

# FAIR GENOMES

A JUNE 2020 UPDATE BY **K. JOERI VAN DER VELDE**

FAIR Genomes is a ZonMw “Personalized Medicine” project, nr. 846003201



**ZonMw**

# Who are FAIR Genomes? WP leads

**Gurnoor Singh<sup>1</sup>, K. Joeri van der Velde<sup>2</sup>, Jeroen Beliën<sup>4</sup>, Jasmin Böhmer<sup>3</sup>, Daphne Stemkens<sup>5</sup>, Lisenka Vissers<sup>1</sup>, Jeroen van Reeuwijk<sup>1</sup>, Saskia Hiltemann<sup>7</sup>, Lennart F. Johansson<sup>2</sup>, Nienke van der Stoep<sup>6</sup>, Daoud Sie<sup>4</sup>, Janneke Weiss<sup>4</sup>, Geert Frederix<sup>3</sup>, Marco Roos<sup>6</sup>, Erik van Iperen<sup>8</sup>, Terry Vrijenhoek<sup>3</sup>, Folkert W. Asselbergs<sup>3</sup>, Joris van Montfrans<sup>3</sup>, Rolf Sijmons<sup>2</sup>, Hanneke van Deutekom<sup>3</sup>, Pieter Neerincx<sup>2</sup>, Fernanda de Andrade<sup>2</sup>, Anna Niehues<sup>1</sup>, Hindrik H.D. Kerstens<sup>10</sup>, Mark Thompson<sup>6</sup>, Rajaram Kaliyaperumal<sup>6</sup>, Annika Jacobsen<sup>6</sup>, Katy Wolstencroft<sup>6,14</sup>, Ies Nijman<sup>3</sup>, Marcel Nelen<sup>1</sup>, Ariaan Siezen<sup>1</sup>, Koen ten Hove<sup>1</sup>, Nine Knoers<sup>2</sup>, Christian Gilissen<sup>1</sup>, Hans Scheffer<sup>1</sup>, Stefan Willems<sup>3</sup>, Wendy van Zelst-Stams<sup>1</sup>, Helger Ijntema<sup>1</sup>, Kim Elsink<sup>3</sup>, Bart de Koning<sup>9</sup>, Bauke Ylstra<sup>4</sup>, Erik Sistermans<sup>4</sup>, Patrick Kemmeren<sup>10</sup>, Henne Holstege<sup>4</sup>, Christine Staiger<sup>11</sup>, Bastiaan Tops<sup>10</sup>, Susanne Rebers<sup>12</sup>, David van Zessen<sup>7</sup>, Valesca Retèl<sup>12</sup>, Edwin Cuppen<sup>13</sup>, Peter van Tintelen<sup>3</sup>, Esther van Enckevort<sup>2</sup>, Lieneke Steeghs<sup>1</sup>, Salome Scholtens<sup>2</sup>, Jeroen Laros<sup>6</sup>, Leon Mei<sup>6</sup>, Cor Oosterwijk<sup>5</sup>, Andrew Stubbs<sup>7</sup>, Peter A.C. 't Hoen<sup>1</sup>, Mariëlle van Gijn<sup>2</sup>, Morris Swertz<sup>2</sup>**

***from:***

<sup>1</sup>Radboud University Medical Center, Nijmegen, The Netherlands, <sup>2</sup>University Medical Center Groningen, The Netherlands, <sup>3</sup>University Medical Center Utrecht, The Netherlands, <sup>4</sup>Amsterdam University Medical Centers, location VUmc, NL, <sup>5</sup>VSOP - Dutch Patient Alliance for Rare and Genetic Diseases, <sup>6</sup>Leiden University Medical Center, The Netherlands, <sup>7</sup>Erasmus Medical Center, Rotterdam, The Netherlands, <sup>8</sup>Durrer Center for Cardiovascular Research, Utrecht, The Netherlands, <sup>9</sup>Maastricht University Medical Center, The Netherlands, <sup>10</sup>Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands, <sup>11</sup>Dutch Techcentre for Life Sciences, Utrecht, The Netherlands, <sup>12</sup>Netherlands Cancer Institute, Amsterdam, The Netherlands, <sup>13</sup>Hartwig Medical Foundation, Amsterdam, The Netherlands, <sup>14</sup>Leiden Institute for Advanced Computer Science, Leiden University, Leiden, NL

# Why do we need FAIR Genomes?

**Goal: a national guideline to promote optimal (re)use of NGS data in research and healthcare**

NGS analysis flow →



What was the phenotype of the patient?



What kind of sharing is allowed by the consent?



Which tissue was sampled?



Which sample prep kit was used?



What type of NGS machine was used?



What software was used to perform read mapping?



Which protocol was used to interpret the data?

***if FAIR:***

*“I would like to please have all nationally available VCF files for PBMC samples from cardiomyopathy patients sequenced whole-exome on HiSeq machines processed with GATK 4.0, for which consent allows re-analysis.”*

**Now: not possible.  
Let's make it possible.**

# What is FAIR Genomes?

**61 people** from **14 institutes** (NL). F2F meetings, Zoom calls, focus workshops. **Interacting** with EJP-RD CDE, Solve-RD RD3, 1+MG, GA4GH, Phenopackets, X-omics, and others.

Currently **9 parts with 82 elements**:  
Personal (12), Clinical (16), Material (12), SamplePrep (7), Sequencing (7), Analysis (7), General Consent (8), Individual Consent (7), Study (6).

**Together** we define what meta data is needed to **find, share and reuse** NGS data in research and healthcare. Forming an evolving **semantic** model for properties and values.

Focus on being **harmonized** with EJP-RD Common Data Elements, RD3, PhenoPackets, MIABIS, etc. Model, codebooks and applications all **free & open source software**.

Join us at: <https://github.com/fairgenomes>

# Main result: a semantic model

## Personal information

Meta-data element	Preferred ontology term for meta-data	Value types	Issue Number	Issue (open / closed)
Individual ID	NCIT:C164337	ID [string]	#3	closed
Gender	SIO:010029	Male SIO:010048 Female SIO:010048 Unknown / Undetermined	#4	closed
Genotypic sex	PATO:0020000	UNKNOWN_KARYOTYPE, XX, XY, XO, XXY, XXX, XXYX, XXXY, XXXX, XYY, OTHER_KARYOTYPE	#69	closed
Country of residence	SNOMEDCT:276205001	ISO 3166 country codes	#5	closed
Ethnicity	LNC:MTHU010275	ISO 3166 country codes	#6	closed
Country of Birth	SNOMEDCT:370159000	snomed ontological term [subclass of SNOMEDCT:370159000]	#7	closed
Year of birth (if allowed)	NCIT:C83164	YYYY [4 digits]	#8	closed
Inclusion status	SNOMEDCT:246097001	Alive / Dead / Lost in follow-up / Opted-out	#9	open
Age at death	NCIT:C135383	[Positive integer]	#10	closed
Inclusion criterion	OBI:0500027	Free text [string]	#11	closed
Primary affiliated institute	SIO:000688	List of Dutch institutes in BBMRI-ERIC	#60	open
Data available in other institutes	SIO:000688	List of Dutch institutes in BBMRI-ERIC	#62	open

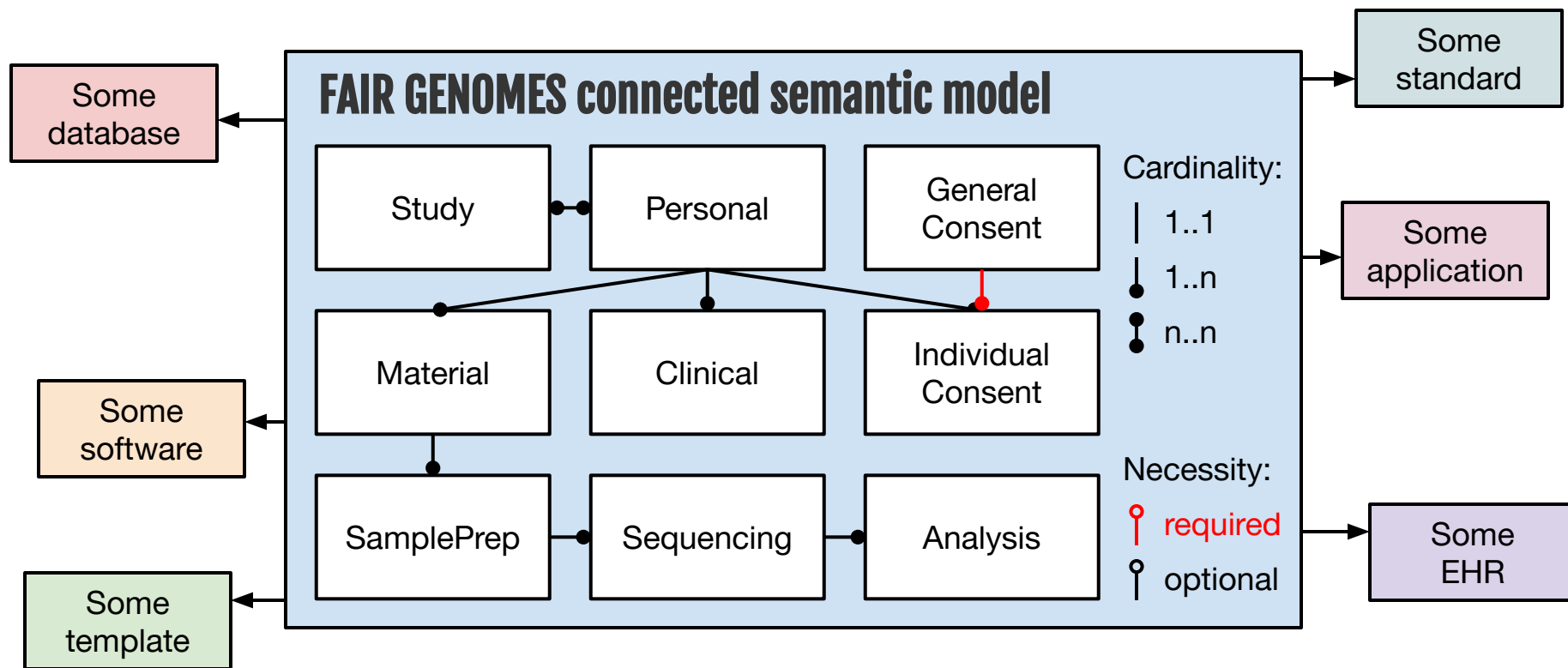
Note: both the elements and values are linked to **ontologies**.

All elements are compulsory, but with HL7 **NullFlavors** to tell why values are missing.

See:

[https://github.com/fairgenomes/information/tree/master/fairgenomes\\_semantic\\_model](https://github.com/fairgenomes/information/tree/master/fairgenomes_semantic_model)

# A model is a means, not a goal



# Application #1: Nictiz

Gurnoor Singh, Jeroen Beliën,  
Sander de Ridder, K. Joeri van der  
Velde

are implementing a FAIR Genomes  
**ART-DECOR** codebook for

**Nictiz** (who develop & manage  
information standards for exchange  
of digital data in healthcare)



<https://www.nictiz.nl/standaardisatie/art-decor/>

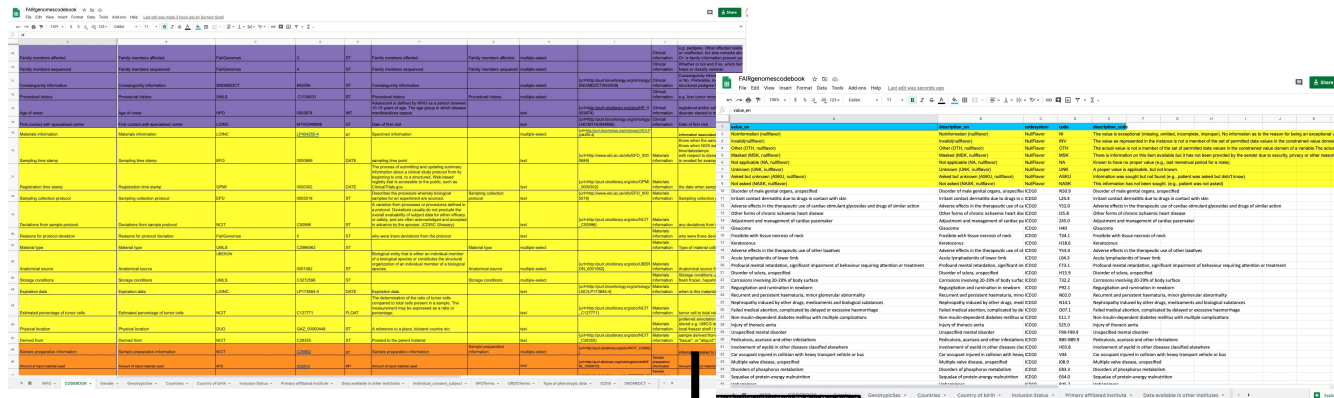


<https://art-decor.org/>

See:

[https://github.com/fairgenomes/information/tree/master/fairgenomes\\_codebook\\_nictiz](https://github.com/fairgenomes/information/tree/master/fairgenomes_codebook_nictiz)

# Why an ART-DECOR codebook?



health **Rijks**  
 The **iCRF Generator**  
 - Generate Interoperable CRFs -  
 Sponsored by KWR (SurT4Health)

Welcome to the Registry in a Box (iCRF Generator)!  
 Select an EDC from the dropdown and press Run

Select EDC  
 Castor  
 REDCap  
 OpenClinica 3

For more useful tools and guidances, keep an eye on the health-ri website (<https://health-ri.org>), which will soon have a Registry in a Box section!

Clear Run Exit Help About

**iCRF Generator** is a Java program that can generate the core of an interoperable electronic **case report form (iCRF)** for several of the major **electronic data capture systems (EDCs)**

de Ridder S and Belien JAM. *The iCRF Generator: Generating interoperable electronic case report forms using online codebooks.* F1000Research 2020, 9:81 (<https://doi.org/10.12688/f1000research.21576.2>)



# Application #2: MOLGENIS

K. Joeri van der Velde, Gurnoor Singh, Fernanda de Andrade, Dieuwke Roelofs-Prins, Lennart F. Johansson, Bart Charbon & MOLGENIS Team

**MOLGENIS:** scientific data platform  
with flexible model, tailor to your  
needs, 100+ instances running



**FAIR Genomes app**

[https://github.com/fairgenomes/information/tree/master/fairgenomes\\_molgenis\\_app](https://github.com/fairgenomes/information/tree/master/fairgenomes_molgenis_app)

**Powering ERNs**

- GENTURIS →
- Ithaca
- Skin
- Cranio

**Solve-RD RD3** ([https://github.com/molgenis/RD3\\_database](https://github.com/molgenis/RD3_database))  
Detailed NGS sample tracking

Value (English)	Description (English)	Codexsystem	Code	Description
16p21.3p21.1 microdeletion syndrome	16p21.3p21.1 microdeletion syndrome	orphanet	ORPHA127613	https://orpha.net/
16p21.3p21.3 microdeletion syndrome	16p21.3p21.3 microdeletion syndrome	orphanet	ORPHA276432	https://orpha.net/
11p22.3p22.3 microdeletion syndrome	11p22.3p22.3 microdeletion syndrome	orphanet	ORPHA444002	https://orpha.net/
12p14.1 microdeletion syndrome	12p14.1 microdeletion syndrome	orphanet	ORPHA9303	https://orpha.net/
14q11.2 microdeletion syndrome	14q11.2 microdeletion syndrome	orphanet	ORPHA261120	https://orpha.net/
14q12.1 microdeletion syndrome	14q12.1 microdeletion syndrome	orphanet	ORPHA261144	https://orpha.net/
16p22.3p22.3 microdeletion syndrome	16p22.3p22.3 microdeletion syndrome	orphanet	ORPHA264200	https://orpha.net/
15q overgrowth syndrome	15q overgrowth syndrome	orphanet	ORPHA143485	https://orpha.net/
16p11.2 microdeletion syndrome	16p11.2 microdeletion syndrome	orphanet	ORPHA235448	https://orpha.net/
16p13.3 microdeletion syndrome	16p13.3 microdeletion syndrome	orphanet	ORPHA180119	https://orpha.net/
16p13.1 microdeletion syndrome	16p13.1 microdeletion syndrome	orphanet	ORPHA261180	https://orpha.net/
16p21.1 microdeletion syndrome	16p21.1 microdeletion syndrome	orphanet	ORPHA9305	https://orpha.net/

Identifier	Tissue Types	QC Inval	organisation	ERN	Sequencing Centre
1	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
2	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
3	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
4	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
5	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
6	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
7	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
8	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
9	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
10	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
11	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
12	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
13	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
14	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
15	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
16	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
17	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
18	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
19	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen
20	Whole Blood	None	skid-synchril	ERN RD	IMMAG-Tuebingen

# Public demo online

FAIR GENOMES Import data ▾ Navigator Data Explorer Plugins ▾



VERSION 0.1 PUBLIC DEMO. THIS IS A PROTOTYPE FOR EVALUATION PURPOSES ONLY. DO NOT IMPORT ANY SENSITIVE DATA.

A national guideline to promote optimal (re)use of NGS data in research and healthcare.



Study



Personal



General Consent



Material



Clinical



Individual Consent



Sampleprep



Sequencing



Analysis



Codebooks



Information



Contribute

**Please visit & give us feedback!**

**<https://molgenis93.gcc.rug.nl>**

FAIR GENOMES Import data ▾ Navigator Data Explorer Plugins ▾

Help Sign in

## CLINICAL

Clinical ID \*

Visit 001

Unique label or human-interpretable identifier for this Clinical record

Phenotypic terms \*

Abnormal B cell count × Abnormality of skull size ×

MESH:D010641

Unobserved phenotypes

10 pairs of ribs × brain

HL7:C0442737

Type of phenotypic data \*

Image × Moving Image ×

DC:DCMIType

Clinical diagnosis

4q21 microdeletion syndrome ×

SNOMEDCT:39154008

Genetic diagnosis

12q14 microdeletion syndrome × syn

Acanthosis nigricans-insulin resistance-muscle cramps-acral enlargement syndrome

Achalasia-microcephaly syndrome

Acral peeling skin syndrome

Acrocallosal syndrome

Acrocardiofacial syndrome

Acrocephalosyndactyly

Acromegaly facial appearance syndrome

# Bonus deliverable: new ontology

We will create a formal 'FairGenomes' ontology for important terms currently **not available** in ontologies

- Genotypic sex (ie. karyotypes, ~10 items)
- NGS kits (~625 items, source: BioCompare)
- Sequencing instruments (adding ~10 items)
- Dutch hospitals (~110 items)

- Discuss open issues with FAIR Genomes & other domain experts
  - ◆ Finalize attributes, descriptions, values, curate codebooks
  - ◆ Further harmonize with other projects & existing standards
- **We need to show that it works in *demonstrator projects*!**
  - ◆ **Testdrive FAIR Genomes on genomics data (1+MG, B1MG, EJP-RD)**
  - ◆ **See if works in daily healthcare & research practice (VKGN, KMBP, ..)**
- **Publish to share our hard work & FAIRify the rest of the world**
- From meta-data to a real semantic data models ?
  - ◆ EMX format (MOLGENIS) works, but is not very semantic yet
  - ◆ Construct true semantic data model (RDF or JSON-LD)
  - ◆ Add rule based validator (RDF Shacl or JSON-LD markups)

Thank you.



# FAIR GENOMES

Time for questions and answers.