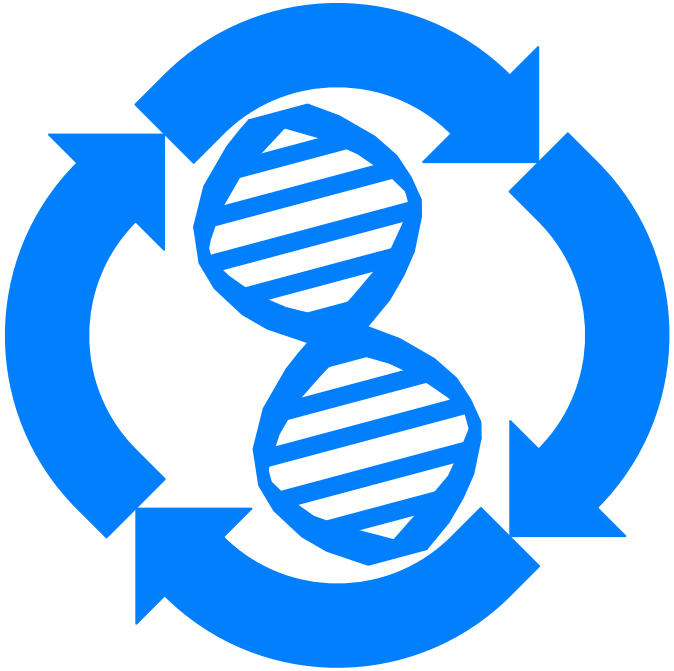


FAIR GENOMES

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[1], K. Joeri van der Velde[2], Jeroen Beliën[4], Jasmin Böhmer[3], Daphne Stemkens[5], Lisenka Vissers[1], Jeroen van Reeuwijk[1], Saskia Hiltemann[7], Lennart F. Johansson[2], Nienke van der Stoep[6], Daoud Sie[4], Janneke Weiss[4], Geert Frederix[3], Marco Roos[6], Erik van Iperen[8], Terry Vrijenhoek[3], Folkert W. Asselbergs[3], Joris van Montfrans[3], Rolf Sijmons[2], Hanneke van Deutekom[3], Pieter Neerincx[2], Fernanda de Andrade[2], Anna Niehues[1], Hindrik H.D. Kerstens[10], Mark Thompson[6], Rajaram Kaliyaperumal[6], Annika Jacobsen[6], Katy Wolstencroft[6,14], Ies Nijman[3], Marcel Nelen[1], Ariaan Siezen[1], Koen ten Hove[1], Nine Knoers[2], Christian Gilissen[1], Hans Scheffer[1], Stefan Willems[3], Wendy van Zelst-Stams[1], Helger Ijntema[1], Kim Elsink[3], Bart de Koning[9], Bauke Ylstra[4], Erik Sistermans[4], Patrick Kemmeren[10], Henne Holstege[4], Christine Staiger[11], Bastiaan Tops[10], Susanne Rebers[12], David van Zessen[7], Valesca Retèl[12], Edwin Cuppen[13], Peter van Tintelen[3], Esther van Enckevort[2], Lieneke Steeghs[1], Salome Scholtens[2], Jeroen Laros[6], Leon Mei[6], Cor Oosterwijk[5], Andrew Stubbs[7], Peter A.C. ‘t Hoen[1], Mariëlle van Gijn[2], Morris Swertz[2]

from:

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

Why do we need FAIR Genomes?

umcg Dept. of Genetics
rijksuniversiteit groningen

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?

umcg Dept. of Genetics
rijksuniversiteit groningen

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.

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.

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).

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

umcg Dept. of Genetics
rijksuniversiteit groningen

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, XXYY, 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

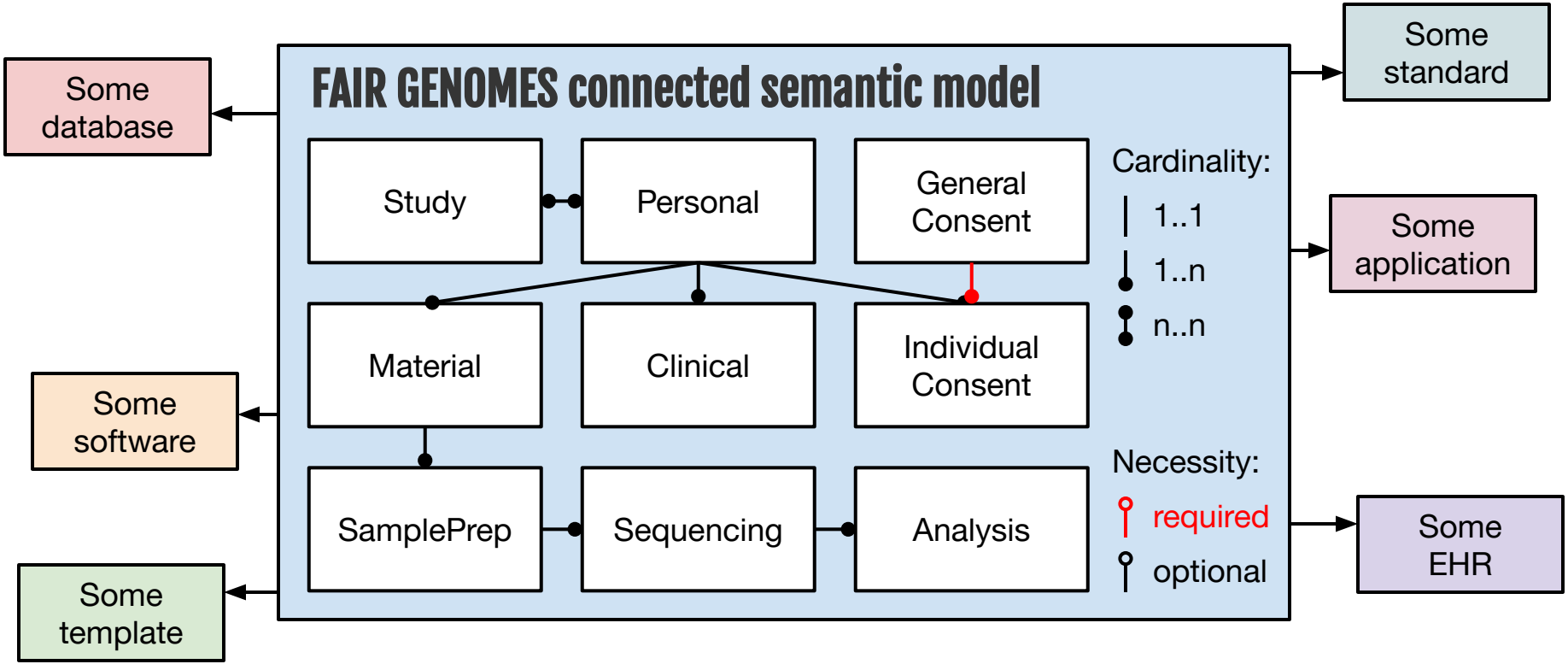
Note: both the elements and values are linked to **ontologies**.

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

A model is a means, not a goal

umcg Dept. of Genetics
rijksuniversiteit groningen



Application #1: Nictiz

umcg
Dept. of Genetics
rijksuniversiteit groningen

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)

Nictiz

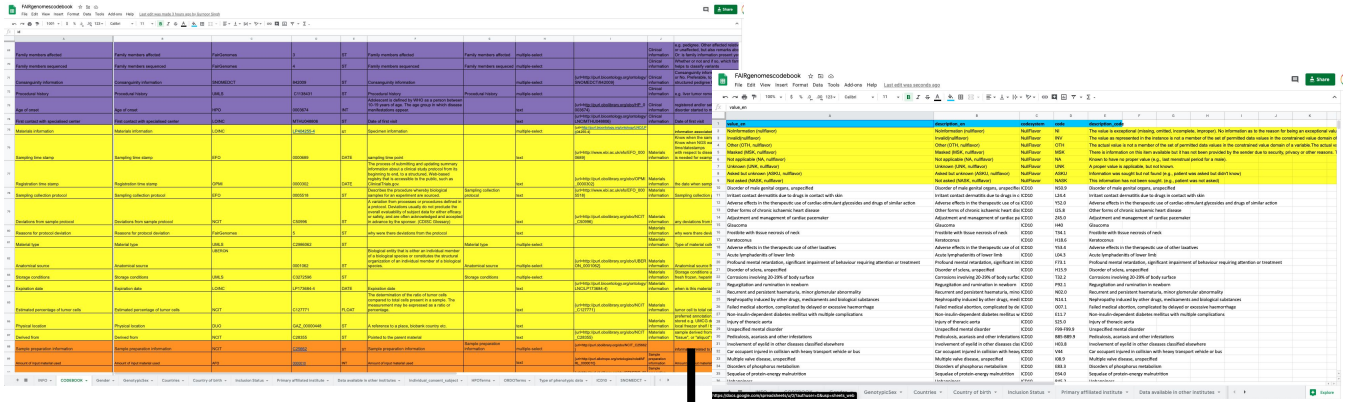
https://www.nictiz.nl/standaardisatie/art-decor/

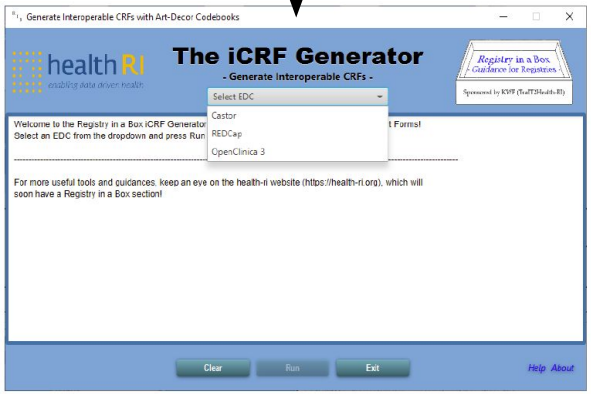
https://art-decor.org/

See:
https://github.com/fairgenomes/information/tree/master/fairgenomes_codebook_nictiz

Why an ART-DECOR codebook?

umcg Dept. of Genetics
rijksuniversiteit groningen





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 Beliën 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

umcg Dept. of Genetics
rijksuniversiteit groningen

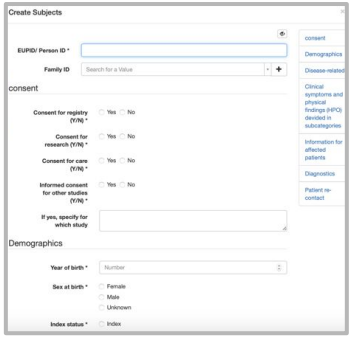
K. Joeri van der Velde, Gurnoor Singh, Fernanda de Andrade, Dieuwke Roelofs-Prins, Lennart F. Johansson, Bart Charbon & MOLGENIS Team

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

MOLGENIS.org

Powering ERNs

- GENTURIS →
- Ithaca
- Skin
- Cranio

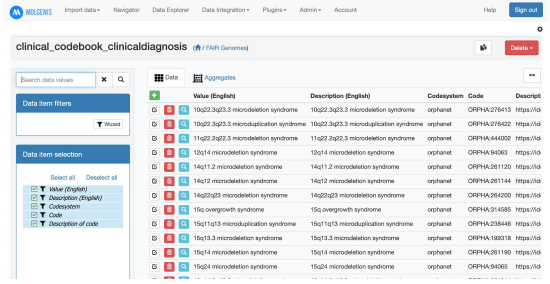


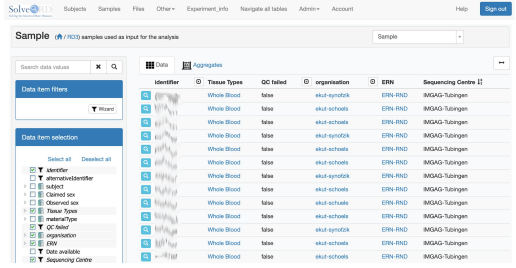
FAIR Genomes app

https://github.com/fairgenomes/information/tree/master/fairgenomes_molgenis_app

Solve-RD RD3 (https://github.com/molgenis/RD3_database)

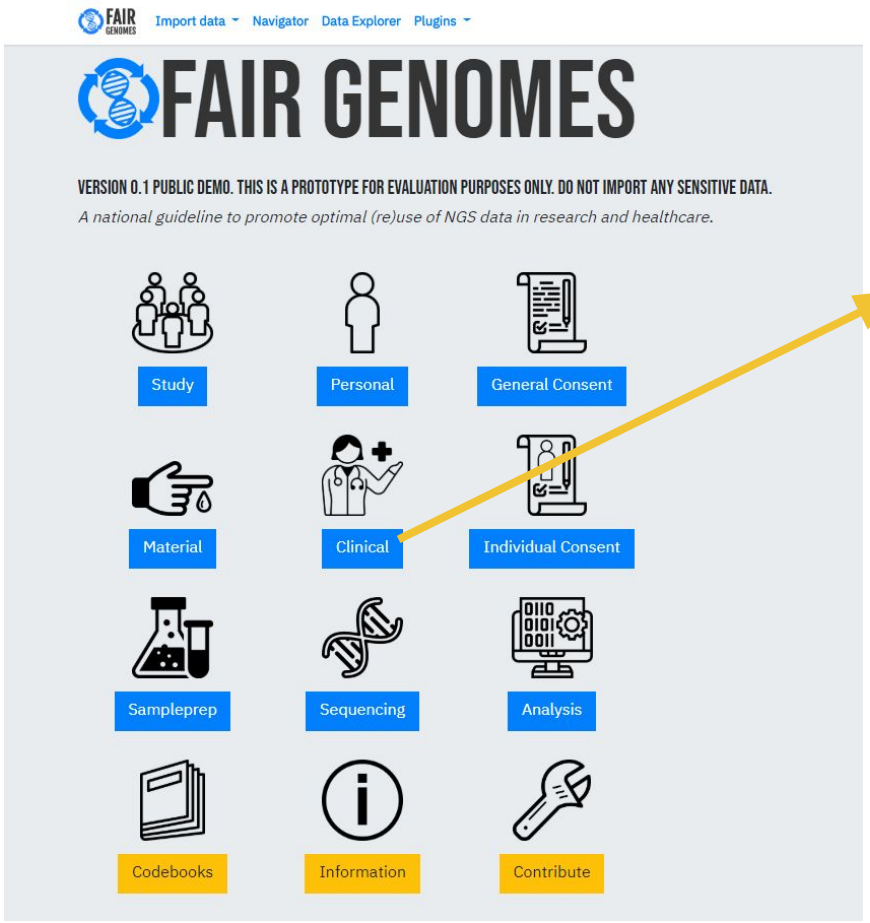
Detailed NGS sample tracking





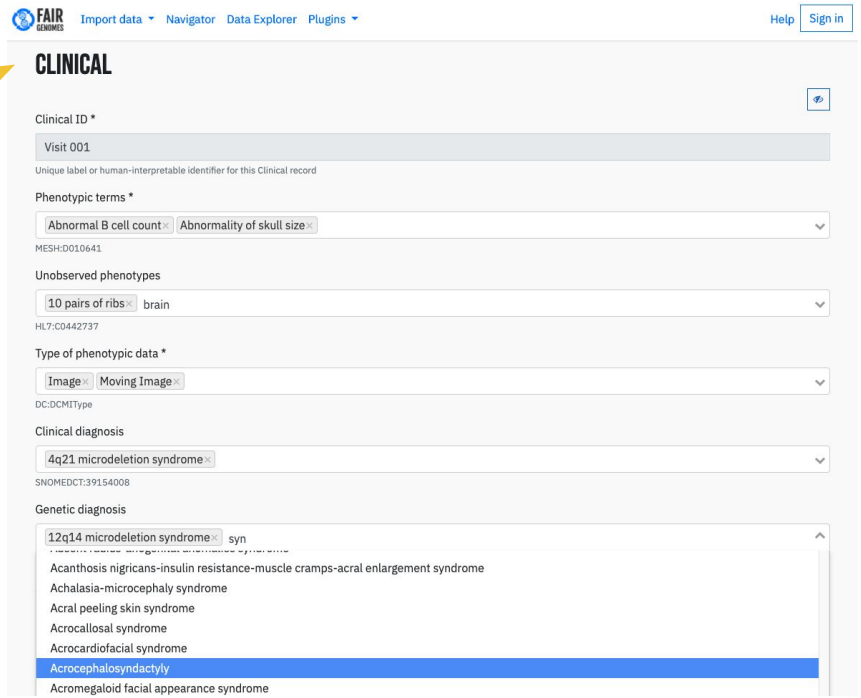
Public demo online

umcg Dept. of Genetics
rijksuniversiteit groningen



Please visit & give us feedback!

https://molgenis93.gcc.rug.nl



Bonus deliverable: new ontology

umcg Dept. of Genetics
rijksuniversiteit groningen

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)

Future work

umcg
Dept. of Genetics
rijksuniversiteit groningen

- 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 anwers.