Phenopcakets proposal - GCAT

This is a description of the exercise of converting the phenotypic data from a dataset from GCAT project to Phenopackets. The resulting proposed phenopacket structure can be found here: https://github.com/clauw87/phenopackets/blob/master/gcat_pheno

1 Description of data

This dataset has the following data:

1. Lifestyle-Socio/demographic [BASELINE]
2. Diagnoses [TIMECOURSE]
3. Medical Interventions [TIMECOURSE]
4. Medications [TIMECOURSE]

2 Reuse of base phenopacket shema and new messages

For this exercise, we used Phenopacket schema v1.0 https://github.com/phenopackets/phenopacket-schema and new messages proposed for v1.1 https://github.com/phenopackets/phenopacket-schema/compare/v1.1

- Data from each individual was stored as a PhenopacketSet associated to that individual, where each Phenopacket is defined by its date (i.e data is in same phenopacket if they shared date value)
- One-time point data (survey data) was stored as one Phenopacket with *Date* and *AgeAtEncounter* set at the beginning of the project, if the actual date is not available.
- Survey lifestyle data on Smoking and Alcohol Drinking and Physical Activity was stored as message *Exposure* with *Evidence* message set as "survey", and including a message *Modifier* to indicate current or past behavior as well as messages *TimeElement* (Age at onset and also Interval for past behaviors), *Quantity* (OntologyClass) and *Frequency* (OntologyClass)
- Timecourse data (Clinical Findings, Diagnoses, Intervention, Medications) was stored in the Phenopackets as messages *Clinical Findings* (based on base schema PhenotypicFeatures), *Diagnoses* (based on base schema Diseases), and *Interventions* (based on new message Procedures) and *Medications* (based on new message PharmaceuticalTreatments) within new message MedicalAction.

3 Adaptations of messages

- A *Frequency* (ontologyClass) block added in *Dose Interval* within *Medications* in addition to *Quantity* and *Interval* to express medication routines as 50 (value) mg (units) "daily"/"every 8 hours" (frequency) for 2 weeks (interval). This could be later converted/ interoperable through summary doses per week, year, etc.
- *Quantity* message and *Frequency* (OntologyClass) were included in *Exposure* message for Lifestyle/ behavior data
- New *Measurements* message with *Parameter* (OntologyClass) and *Quantity* used for quantitative measures (anthropometric measures and medical measures)

4 Metadata

Are these standard enough? should we map to some other standards?

- Units: IS units, non-IS units if necessary can defined in Metadata message?, e.g smoking fagerstrom score, MET
- Standards ATC for medications, ICD9 for diagnoses, X GCAT codes for Interventions