set element **Lower Limb Dysmetria Left** to **Severe**
Rule Midline and Gait **When** Element **Gait severity value** is greater than or equals to **Stance severity value** Element **Gait severity value** is greater than or equals to **Sitting severity value** **Then** Set element **Midline severity value** to **Gait severity value**
Rule Midline and Stance **When** Element **Stance severity value** is greater than **Midline severity value** **Then** Set element **Midline severity value** to **Stance severity value**
Rule Midline and Sitting **When** Element **Sitting severity value** is greater than **Midline severity**

value **Then** Set element **Midline severity value** to **Sitting severity value**
Rule Midline normal **When** ((Element **Midline severity value** equals to **0**) **or** (Element **Sara Total Score** is less than **3**)) **Then** Set element **Midline Ataxia** to **Normal**
Rule Midline Borderline **When** Element **Sara Total Score** is greater than or equals to **3** Element **Midline severity value** equals to **1** **Then** Set element **Midline Ataxia** to **Borderline**
Rule Midline Mild **When** Element **Sara Total Score** is greater than or equals to **3** Element **Midline severity value** equals to **2** **Then** Set element **Midline Ataxia** to **Mild**
Rule Midline moderate **When** Element **Sara Total Score** is greater than or equals to **3** Element **Midline severity value** equals to **3** **Then** Set element **Midline Ataxia** to **Moderate**
Rule Midline severe **When** Element **Sara Total Score** is greater than or equals to **3** Element **Midline severity value** equals to **4** **Then** Set element **Midline Ataxia** to **Severe**
Rule Midline Profound **When**

Element **Sara Total Score** is greater than or equals to **3** Element **Midline severity value** equals to **5** **Then** Set element **Midline Ataxia** to **Profound**
Rule Has Abasia yes **When** Element **Gait** equals to **Unable to walk, even supported** **Then** Set element **Abasia** to **Yes**
Rule Has Abasia no **When** Element **Gait** is not equal to **Unable to walk, even supported** **Then** Set element **Abasia** to **No**
Rule Has Astasia yes **When** Element **Stance** equals to **Unable to stand for >10 s even with constant support of one arm** **Then** Set element **Astasia** to **Yes**
Rule Has Astasia no **When** Element **Stance** is not equal to **Unable to stand for >10 s even with constant support of one arm** **Then** Set element **Astasia** to **No**
Rule Has Anarthria yes **When** Element **Speech Disturbance** equals to **Speech unintelligible / anarthria** **Then** Set element **Anarthria** to **Yes**
Rule Has Anarthria no **When** Element **Speech Disturbance** is not equal to **Speech unintelligible / anarthria** **Then** Set element **Anarthria** to **No**

Rule Cerebellar syndrome no **When** ((Element **Counter** is less than or equals to **1**) **or** (Element **Sara Total Score** is less than or equals to **1**)) **Then** Set element **Cerebellar Syndrome** to **No**
Rule Cerebellar syndrome no significant **When** Element **Counter** is greater than **1** Element **Sara Total Score** is greater than **1** Element **Sara Total Score** is less than **3** **Then** Set element **Cerebellar Syndrome** to **No Significant**
Rule Cerebellar syndrome mild **When** Element **Counter** is greater than **1** Element **Sara Total Score** is greater than or equals to **3** Element **Sara Total Score** is less than or equals to **8** **Then** Set element **Cerebellar Syndrome** to **Mild**
Rule Cerebellar syndrome moderate **When** Element **Counter** is greater than **1** Element **Sara Total Score** is greater than **8** Element **Sara Total Score** is less than or equals to **15** **Then** Set element **Cerebellar Syndrome** to **Moderate**
Rule Cerebellar syndrome severe **When** Element **Counter** is greater than **1** Element **Sara Total Score** is greater than **15** **Then** Set element **Cerebellar Syndrome** to **Severe**
Rule Appendicular right and finger chase right **When**

Element **Appendicular right severity value** is less than **Finger chase right severity value** **Then** Set element **Appendicular right severity value** to **Finger chase right severity value**
Rule Appendicular right and nose finger test right **When** Element **Appendicular right severity value** is less than **Nose finger test right severity value** **Then** Set element **Appendicular right severity value** to **Nose finger test right severity value**
Rule Appendicular right and fast alternating hand movements right **When** Element **Appendicular right severity value** is less than **Fast alternating hand movements right severity value** **Then** Set element **Appendicular right severity value** to **Fast alternating hand movements right severity value**
Rule Appendicular right and heel shin slide right **When** Element **Appendicular right severity value** is less than **Heel-shin slide right severity value** **Then** Set element **Appendicular right severity value** to **Heel-shin slide right severity value**
Rule Appendicular right normal **When** ((Element **Sara Total Score** is less than **3**) **or** (Element **Appendicular right severity value** equals to **0**)) **Then** Set element **Appendicular Ataxia Right** to **Normal**
Rule Appendicular right mild **When**

Element **Sara Total Score** is greater than or equals to **3** Element **Appendicular right severity value** equals to **2** **Then** Set element **Appendicular Ataxia Right** to **Mild**
Rule Appendicular right moderate **When** Element **Appendicular right severity value** equals to **3** Element **Sara Total Score** is greater than or equals to **3** **Then** Set element **Appendicular Ataxia Right** to **Moderate**
Rule Appendicular right severe **When** Element **Sara Total Score** is greater than or equals to **3** Element **Appendicular right severity value** equals to **4** **Then** Set element **Appendicular Ataxia Right** to **Severe**
Rule Appendicular left and finger chase left **When** Element **Appendicular left severity value** is less than **Finger chase left severity value** **Then** Set element **Appendicular left severity value** to **Finger chase left severity value**
Rule Appendicular left and nose finger test left **When** Element **Appendicular left severity value** is less than **Nose finger test left severity value** **Then** Set element **Appendicular left severity value** to **Nose finger test left severity value**
Rule Appendicular left and fast alternating hand movements left **When** Element **Appendicular left severity value** is less than **Fast alternating hand movements right severity value** **Then** Set element **Appendicular left severity value** to **Fast alternating hand movements right severity value**

Rule Appendicular left and heel shin slide left **When** Element **Appendicular left severity value** is less than **Heel-shin slide left severity value** **Then** Set element **Appendicular left severity value** to **Heel-shin slide left severity value**
Rule Appendicular left normal **When** ((Element **Sara Total Score** is less than **3**) **or** (Element **Appendicular left severity value** equals to **0**)) **Then** Set element **Appendicular Ataxia Left** to **Normal**
Rule Appendicular left mild **When** Element **Sara Total Score** is greater than or equals to **3** Element **Appendicular left severity value** equals to **2** **Then** Set element **Appendicular Ataxia Left** to **Mild**
Rule Appendicular left moderate **When** Element **Appendicular left severity value** equals to **3** Element **Sara Total Score** is greater than or equals to **3** **Then** Set element **Appendicular Ataxia Left** to **Moderate**
Rule Appendicular left severe **When** Element **Sara Total Score** is greater than or equals to **3** Element **Appendicular left severity value** equals to **4** **Then** Set element **Appendicular Ataxia Left** to **Severe**

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ACRONYMS AND ABBREVIATIONS

ADL	Archetype Definition Language
AM	Archetype Model
APGAR	Appearance, Pulse, Grimace, Activity and Respiration
CDA	Clinical Document Architecture
CDE	Common Data Element
CDS	Clinical Decision Support
CKM	Clinical Knowledge Manager
DL	Description Logic
ECG	Electrocardiography
EHR	Electronic Health Record
GCS	Glasgow Comma Scale
GDL	Guideline Definition Language
HL7	Health Level 7
HPO	Human Phenotype Ontology
ICD	International Classification of Diseases
MMSE	Mini-Mental State Examination
MRS	Modified Rankin Scale
MS	Multiple Sclerosis
NINDS	National Institute of Neurological Disorders and Stroke
OBO	Open Biomedical Ontology
OWL	Web Ontology Language
RM	Reference Model
RM	Reference Model
SARA	Scale for the Assessment and Rating of Ataxia
SCA36	Spinocerebellar Ataxia Type 36
SMS	SARA Management System
SNHL	Slowly progressive sensorineural hearing loss
UPDRS	Unified Parkinson's Disease Rating Scale

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA