pedigree

Release 0.1

GA4GH Clinical and Phenotypic Data Capture Workstream

Feb 08, 2023

TABLE OF CONTENTS

Note: This project is under active development.

CHAPTER

ONE

INTRODUCTION

1.1 The GA4GH Pedigree Standard

The GA4GH Pedigree Standard supports the computable exchange of family structure and relationships for family health history and pedigree use cases. It does this by providing a common conceptual model, implementation guides in common standards, and tooling to support the adoption within existing pipelines. Relationships between individuals are codified using the Kinship Ontology (KIN) to allow for inference, semantic interoperability, and reasoning.

The primary goal of the GA4GH Pedigree Standard is to support the computable exchange of familial health information between the following healthcare settings:

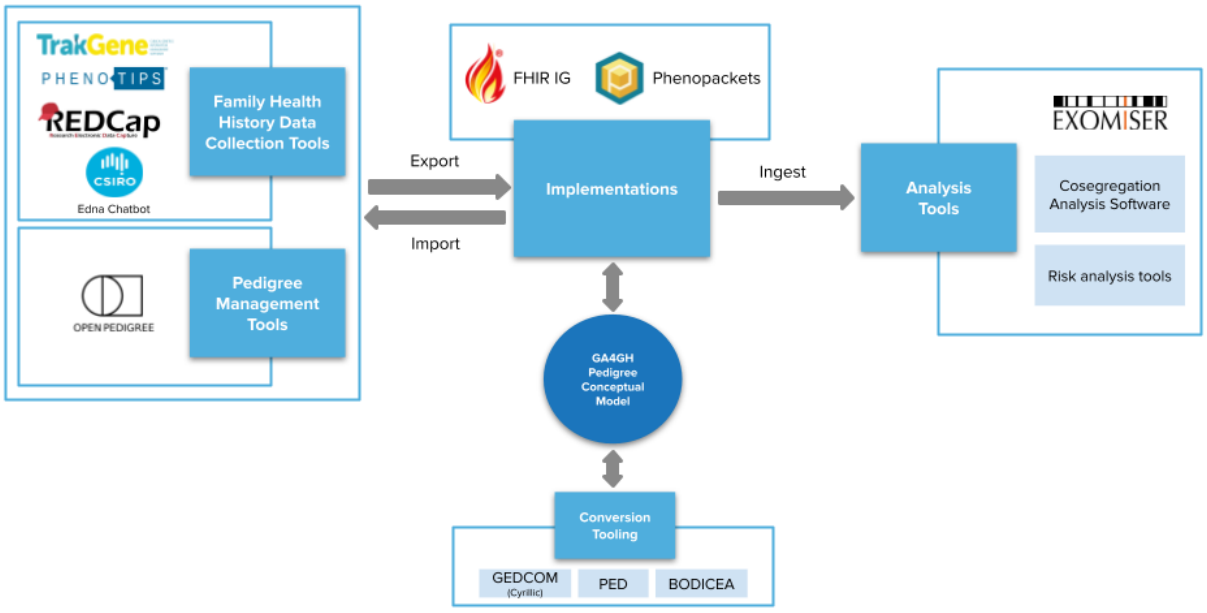
- self-reported by patients in portals and intake forms
- collected by nurses and genetic counselors from patients
- visualized as a pedigree for clinicians, counselors, and specialists
- used for sample relationships in genomic data analysis
- used by risk algorithms, e.g. hereditary cancer predisposition

Because the standardization of detailed clinical and genetic data is well-supported by other formats and efforts, the GA4GH Pedigree standard focuses on improving the core representation of individuals and their relationships in the context of a family. Additional clinical and genetic data associated with individuals in the family are expected to be represented in other standards and linked to individuals within the pedigree representation.

1.2 Key terminology

A **family health history** is the description of the health conditions in a person's family, usually from a single historian (usually the patient or their caregiver). Family health histories are routinely collected as part of health care and typically stored in the patient's medical record.

A **pedigree** is a standardized representation of the individuals, relationships, and health conditions in a family. This is usually drawn or visualized using standardized symbols (such as circles for women, squares for men, and diamonds for non-binary individuals). This information often comes from the family health histories reported by family members, but is usually curated by the clinical team. In clinical genetics, information from multiple family members is frequently combined to form a single pedigree which is stored in a family record that is separate from any individual's medical record.



1.3 Motivation

The need for high quality, unambiguous, computable pedigree and family information is critical for scaling genomic analysis to larger, complex families.

Pedigree data is currently represented in heterogeneous formats that frequently result in the use of lowest-common-denominator formats (e.g., PED) or custom JSON formats for data transfer. The HL7 FHIR standard core data models do not support pedigrees, but there is a draft extension to support genomic pedigrees for a single, fixed proband patient.

By standardizing the way systems represent family structure, patients will be able to share this information more easily between healthcare systems, and software tools will be better able to use this information to improve genome analysis and diagnosis.

We asked our stakeholders about their use of family health history and pedigree data - How are you using it? How is it stored? What do you wish you could do with your data that you currently can't? The results of the survey can be found here. A significant percentage of respondents were using a non-computable or non-interoperable format, and there was no common tool or format with which they intended to import or export data. Importantly, 57% of respondents were experiencing challenges with standardization, including lack of computability and integration with analysis tools, and inability to represent complex families and share data easily.

A full listing of the use cases that informed development can be reviewed here.

1.4 Existing Standards

1.4.1 Pedigree

The PED format is a simple text file with 6 columns - IDs, a binary sex field, the phenotype (singular) and SNP genotypes. It can represent parent-child relationships only. It is unable to communicate twins, adoption or donors, pregnancy, vital status, multiple phenotypes and data provenance. All of this type of data is important for genetic counseling and risk assessments where richer representations of relationships are valuable.

1.4.2 Family History

The HL7 FamilyMemberHistory resource and FamilyMemberHistory-genetic profile allow for capturing a proband's family health history. All data and relationships are relative to that single proband and are in the context of a single patient. This limits its use to representing the family history of a single individual rather than the complete pedigree for a family.

The HL7 FHIR FamilyMember ValueSet is a taxonomy, not an ontology, which limits its utility for computation and reasoning. It is also Anglo-centric in its construction, which limits the global adoption, for example, there aren't equivalent terms for "aunt" and "uncle". By creating the Kinship Ontology, we were able to define deeper semantics between relationships, allowing these terms to be used for inference or validation.

1.5 The Common Dataset for Family Health History

The collection and use of family health histories span medical activities from genetic research to heritable risk assessment in patient care. For all the stakeholders in this process, the goal must be data that is accurate and coded for effective analysis, and transferable between systems. To achieve this, a globally accepted and universally implemented family health history (FHH) data set should be established as a benchmark. The purpose of the common dataset document is to create an updated recommended data set that can be used not only in both research and clinical settings, but to eliminate the gap between the two disciplines. This recommendation should also guide the development of research, clinical, and patient-facing FHH data and information collection tools, applications, and data repositories. This document should only be used as informative.

Common Dataset Document

This work was inspired by the efforts of the Personalized Health Care Workgroup of the American Health Information Community, which first released its recommendation on a core family health history (FHH) minimum data set on October 25, 2007. A peer-reviewed paper was published in December 2008.

1.6 Example Use Cases

The overarching use case is to enable the exchange of information collected through family health histories and clinical genetics pedigrees across pedigree tools and algorithms that operate on pedigrees and family health histories.

Specific use cases considered in the development of the standard include:

- Representing relationships necessary for counseling (e.g., adoption), risk assessment (e.g., infertility, miscarriage, health history), and assisted reproduction (e.g., IVF, MRT)
- Allowing the exchange of pedigree information required to inform clinical and research genomic data analysis, noting that the majority of testing involves singletons, <5% as trios, and other family configurations are extremely rare (parent/child duo, sib pair, half-sib pair, quad)

- Allowing sharing collected clinical and family health history information with bioinformatics systems and research environments (or other services) to unambiguously document relationships between sequenced individuals to support joint calling of variants and filtering of variants based on segregation, as well as describing wider family history (re: non-sequenced individuals).
- Allowing the exchange of the necessary family health history, genotype, and phenotype results of a patient or relative to computational tools for assessing whether the patient needs further testing or sequence analysis, and/or if a relative needs the same
- Representing family history and pedigree data in a programmatic standard that people can consume across a number of resources in both a format for analysis as well as for building algorithms and tools over would be of high utility for secondary analysis and research purposes

1.7 Requirement Levels

The Pedigree model uses two requirement levels.

1.7.1 Required

If a field is required, its presence is an absolute requirement of the specification, failing which the entire model is regarded as malformed. This corresponds to the key words `MUST`, `REQUIRED`, and `SHALL` in RFC2119.

1.7.2 Optional

A field is truly optional. This category can be applied to fields that are only useful for a certain type of data. For instance, the Proband ID and Type field is only required when the pedigree is used to focus on heritable risk for a specific person in the pedigree. For other use cases such as research, a Proband type may be needed.

CHAPTER

TWO

CONCEPTUAL MODEL

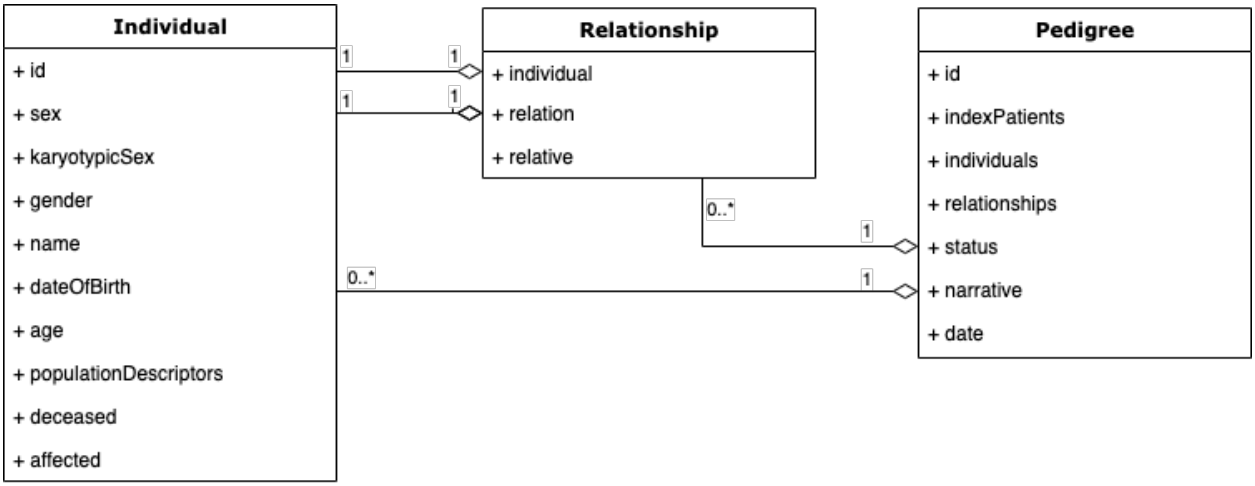
2.1 Overview

To support the interoperability of family health history data within and between existing standards (such as HL7 FHIR and Phenopackets), the GA4GH Clinical and Phenotypic Data Capture Workstream developed the Pedigree Conceptual Model.

The Pedigree Conceptual Model defines core concepts and their properties, and is based on A Recommendation for The Common Data Set for Family Health History.

2.2 Concepts

The diagram below shows an overview of the pedigree concepts. Lines between concepts indicate composition.



2.2.1 Individual

The **Individual** concept represents an individual person or patient who is a member of the pedigree being investigated.

Field	Multiplicity	Description
id	1..1	External identifier for the individual
sex	1..1	Sex assigned at birth
karyotypicSex	0..1	The chromosomal sex of the individual; See Phenopacket KaryotypicSex.
gender	0..1	Presumed or reported gender identity
name	0..1	Name of the individual
dateOfBirth	0..1	Birth date of the individual, can be just birth year in most cases
age	0..1	Age of the individual, can be either Age, Estimated Age (or Ontology Class), Age Range, and/or Gestational Age; See also Phenopackets' TimeElement.
populationDescriptors	0..*	Information about the individual's ancestry, ethnicity, race, tribe, etc.,; terms from the Human Ancestry Ontology (HANCESTRO) are recommended, but freetext must be supported
deceased	0..1	The presumed/accepted life status of the individual as of the pedigree collection date
affected	0..1	Whether or not the individual is affected

2.2.2 Relationship

The *Relationship* concept represents the relationship that one individual has to another individual.

Field	Multiplicity	Description
individual	1..1	Identifier of the subject `Individual`; equivalent to the Biolink "subject"
relation	1..1	The relationship the `individual` has to the `relative` (*e.g.*, if the `individual` is the `relative`'s biological mother, then relation could be `isBiologicalMotherOf [KIN:027]`); terms should come from the KIN Ontology.
relative	1..1	Identifier of the relative `Individual`; equivalent to the Biolink "object"

2.2.3 Pedigree

A **Pedigree** is a set of individuals and the relationships between them.

Field	Multiplicity	Description
id	1..1	External identifier for the family being investigated
indexPatients	0..*	Identified `Individual` in the family of a health condition of focus being investigated: `Proband`, `Consultand`, `First Person Tested Positive`
individuals	0..*	Collection of `Individual` who are the members of this pedigree
relationships	0..*	Collection of `Relationship` between the `individuals` who are the members of this pedigree
status	0..1	Status of the pedigree resource collection
narrative	0..1	Summary of the pedigree resource for human interpretation
date	0..1	The date the pedigree was collected or last updated, as ISO full or partial date, *i.e.* `YYYY`, `YYYY-MM`, or `YYYY-MM-DD`

2.3 Design motivations

Design motivation:

- avoid overlap with other standards (fhir, phenopacket)
- focus on relationship
- graphical model, bringing relationships as top-level entities
- allow for the synthesizing of patient-reported family history data, such as comes out of family history questionnaires and EHR records (and can be represented with the FamilyMemberHistoryResource), and support this information through to risk models
- provide a standard interface for validation
- facilitate conversion among existing standards for pedigree data

Relationships between individuals are standardized using concepts from the newly developed Kinship Ontology. To allow existing workflows and tools to gracefully add interoperability with this standard, we developed an open-source pedigree data interoperability library, pedigree-tools.

CHAPTER

THREE

KINSHIP ONTOLOGY (KIN)

3.1 The Kinship Ontology

The Kinship Ontology (KIN) is a family relations ontology developed as part of the Global Alliance for Genomics and Health Pedigree Standard project. It allows using an OWL reasoner to automatically validate a family history graph and infer new relations.

The latest version of the ontology can be found at: http://purl.org/ga4gh/kin.owl.

The Ontology is open-source and managed in this GitHub repo: https://github.com/GA4GH-Pedigree-Standard/family_history_terminology

Note: We are working with colleagues to explore migrating KIN to the Relations Ontology (RO).

CHAPTER

FOUR

USING THE PEDIGREE STANDARD

4.1 Compatible standards

The GA4GH Pedigree Standard is a conceptual model and recommendations for transferring family history and pedigree data. It is not a standalone data format, but is intended to be implemented by compatible standards to facilitate the transfer and interoperability of this data.

Compatible standards provide an implementation guide for capturing and representing pedigree data in a manner that is compatible with this model.

The representation of each core concept within each standard is summarized in *Conceptual Model*.

The current list of compatible standards are:

- Phenopackets
- HL7 FHIR

4.1.1 Phenopackets

The Phenopackets "Implementation Guide" - an implementation of the GA4GH pedigree spec which is partly composed of phenopacket-schema messages. It is not 'part' of the Phenopackets spec, but sits in its own org.ga4gh.pedigree namespace.

For tools like Exomiser, it is possible to convert to PED format using pedigree-tools and ingest via a Phenopacket.

Phenopackets schema uses protobuf, an exchange format developed in 2008 by Google. It is recommended to review the Wikipedia page on Protobuf and to Google's documentation for details. This page intends to get curious readers who are unfamiliar with protobuf up to speed with the main aspects of this technology, but it is not necessary to understand protobuf to use the phenopacket or pedigree schemas.

Learn more about the Phenopackets here.

4.1.2 HL7 FHIR

Note: Our FHIR-based Implementation Guide of the GA4GH Pedigree conceptual model is under development. The website linked above states the Guide is a "Local Development Build (v0.1.0)". As the development proceeds, all artifacts in the GA4GH Pedigree specification will be assigned a "Maturity Level". When completed, this IG will go through the HL7 balloting process to become part of the normative version of the FHIR standard.

The Pedigree FHIR Implementation Guide

Fast Health Interoperability Resources (FHIR) is a loosely defined base model describing things in healthcare (e.g. Patient, Specimen) and how they relate to each other, developed by Health Level 7 (HL7). The FHIR specification is

completely technology agnostic. Thus, it does not depend on programming languages or include things like relational database schemas. It is up to the implementers to decide how to implement the data model (i.e. relational database, nosql database, etc) and RESTful API.

To learn more about FHIR, we recommend you check out the following resources: HL7.org, FHIR Basics, and this excellent FHIR 101 Jupyter Notebook developed by NIH Cloud-based Platform Interoperability (NCPI) Working Groups.

4.2 Direction of Relationships

A Relationship defines a relationship between one individual and another, such as *isBiologicalMotherOf* or *isTwinOf*. Only one of the two directions needs to be specified, and it does not matter which.

Symmetric relationships are those where both individuals share the same relationship with one another. These include: *isTwinOf* and *isPartnerOf*.

Non-symmetric relationships are those where the relationship that individual X has to individual Y is not the same as the relationship that individual Y has to individual X. For example, if individual X has relationship *isBiologicalParentOf* to individual Y, then individual Y has relationship *isBiologicalChildOf* individual X.

Because of this inherent flexibility in the way that relationships can be described, there is no single representation for a particular pedigree. However, pedigrees can be represented in a **reduced form**, in which implied relationships are excluded. A pedigree in reduced form:

1. Has explicit parent-child relationships between all parents and their offspring, and they are directed downwards, with the parent as the individual and the child as the relative.
2. Has sibling relationships only when this is not implied by having shared parents, and in the event of multiple siblings, all sibling relationships are defined relative to the same individual
3. Defines all twin relationships relative to the same individual
4. Has partnership relationships only when this is not implied by having shared children
5. Has extended relative relationships only when this is not implied by the previously-defined relationships, and they are directed downwards, with the ancestor as the individual and the descendant as the relative

4.3 Pedigree Regulatory & Ethics Disclaimer

This model has been designed for use in clinical and research settings. The model may be implemented differently depending on the use cases and setting within which it will be used. While a stand alone regulatory and ethics review has been performed on the model itself, an independent regulatory and ethics review by the implementer may be required depending on the context of use to consider specific issues such as privacy, confidentiality and/or data security and ensure that the model's implementation and usage is in compliance with applicable legislation and ethical requirements in their jurisdiction. Given that this model is designed to represent family health history data, information which carries potential for personal identification, it is the duty of the implementer to address these risks in the implementation and use of this model. When used in clinical research settings please refer to the Global Alliance for Genomics and Health Policy on Clinically Actionable Genomic Research Results for guidance in managing the return of results.

CHAPTER

FIVE

EXAMPLES

5.1 Examples

The following examples demonstrate the way in which pedigrees of various complexity can be represented using the pedigree conceptual model.

Any pedigree more complex than would be represented with a PED file should use the conceptual model implemented within a compatible standard, such as FHIR or Phenopacket.

5.1.1 Basic Trio

A basic family trio consists of one male parent, one female parent, and a proband child. This would be represented as a Pedigree with three Individuals and two parent-child Relationships:

As a Phenopacket *GA4GHPedigree* message:

```
id: FAM1
narrative: A Phenopacket GA4GHPedigree of a trio with an affected child
date: 2022-06-23
individuals:
  - id: 1
    subject:
      id: MOTHER
      sex: FEMALE
  - id: 2
    subject:
      id: FATHER
      sex: MALE
  - id: 3
    subject:
      id: CHILD
      sex: UNKNOWN
relationships:
  - individual_id: MOTHER
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: CHILD
  - individual_id: FATHER
    relation:
      id: KIN:028
```

(continues on next page)

(continued from previous page)

```
      label: isBiologicalFatherOf
    relative_id: CHILD
index_patients:
  - CHILD
```

5.1.2 Twins

The relationship between twins (TWIN1 and TWIN2) can be represented by adding another Individual, parent-child relationships and a twin Relationship to the Pedigree:

```
id: FAM2
narrative: A Phenopacket GA4GHPedigree of a couple with identical twins
date: 2022-06-23
individuals:
  - id: 1
    subject:
      id: MOTHER
      sex: FEMALE
  - id: 2
    subject:
      id: FATHER
      sex: MALE
  - id: 3
    subject:
      id: TWIN1
      sex: UNKNOWN
  - id: 4
    subject:
      id: TWIN2
      sex: UNKNOWN
relationships:
  - individual_id: MOTHER
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: TWIN1
  - individual_id: FATHER
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: TWIN1
  - individual_id: TWIN1
    relation:
      id: KIN:010
      label: isMonozygoticMultipleBirthSiblingOf
    relative_id: TWIN2
```

The parent-child relationships for TWIN2 are not strictly necessary. Because the *isMonozygoticTwinOf* relationship is symmetric, it would be equally valid to have said that TWIN2 isMonozygoticTwinOf TWIN1.

5.1.3 Adoption

```
id: FAM3
narrative: A Phenopacket GA4GHPedigree of a child with an adoptive mother
date: 2022-06-23
individuals:
  - id: 1
    subject:
      id: MOTHER
      sex: FEMALE
  - id: 2
    subject:
      id: BIOLOGICAL_MOTHER
      sex: FEMALE
  - id: 3
    subject:
      id: FATHER
      sex: MALE
  - id: 4
    subject:
      id: CHILD
      sex: UNKNOWN
relationships:
  - individual_id: MOTHER
    relation:
      id: KIN:022
      label: isAdoptiveParentOf
    relative_id: CHILD
  - individual_id: BIOLOGICAL_MOTHER
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: CHILD
  - individual_id: FATHER
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: CHILD
```

5.1.4 IVF

```
id: FAM4
narrative: A Phenopacket GA4GHPedigree of a child with an egg donor, gestational carrier,
↪ and biological father
date: 2022-06-23
individuals:
  - id: 1
    subject:
      id: MOTHER
      sex: FEMALE
  - id: 2
```

(continues on next page)

(continued from previous page)

```
    subject:
      id: SURROGATE
      sex: FEMALE
  - id: 3
    subject:
      id: FATHER
      sex: MALE
  - id: 4
    subject:
      id: CHILD
      sex: UNKNOWN
relationships:
  - individual_id: MOTHER
    relation:
      id: KIN:038
      label: isOvumDonorOf
    relative_id: CHILD
  - individual_id: SURROGATE
    relation:
      id: KIN:005
      label: isGestationalCarrierOf
    relative_id: CHILD
  - individual_id: FATHER
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: CHILD
```

5.1.5 Complete cancer family

```
id: FAM5
narrative: A Phenopacket GA4GHPedigree of a classic BRCA1 pedigree
date: 2022-06-23
individuals:
  - id: 1
    subject:
      id: 1
      sex: MALE
      vital_status: DECEASED
  - id: 2
    subject:
      id: 2
      sex: FEMALE
      vital_status: DECEASED
  - id: 3
    subject:
      id: 3
      sex: MALE
      vital_status: DECEASED
  - id: 4
```

(continues on next page)

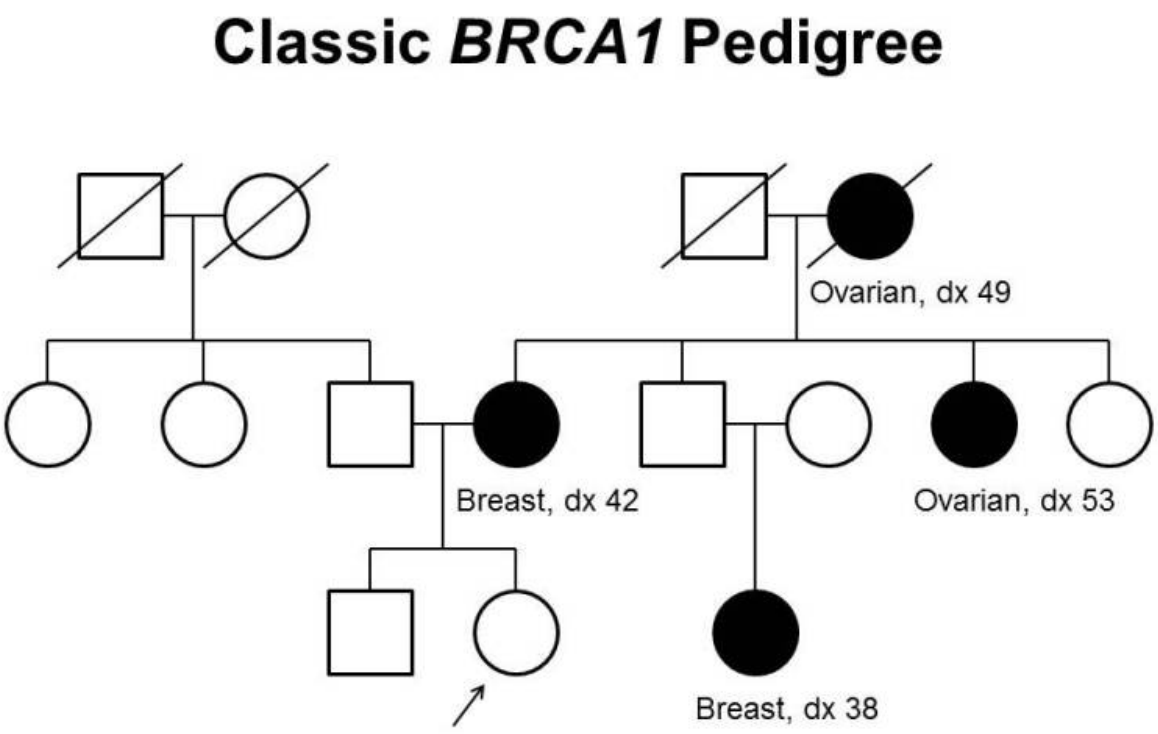


Fig. 1: Example BRCA1 pedigree. Source: https://visualsonline.cancer.gov/details.cfm?imageid=10436

(continued from previous page)

```
    subject:
      id: 4
      sex: FEMALE
      vital_status: DECEASED
      diseases:
        - term:
            id:
            label: Ovarian cancer
          onset:
            age: P49Y
  - id: 5
    subject:
      id: 5
      sex: FEMALE
  - id: 6
    subject:
      id: 6
      sex: FEMALE
  - id: 7
    subject:
      id: 7
      sex: MALE
  - id: 8
    subject:
      id: 8
      sex: FEMALE
      diseases:
        - term:
            id:
            label: Breast cancer
          onset:
            age: P42Y
  - id: 9
    subject:
      id: 9
      sex: MALE
  - id: 10
    subject:
      id: 10
      sex: FEMALE
  - id: 11
    subject:
      id: 11
      sex: FEMALE
      diseases:
        - term:
            id:
            label: Ovarian cancer
          onset:
            age: P53Y
  - id: 12
    subject:
```

(continues on next page)

(continued from previous page)

```
      id: 12
      sex: FEMALE
  - id: 13
    subject:
      id: 13
      sex: MALE
  - id: 14
    subject:
      id: 14
      sex: FEMALE
  - id: 15
    subject:
      id: 15
      sex: FEMALE
      diseases:
        - term:
            id:
            label: Breast cancer
          onset:
            age: P38Y
relationships:
  - individual_id: 1
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 5
  - individual_id: 2
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 5
  - individual_id: 1
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 6
  - individual_id: 2
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 6
  - individual_id: 1
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 7
  - individual_id: 2
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 7
  - individual_id: 3
```

(continues on next page)

(continued from previous page)

```
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 8
  - individual_id: 4
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 8
  - individual_id: 3
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 9
  - individual_id: 4
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 9
  - individual_id: 3
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 11
  - individual_id: 4
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 11
  - individual_id: 3
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 12
  - individual_id: 4
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 12
  - individual_id: 7
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 13
  - individual_id: 8
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 13
  - individual_id: 9
    relation:
      id: KIN:028
```

(continues on next page)

(continued from previous page)

```
      label: isBiologicalFatherOf
    relative_id: 14
  - individual_id: 10
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 14
  - individual_id: 7
    relation:
      id: KIN:028
      label: isBiologicalFatherOf
    relative_id: 15
  - individual_id: 8
    relation:
      id: KIN:027
      label: isBiologicalMotherOf
    relative_id: 15
index_patients:
  - 14
```

CHAPTER

SIX

TOOLING

6.1 Pedigree Tools

Pedigree-tools is a library for supporting the conversion of pedigree data between various file formats.

It can currently support importing from the following formats:

- PED/Linkage
- GEDCOM (Cyrillic)
- BOADICEA

In can currently export into the following formats:

- PED/Linkage

This tool is available at the following GitHub repository: https://github.com/GA4GH-Pedigree-Standard/pedigree-tools

6.2 Pedigree Validator

Pedigree Validator is a simple command line application that shows how validation of a FHIR pedigree file can be implemented using the HAPI FHIR libraries and the artifacts produced by the FHIR implementation guide.

It also shows how an OWL reasoner can be used to implement additional validation based on the KIN ontology.

The application is available at the following GitHub repository: https://github.com/GA4GH-Pedigree-Standard/pedigree-validator

CHAPTER

SEVEN

IMPLEMENTATIONS

7.1 Known Implementations

The following systems have implemented the GA4GH Pedigree Standard:

7.1.1 FHIR implementations:

CSIRO Redcap Pedigree Plugin (Open Source)
Open Pedigree (Open Source)

The REDCap Pedigree editor External Module is a third party add on to REDCap, a web-based application for building and managing online surveys. The external module allows a field on a survey to be marked as a 'pedigree diagram'. Clicking the field will open a new web browser to the open pedigree web-based pedigree editor, where the pedigree diagram can be entered. The resulting diagram will then be placed into the survey in REDCap as a JSON string using the GA4GH FHIR IG format.

Kids First Data Resource Center (currently in testing)

The GA4GH Pedigree model was test-implemented with one of the research studies registered in and publicized by Kids First DRC (hereinafter "KFDRC"). This implementation was done in KFDRC's development environment based on the GA4GH Pedigree FHIR Implementation Guide and demonstrated at the GA4GH December 2021 Connect. The research study chosen for this use case is titled "GMKF: Kids First Pediatric Research Program on Congenital Cranial Dysinnervation Disorders and Related Birth Defects" (dbGaP Study Accession: phs001247.v1.p1). This implementation has the following FHIR resources:

FHIR Resource Type	Pedigree Profile	# of Resources	Note
Patient	Individual	899	270 probands
FamilyMemberHistory	Relationship	772	27 KIN codes
Condition	–	359	5 diseases
Composition	Pedigree	270	23 families with more than trios

The figure below shows one example of pedigree (KF Family ID: FM_C0YWP4XR) that has 8 family members (the indexing proband's ID: 5047) and the phenotypic abnormality being investigated is with CFP (Complement Factor Properdin). Between Individual resources, 12 Relationship resources were created. The left part of the figure represents this family's pedigree chart while the right part shows a Relationship resource between Individual 5047 (the indexing patient) and Individual 5037, who is a parental sibling (i.e. a paternal aunt) of 5047.

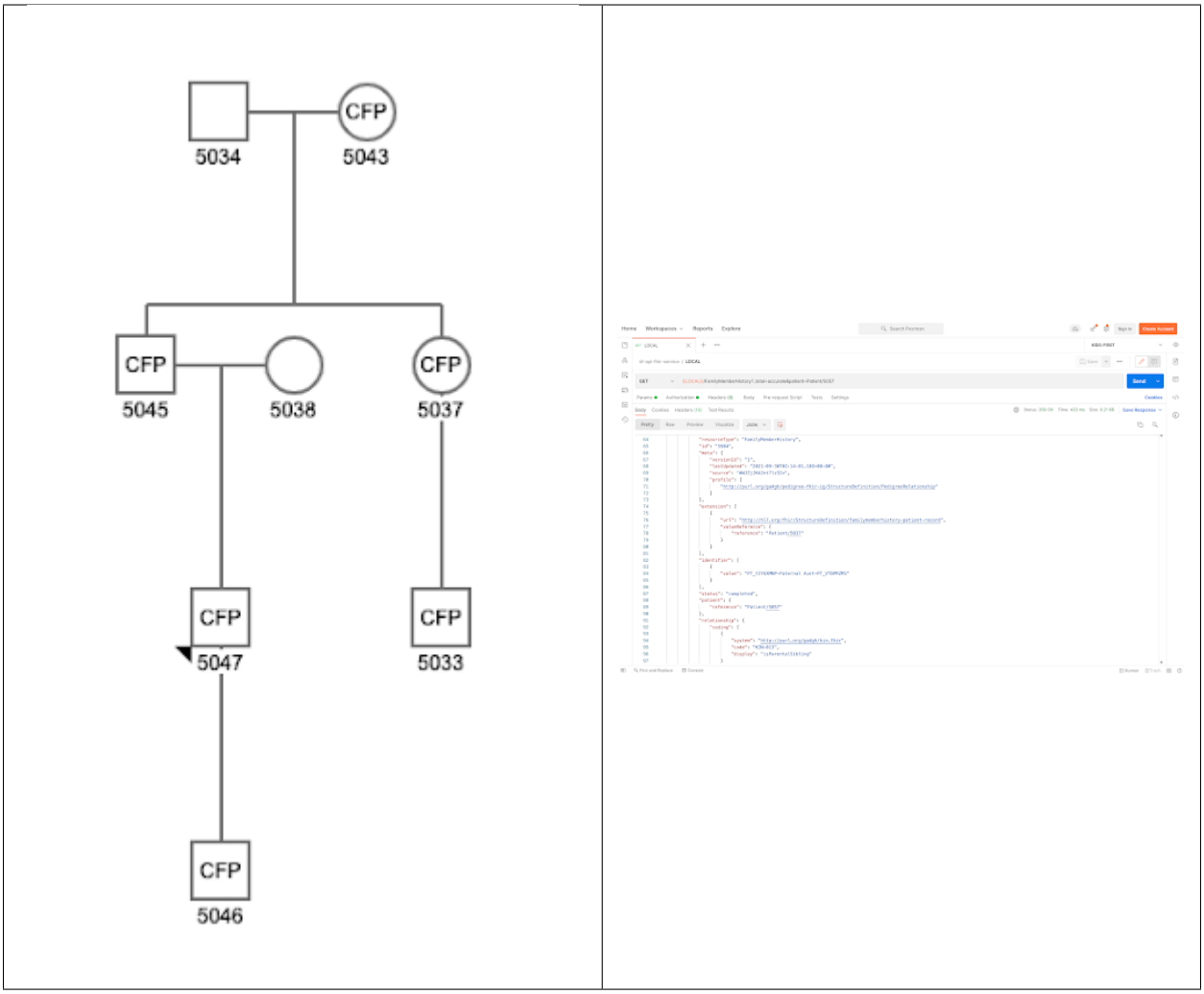


Figure: KFDRC Pedigree FHIR IG Implementation Example showing the family's pedigree chart and the associated relationship resource.

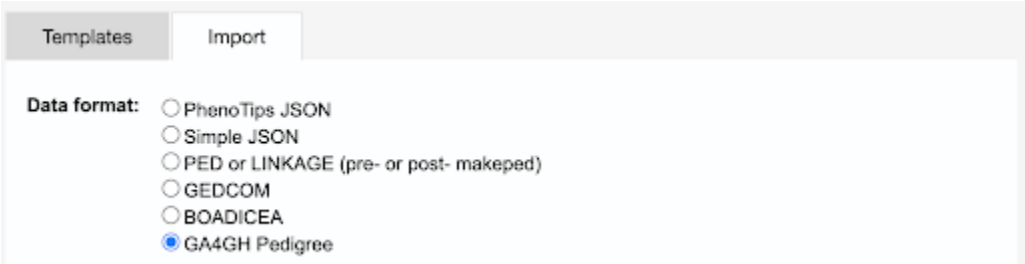
7.1.2 Phenopacket implementations:

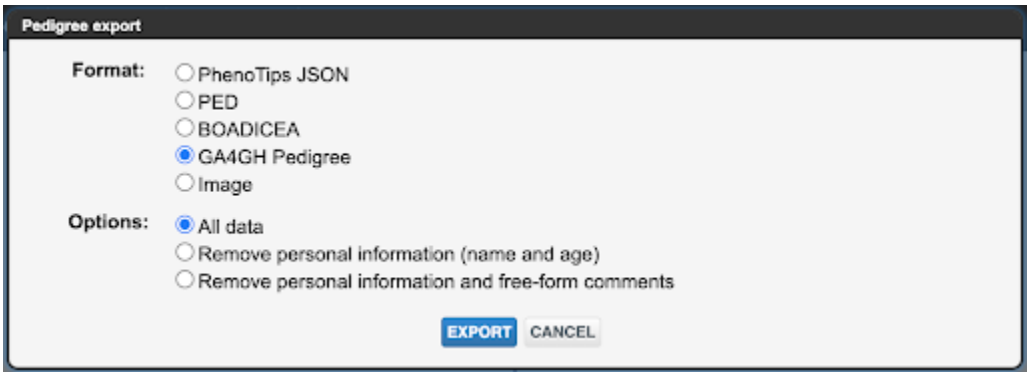
PhenoTips (not yet fully implemented)

PhenoTips is a commercial clinical software platform which includes a comprehensive pedigree editor. https://phenotips.com/. PhenoTips supports importing and export family records in the Phenopacket format through the user interface and REST APIs. This capability now includes the GA4GH Pedigree representation within the Family object. The PhenoTips pedigree editor user interface supports importing and exporting pedigrees in the GA4GH Pedigree format. This supports a direct JSON implementation of the core pedigree structure.

Figure: PhenoTips Implementation Example showing the import and export format options.

`Do you have an implementation to share? Make a pull request via GitHub to add it.`


Templates
Import
Data format:
PhenoTips JSON
Simple JSON
PED or LINKAGE (pre- or post- makeped)
GEDCOM
BOADICEA
GA4GH Pedigree


Pedigree export
Format:
PhenoTips JSON
PED
BOADICEA
GA4GH Pedigree
Image
Options:
All data
Remove personal information (name and age)
Remove personal information and free-form comments
EXPORT
CANCEL

CHAPTER

EIGHT

ACKNOWLEDGEMENTS

This standard was developed by Clinical and Phenotpic Data Capture Work Stream of the GA4GH, and is the result of the collaborative work, comments, and input of many individual and organizational contributors. We thank all contributors for their time and expertise.

8.1 Pedigree Standard Contributors (in alphabetical order)

Louis Bergelson (Broad Institute)
Eva Bermejo (EJP RD)
Chris Bun (CancerIQ)
Orion Buske (PhenoTips)
Hannah Calkins (CHOP)
Chen Chen (GA4GH)
Melissa Cline (UCSC, BRCA Exchange)
Melissa Cook (NCI, CRDC)
Shahim Essaid (OHSU)
Alex Felmeister (Illumina)
Bingjian Feng (Univeristy of Utah, BRCA Exchange)
Dietmar Fernandez (EGA)
Michael Franklin (Centre for Population Genomics, Garvan Institute)
Sean Garin (NIH)
Lisa Glaspie (CancerIQ)
Melissa Haendel (University of Colorado, Monarch Initiative)
David Hansen (CSIRO, Australian Genomics)
Allison Heath (CHOP, Kids First DRC)
Tim Jackson (TrakGene)
Julius Jacobsen (QMUL, Monarch Initiative)
Katherine Johnston (H3Africa)
Meen Chul Kim (CHOP, Kids First DRC)
Guida Landouré (H3Africa)
Steven Laurie (CNAG-CRG)
Tara Lichtenberg (University of Chicago)
Mamana Mbiyavanga (University of Cape Town, H3Africa)
Alejandro Metke (CSIRO, Australian Genomics)
Moni Munoz-Torres (University of Colorado, Monarch Initiative)
Thanh-Phuong Nguyen (Megeno S.A.)

Soichi Ogishima (Tohoku University, GEM-Japan)
Kevin Power (Children's Mercy Hospital)
Peter Robinson (Jackson Laboratory, Monarch Initiative)
Richard Scott (Genomics England)
Natasha Singh (CHOP, D3B)
Neerjah Skantharajah (GA4GH)
Lindsay Smith (GA4GH)
Amanda Spurdle (QIMR Berghofer Medical Research Institute, BRCA Exchange)
Zornitza Stark (Australian Genomics)
Deanne Taylor (CHOP)
Alex Tsai (GA4GH)
Katheryn Van Diemen (TrakGene)
Grant Wood (MyGenomeTrust)
Teruhiko Yoshida (National Cancer Center, Japan)

8.2 Driver Project Survey Participants (if not listed above)

Luca Barcella (EJP RD)
Sergi Beltran (CRG, EJP RD)
Eva Brezinova (EJP RD)
Erwin Brosens (EJP RD)
Candice Feben (DDD-Africa)
Krisztian Gaspar
James Gyamfi (Genomics England)
Vesta Kucinskiene (EJP RD)
Audald Lloret-Villas (EGA)
Balzas Mayer (Semmelweis University)
Anna Need (Genomics England)
Oscar Nyangiri (H3Africa)
Olusola Omosaiye (Genomics England)
Suzanne Pasmans (ErasmusMC, EJP RD)
Alessia Pepe
Anita Rauch (Universität Zürich)
Elzbieta Radzikowska
Marina Vivarelli (EJP RD)
Daryl Waggott (Genome Canada)
Zhenyu Zhang (Univesity of Chicago, NCI GDC)

8.3 Special Thanks To

Robert Freimuth (Mayo Clinic)
Hoa Ngo (CSIRO)

8.4 Funding

J.J. would like to acknowledge National Institutes of Health (NIH) grants 1R24OD011883 and NIH, National Institute of Child Health and Human Development 1R01HD103805-01

A.B.S. was supported by an NHMRC Investigator Fellowship (APP177524)