



# WELCOME: DAY 3

## Session 4

### FAIRifier



## FAIR DATA STEWARDSHIP COURSE

TOWARDS A GO FAIR READINESS PROGRAM



## RECAP – FAIRIFIER – MAJOR FUNCTIONALITY BLOCKS

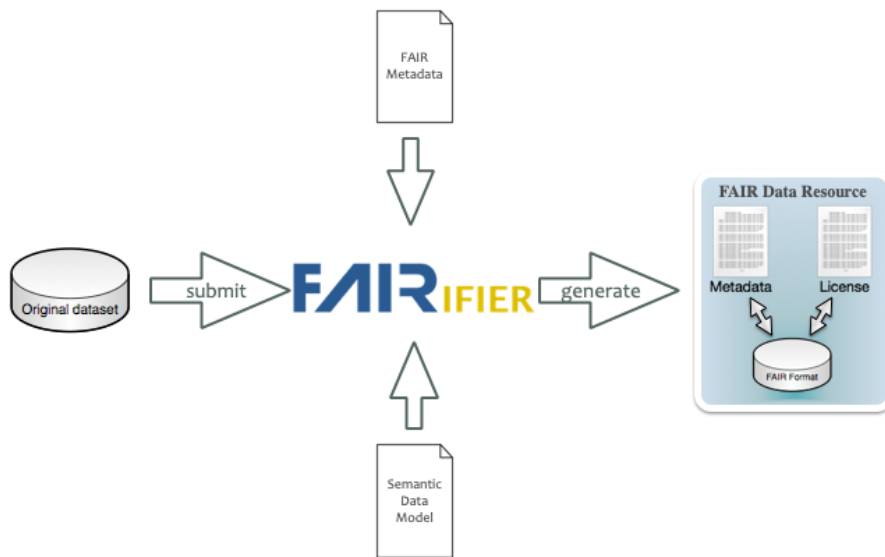
**Data Wrangling**

**Semantic Data  
Modeling**

**Metadata Definition  
and Extraction**

**FAIR Publication**

FAIR



FAIR

## DATA FAIRIFICATION

- **Dataset:**

- Name: GoNL vcf dump
- URL: [https://molgenis26.target.rug.nl/downloads/gonl\\_public/variants/release5/](https://molgenis26.target.rug.nl/downloads/gonl_public/variants/release5/)
- Description: [http://www.nlgenome.nl/?page\\_id=9](http://www.nlgenome.nl/?page_id=9)
- Format: Variant Call Format (VCF) 4.1: <http://samtools.github.io/hts-specs/VCFv4.1.pdf>
- Size: 27.896 rows of data

2024

## RETRIEVE NON-FAIR DATA

Retrieve  
non-FAIR  
data

Download compressed file from [https://molgenis26.target.rug.nl/downloads/gonl\\_public/variants/releases/gonl.SV.r5.vcf.gz](https://molgenis26.target.rug.nl/downloads/gonl_public/variants/releases/gonl.SV.r5.vcf.gz)

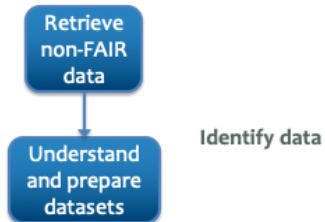
Extract file gonl.SV.r5.vcf

# ANALYZE AND PREPARE DATASETS



DATA

## ANALYZE AND PREPARE DATASETS



```
##fileformat=VCFv4.1
##ALT=<ID=DEL,Description="Deletion">
##FILTER=<ID=Inaccessible,Description="Overlaps a user-input mask">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype quality (#subdomain
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT a
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in cal
##INFO=<ID=CIEND,Number=2,Type=Integer,Description="Confidence interval around
##INFO=<ID=CIP0S,Number=2,Type=Integer,Description="Confidence interval around
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP Membership">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some r
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant s
```

## ANALYZE AND PREPARE DATASETS



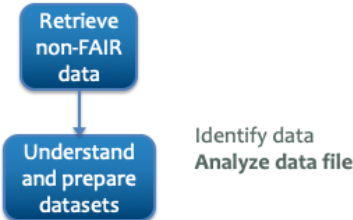
Identify data

VCFv4.1: <http://samtools.github.io/hts-specs/VCFv4.1.pdf>

```
##fileformat=VCFv4.1
##ALT=<ID=DEL,Description="Deletion">
##FILTER=<ID=Inaccessible,Description="Overlaps a user-input mask">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype quality (#subdomain
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT a
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in cal
##INFO=<ID=CIEND,Number=2,Type=Integer,Description="Confidence interval around
##INFO=<ID=CIPQS,Number=2,Type=Integer,Description="Confidence interval around
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP Membership">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some r
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant s
```



# ANALYZE AND PREPARE DATASETS



# ANALYZE AND PREPARE DATASETS



Identify data  
Analyze data file

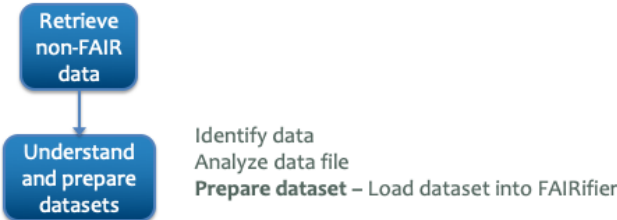
Although not explicit, in row 108 we can find information about the reference file, which allowed us to *infer* that the variant information refers to the human reference genome version 19

According to the VCF spec, the #CHROM column is required and refers to the identifier of the chromosome where the variant occurred

Actual data starting at row 113 with headers at row 112

```
103 ##contig=GL000193.1,length=189789,assembly=b37
104 ##contig=GL000194.1,length=191469,assembly=b37
105 ##contig=GL000225.1,length=211173,assembly=b37
106 ##contig=GL000192.1,length=547496,assembly=b37
107 ##fileDate=20130429
108 ##reference=file:///home/cog/lfrancioli/resources/hg19/human_g1k_v37.fa
109 ##source=GATKUnifiedGenotyper_2.1.8,Pindel_0.2.4t,MATE-CLEVER_rev.3097f2,1235V_0.9,Breakd
110 ##source=SelectVariants
111 ##source_20130804.1=vcf-annotate(r797) -r INFO/TDT -r INFO/ADOS -r INFO/pop -r INFO/Calie
112 #CHROM POS ID REF ALT QUAL FILTER INFO
113 1 776769 . T <DEL> . PASS AC=8;AF=5.028e-03;AN=1534;CIEND=0,708;CIPOS=-110;
114 1 829169 rs201003251 AAAAAAAAAAATATATATATATATATATAT A . Inaccessible AI
115 1 869368 . CGCGTGGTTACAGTGGGAGGGGGGCGGCTTACAGTGGGAGGGGAGGGGCGGCTTACAGTGG
116 1 943126 rs34521632 CTTTTTTTTTTTTTTTTTTT C . PASS AC=121;AF=0.082;AN=15;
117 1 965017 . GTGTGTGCAGTGATGCTGCTGATGATGAGATGAGATGAGTGTGTGTGTGTGTGCGAGTGTGTGTGATC
118 1 988572 . GGTGCTGCAGTGGTGTGCTGATGTTGGTGTCTGTGCCCTCAAGTGTCTC G . PASS AI
119 1 997436 rs145846158 CTCCCTCCCTTTGCCCGTTCCCTCCG C . Inaccessible AC=382;AF=0.21
120 1 1044488 rs142246657 ACCACAGCCAAAGGTGGGAGCAGGTGCCAC A . PASS AC=127;AF=0.01
121 1 1058898 . CGCACACGCCACACACTCGGCACACTCTGCACACACAGTGCACACACTGCACACACTCTGCACAC
122 1 1142719 . GGAGACTGCTTATGTTCTTTCTGAGCTCAGTTCCCTGTGGGACCCAGGGGTTCTGGACCTGCCTCA
123 1 1161100 rs3834014 CGCCACGACAGGCGCCACACTCCACA C . Inaccessible AC=60;AF=1
124 1 1162672 . GCGGGAGGGGAGCTGTGGCCAGGCTCGGGAGGGAGCTCGTGGCCAGGCCGGGGAGGGGAGC
```

# ANALYZE AND PREPARE DATASETS




FAIR

# ANALYZE AND PREPARE DATASETS

The screenshot displays the FARIFIER web application interface. At the top, there is a navigation bar with 'Create Project', 'Start Over', and 'Configure Parsing Options' buttons. The main area shows a list of projects with columns for 'ID', 'REF', and 'CHARS'. Below the list, there is a 'Parse data as' section with various file format options like 'CSV / TSV / separator-based file', 'JSON files', and 'XML files'. The 'CSV / TSV / separator-based file' option is selected, and its configuration panel is visible. This panel includes a 'Character encoding' dropdown set to 'UTF-8', a 'Columns are separated by' dropdown set to 'Tab (TSV)', and several checkboxes for parsing options: 'Ignore first 111 line(s) at beginning of file', 'Parse first 1 line(s) as column headers', 'Quote empty cells', 'Load at most 0 row(s) of data', 'Parse cell text into numbers, dates...', 'Quote marks are used to enclose cells containing column separators', 'Store blank rows', 'Store blank cells as null', 'Store file source (file names, URLs) in each row', and 'Update Preview'.

## ANALYZE AND PREPARE DATASETS

From VCF spec: “ID - identifier: Semi-colon separated list of unique identifiers where available. If this is a dbSNP variant it is encouraged to use the rs number(s).”



All	#CHROM	POS	ID	REF
★	1.	776769	.	T
★	2.	829169	rs201003251	AAAAAAAAAAAAATATATATATATATAT/
★	3.	869368	.	CGGCTGCGTTACAGGTGGGCAGGGGAGG
★	4.	943126	rs34521632	CTTTTTTTTTTTTTTTTTTTTTTTTTTT
★	5.	965017	.	GTGTGTGTGCAGTGCATGGTGTGTGAGA
★	6.	988572	.	GGTGTCTGCACGTGGGTGTCTGCATGTGC
★	7.	997436	rs145846158	CTCCCTCCCTTGTCCCGTCCCTCCG
★	8.	1044488	rs142246657	ACCACAGCCAAAAGTGGGAGCAAGTGT
★	9.	1058898	.	CGCACACGCCACACACCTGCGCACACT
★	10.	1142719	.	GGAGACTGTCTATGTCTTTCTGAGCCTC/
★	11.	1161100	rs3834014	CGCCACAGACACGGGCCACACTCCAC/

## ANALYZE AND PREPARE DATASETS



Identifiers.org

Search

[Home](#) | [Documentation](#) | [Services](#) | [About](#)

Advanced search

[nucleotide](#)

[gene](#)

[Feedback](#)

### Data collection: *dbSNP*

[Overview](#)

[Miscellaneous](#)

[RDF/XML](#)

[Turtle](#)

#### General information

Recommended name	dbSNP
Description	The dbSNP database is a repository for both single base nucleotide substitutions and short deletion and insertion polymorphisms.
Identifier pattern	<code>^rs\d+\$</code>
Registry identifier	MIR:0000161

#### Identification schemes

Namespace	dbsnp
URI	<code>https://identifiers.org/dbsnp</code>

## ANALYZE AND PREPARE DATASETS

Custom text transform on column ID

Expression Language Python / Jython

```
if value == ".":  
    raise Exception  
return value
```

No syntax error.

Preview History Starred Help

row	value	if value == ".": raise Exception return value
1.	.	Error: Traceback (most recent call last): File "<string>", line 3, in __temp_1639243256__ Exception
2.	rs201003251	rs201003251
3.	.	Error: Traceback (most recent call last): File "<string>", line 3, in __temp_1639243256__ Exception
4.	rs34521632	rs34521632
5.	.	Error: Traceback (most recent call last): File "<string>", line 3, in __temp_1639243256__ Exception

On error  keep original  set to blank  store error

Re-transform up to 10 times until no change

OK Cancel

DBSNP URI preparation: remove the . (VCF blank) from the ID column

```
if value == ".":  
    raise Exception  
return value
```

On error: set to blank

## ANALYZE AND PREPARE DATASETS

27896 rows

Show as: **rows** records Show: 5 10 25 50 rows

All	#CHROM	POS	ID	REF
1.	1	776769		
2.	1	829169		AAAAAAAAATATATATATATATATATATATAT
3.	1	869368		CGTTACAGGTGGGCAGGGGAGGCGGCTGCGTTACAGGT
4.	1	943126		
5.	1	965017		TGTGTG
6.	1	988572		CCCTCA
7.	1	997436		
8.	1	1044488		
9.	1	1058898		CAGTG
10.	1	1142719		TGGGC
11.	1	1161100	rs3834014	CGCCAC
12.	1	1162672		GGCGG
13.	1	1164440		CGCCTT
14.	1	1183415		CCTGTG
15.	1	1194505		ACGTGT
16.	1	1223700		T

DBSNP URI preparation: create a new column for the DBSNP URIs



## ANALYZE AND PREPARE DATASETS

**Add column based on column ID**

New column name:

set to blank  store error  copy value from original column

Expression:  Language:  No syntax error.

**Preview** History Starred Help

row	value	return "http://identifiers.org/dbsnp/" + value
1.	null	Error: Traceback (most recent call last): File "<string>", line 2, in __temp_1235182231__ TypeError: cannot concatenate 'str' and 'NoneType' objects
2.	rs201003251	http://identifiers.org/dbsnp/rs201003251
3.	null	Error: Traceback (most recent call last): File "<string>", line 2, in __temp_1235182231__ TypeError: cannot concatenate 'str' and 'NoneType' objects
4.	rs34521632	http://identifiers.org/dbsnp/rs34521632

OK Cancel

DBSNP URI creation: concatenate the identifiers.org URL with the rs# values

```
return "http://identifiers.org/dbsnp/" + value
```

## ANALYZE AND PREPARE DATASETS

#CHROM	POS	ID	DBSNP	REF
1	776769			T
1	829169	rs201003251	<a href="http://identifiers.org/dbsnp/rs201003251">http://identifiers.org/dbsnp/rs201003251</a>	AAAAAAAAAAAAATATATATATATATATATATATATAT
1	869368			CGGCTGCGTTACAGGTGGGCAGGGAGGCGGCTGCGTTACAGGTGGGCAGG
1	943126	rs34521632	<a href="http://identifiers.org/dbsnp/rs34521632">http://identifiers.org/dbsnp/rs34521632</a>	CTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
1	965017			GTGTGTGTGCAGTGCATGGTGTGTGAGATCAGCATGTGTGTGTGTGTGTGC
1	988572			GGTGTCTGCACGTGGGTGTCTGCATGTGGGTGTCTGTGCCCTCAAGTGTCTC
1	997436	rs145846158	<a href="http://identifiers.org/dbsnp/rs145846158">http://identifiers.org/dbsnp/rs145846158</a>	CTCCCTCCCTTGTCCTCCCTCCCTCCG
1	1044488	rs142246657	<a href="http://identifiers.org/dbsnp/rs142246657">http://identifiers.org/dbsnp/rs142246657</a>	ACCACAGCCAAAAGGTGGGAGCAAGTGCCAC
1	1058898			CGCACAGCCACACACACCTGCGCACACTCCTGCACACACAGTGCACACACC
1	1142719			GGAGACTGTCCTATGTCTTTCTGAGCCTCAGTTTCCCTGTGGGCACCGAGG
1	1161100	rs3834014	<a href="http://identifiers.org/dbsnp/rs3834014">http://identifiers.org/dbsnp/rs3834014</a>	CGCCACAGACACGGCCACACACTCCACA
1	1162672			GGCGGGAAGGCGAGCTCGTGGCCAGGCCCTGCGGGAAGGCGAGCTCGTGC
1	1164440			CGCCTTCTCCAGACCACAGTGGCAC
1	1183415			CCTGTGACTCTAGACAGAACAGGCTGGATCTCCAACCTGACCTACAGGCAG
1	1194505			ACGTGTGTGCATGCCATCAGGAC

2/20

# ANALYZE AND PREPARE DATASETS

## Reference SNP (rs) Report

Download [f](#) [t](#) [g+](#) [v](#)

[Switch to classic site](#)

**rs145846158**

Current Build 152  
Released October 2, 2018

FEEDBACK

<b>Organism</b>	Homo sapiens	<b>Clinical Significance</b>	Not Reported in ClinVar
<b>Position</b>	chr1:1062057-1062088 (GRCh38.p12) <a href="#">i</a>	<b>Gene : Consequence</b>	LOC100288175 : Intron Variant
<b>Alleles</b>	delCCTTGT(C) <sub>6</sub> GTTCCCTCCGTCCTC	<b>Publications</b>	0 citations
<b>Variation Type</b>	Indel Insertion and Deletion	<b>Genomic View</b>	<a href="#">See rs on genome</a>
<b>Frequency</b>	delCCTTGT(C) <sub>6</sub> GTTCCCTCCGTCCTC=0.823 (4121/5008, 1000G) delCCTTGT(C) <sub>6</sub> GTTCCCTCCGTCCTC=0.880 (3390/3854, ALSPAC) delCCTTGT(C) <sub>6</sub> GTTCCCTCCGTCCTC=0.877 (3251/3708, TWINSUK)		

### Variant Details

- Clinical Significance
- Frequency
- Aliases
- Submissions
- History
- Publications

### Genomic Placements

Sequence name	Change
GRCh37.p13 chr 1	NC_000001.10.g.997443_997468del
GRCh38.p12 chr 1	NC_000001.11.g.1062063_1062088del

**Gene: LOC100288175, uncharacterized LOC100288175 (plus strand)**

Molecule type	Change	Amino acid[Codon]	SO Term
LOC100288175 transcript	NR_148960.1:n.	N/A	Intron Variant

## ANALYZE AND PREPARE DATASETS

▼ All	▼ #CHROM	▼ POS	▼ ID	▼ REF
☆	1	776769	.	T
☆	2	829169	rs201003251	AAAAAAAAAAAAATATATATATATATATAT/
☆	3	869368	.	CGGCTGCGTTACAGGTGGGCAGGGGAGG
☆	4	943126	rs34521632	CTTTTTTTTTTTTTTTTTTTTTTTTTTT
☆	5	965017	.	GTGTGTGTGCAGTGCATGGTCTGTGAGA
☆	6	988572	.	GGTGTCTGCACGTGGGTGTCTGCATGTGC
☆	7	997436	rs145846158	CTCCCTCCCTTGTCCTCCGTTCCCTCCG
☆	8	1044488	rs142246657	ACCACAGCCAAAAGGTGGGAGCAAGTGTG
☆	9	1058898	.	CGCACACGCCACACACACCTGCGCACAC
☆	10	1142719	.	GGAGACTGTCCTATGTCCTTCTGAGCCTC/
☆	11	1161100	rs3834014	CGCCACAGACACGGGCCACACACTCCAC/

20/20

# ANALYZE AND PREPARE DATASETS

## National Cancer Institute Thesaurus

Last uploaded: March 25, 2019



Summary Classes Properties Notes Mappings Widgets

Jump to:

- Abnormal Cell
- Activity
- Anatomic Structure, System, or Substance
  - Body Cavity
  - Body Fluid or Substance
  - Body Part
  - Body Region
  - Embryonic Structure or System
  - Microanatomic Structure
  - Cell
    - Extracellular Space
    - Genomic Feature Physical Location
    - Macromolecular Structure
      - Acetyl Group
      - Amine Group
      - Base Pair
      - Carboxyl Group
    - Chromatin Structure
    - Cross Link
    - DNA Structure
      - A-DNA
      - B-DNA
      - Chromosome
        - Autosome
        - Human Chromosome
          - Chromosome 1**
          - Chromosome 10
          - Chromosome 11
          - Chromosome 12
          - Chromosome 13
          - Chromosome 14
          - Chromosome 15
          - Chromosome 16

Details Visualization Notes ( 0 ) Class Mappings ( 17 )

Preferred Name	Chromosome 1
Synonyms	Chromosome 1
Definitions	The designation for each member of the largest human autosomal chromosome pair. Chromosome 1 spans about 247 million nucleotide base pairs and represents about 8% of the total DNA in normal diploid cells.
ID	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204</a>
code	C13204
Concept_In_Subset	<a href="#">CTRP Terminology</a> <a href="#">CTRP Biomarker Terminology</a> <a href="#">CTRP Molecular Genetic Biomarker Terminology</a>
Contributing_Source	CTRP
DEFINITION	The designation for each member of the largest human autosomal chromosome pair. Chromosome 1 spans about 247 million nucleotide base pairs and represents about 8% of the total DNA in normal diploid cells.
Display_Name	Chromosome 1
FULL_SYN	Chromosome 1
label	Chromosome 1
Legacy_Concept_Name	Chromosome_1
Preferred_Name	Chromosome 1
prefixIRI	ncit:C13204
Semantic_Type	Cell Component

## **ANALYZE AND PREPARE DATASETS**

- We should try to match the data we have (in the #CHROM column) with the correspondent chromosome identifiers;
- This is the role of the reconciliation service;
- The goal is to reconcile our chromosome names with the chromosome labels from the NCI Thesaurus.
  - However, there is no public SPARQL endpoint offered by NCI;
  - The BioPortal SPARQL endpoint is bandwidth limited;
- One possible solution is to create our own reconciliation service based on a RDF file.

21/22

## ANALYZE AND PREPARE DATASETS

National Cancer Institute Thesaurus  
Last updated: February 25, 2019

Summary Classes Properties Notes Mappings Widgets

Jump to:

- Abnormal Cell
- Activity
- Anatomic Structure, System, or Substance
- Body Cavity
- Body Fluid or Substance
- Body Part
- Body Region
- Embryonic Structure or System
- Microanatomic Structure
- Cell
  - Extracellular Space
  - Genomic Feature Physical Location
  - Macromolecular Structure
    - Acetyl Group
    - Amine Group
    - Base Pair
    - Carboxyl Group
  - Chromatin Structure
  - Cross Link
  - DNA Structure
    - A-DNA
    - B-DNA
    - Chromosome
      - Autosome
        - Human Chromosome
          - Chromosome 1**
          - Chromosome 10
          - Chromosome 11
          - Chromosome 12
          - Chromosome 13
          - Chromosome 14
          - Chromosome 15
          - Chromosome 16
          - Chromosome 17
          - Chromosome 18
          - Chromosome 19
          - Chromosome 2
          - Chromosome 20
          - Chromosome 21
          - Chromosome 22
          - Chromosome 3
          - Chromosome 4
          - Chromosome 5
          - Chromosome 6
          - Chromosome 7
          - Chromosome 8
          - Chromosome 9
          - Chromosome X
          - Chromosome Y
        - Interphase Chromosome
        - Metaphase Chromosome
        - Pre-Condensed Chromosomes
        - Sex Chromosome
        - DNA Replication Fork

Details Visualization Notes ( 0 ) Class Mappings ( 17 )

Preferred Name	Chromosome 1
Synonyms	Chromosome 1
Definitions	The designation for each member of the largest in normal diploid cells.
ID	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus</a>
code	C13204
Concept_In_Subset	CTRP Terminology CTRP Biomarker Terminology CTRP Molecular Genetic Biomarker Terminology
Contributing_Source	CTRP
DEFINITION	The designation for each member of the largest in normal diploid cells.
Display_Name	Chromosome 1
FULL SYN	Chromosome 1
label	Chromosome 1
Legacy_Concept_Name	Chromosome_1
Preferred_Name	Chromosome 1
prefixIRI	ncit:C13204
Semantic_Type	Cell Component
UMLS_CUI	C0008651
subClassOf	Human Chromosome

From the BioPortal [UI](#) we already know that all chromosomes are subclasses of the class “Human Chromosome” with URI:

<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13203>

We can also see that the chromosome label is in the format “Chromosome + the chromosome number/letter.

## ANALYZE AND PREPARE DATASETS

**Add column based on column #CHROM**

New column name:

set to blank  store error  copy value from original column

Expression:   
Language:  No syntax error.

[Preview](#) [History](#) [Starred](#) [Help](#)

row	value	return "Chromosome " + str(value)
1.	1	Chromosome 1
2.	1	Chromosome 1
3.	1	Chromosome 1
4.	1	Chromosome 1
5.	1	Chromosome 1
6.	1	Chromosome 1

**Chromosome URI preparation: add Chromosome ID column appending "Chromosome"**

2/2/2014



## ANALYZE AND PREPARE DATASETS

BioPortal SPARQL  
Beta

Feedback

BioPortal SPARQL is a service to query BioMedical ontologies using the SPARQL standard. Ontologies have been transformed into RDF triples from their original formats (OWL, OBO and UMLS/RRF, ...) and asserted into a triple store.

**Notice:** This SPARQL endpoint is maintained by NCBO for demo purposes. It serves as playground to explore BioPortal's ontologies in RDF and we do not recommend its use for production applications or heavy batch processing. As an alternative, our virtual appliance is packaged with a SPARQL endpoint that can be used for local deployments.

Documentation

SPARQL Examples

```

1 PREFIX owl: <http://www.w3.org/2002/07/owl#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3
4 CONSTRUCT {
5   ?s ?p ?o .
6 } WHERE {
7   GRAPH <http://bioportal.bioontology.org/ontologies/NCIT> {
8     ?s rdfs:subClassOf <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13203> .
9     ?s ?p ?o .
10  }
11 }
12

```

Results: Browse  reset Database:  ontologies  mappings

```

<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P106> "Cell Component"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P366> "Chromosome_13"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208> <http://www.w3.org/2000/01/rdf-schema#subClassOf> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208*^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P108> "Chromosome 13"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#Class> .
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P325> "ncicp:ComplexDefinition xmlns:ncicp="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P207"> "C0008655"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13208> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P97> "The designation for each member of the thirteen chromosomes of the human karyotype."^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13207> <http://www.w3.org/2000/01/rdf-schema#label> "Chromosome 13"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13207> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P106> "Cell Component"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13207> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P366> "Chromosome_12"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13207> <http://www.w3.org/2000/01/rdf-schema#subClassOf> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13207*^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13207> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#P108> "Chromosome 12"^^<http://www.w3.org/2001/XMLSchema#string>
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13207> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#Class> .

```

Chromosome URI preparation: run the SPARQL query on BioPortal SPARQL endpoint

Run the SPARQL query on <http://sparql.bioontology.org/>:

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

CONSTRUCT {

  ?s ?p ?o

} WHERE {

  GRAPH <http://bioportal.bioontology.org/ontologies/NCIT> {

    ?s rdfs:subClassOf

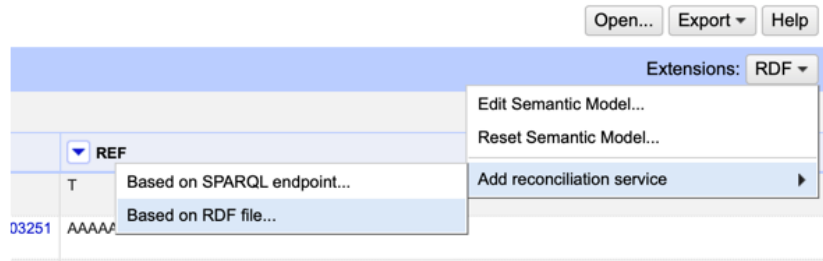
    <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13203> .

    ?s ?p ?o .

  }

}

## ANALYZE AND PREPARE DATASETS



**Chromosome URI preparation:** add a new reconciliation service based on our RDF file

2024

## ANALYZE AND PREPARE DATASETS

### Add file-based reconciliation service

This will set up a new reconciliation service based on an RDF file that provides entity URIs and entity labels.

**Name:**   
A human readable name

#### File details

Load file from URL:

Upload file:  chromosome\_names.ttl

#### Label properties

Select properties that are used to label resources in the RDF data. These properties will be used to match resources:

rdfs:label    skos:prefLabel    dcterms:title    dc:title  
 foaf:name  
 Other...

**Chromosome URI preparation:** add a new reconciliation service based on our RDF file

2/2/2024

## ANALYZE AND PREPARE DATASETS

**Reconcile column "CHROMID"** » Access Service API

Reconcile each cell to an entity of one of these types:

- owl:Class  
http://www.w3.org/2002/07/owl#Class

Also use relevant details from other columns:

Column	Include? As Property
#CHROM	<input type="checkbox"/>
POS	<input type="checkbox"/>
ID	<input type="checkbox"/>
DBSNP	<input type="checkbox"/>
REF	<input type="checkbox"/>
ALT	<input type="checkbox"/>
QUAL	<input type="checkbox"/>
FILTER	<input type="checkbox"/>
INFO	<input type="checkbox"/>

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Maximum number of candidates to return

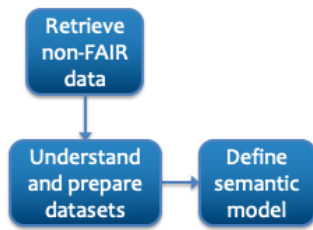
**Chromosome URI preparation: start reconciliation using our new reconciliation service**

## ANALYZE AND PREPARE DATASETS

#CHROM	CHROMID	POS	ID	DBSNP	REF
1		776769			T
1		829169	rs201003251	<a href="http://identifiers.org/dbSNP/rs201003251">http://identifiers.org/dbSNP/rs201003251</a>	AAAAAA
1	Chromosome 1				CGGGTC
1	Chromosome 1	943126	rs34521632		T

**Chromosome URI preparation: match all cells to their best candidate**

## FAIRIFICATION WORKFLOW



FAIR

# DEFINE SEMANTIC MODEL

## RDF Schema Alignment

The Semantic model below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the model. Configure the model by specifying which column to substitute into which node.

**Base URI:** <http://www.mydomain.org/variants/> [edit](#)

**Semantic model** [RDF Preview](#)

Available Prefixes: [rdf](#) [owl](#) [rdfs](#) [foaf](#) [+add prefix](#) [manage prefixes](#)

<input type="checkbox"/> (row index) URI	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> #CHROM cell
<input type="checkbox"/> add rdf:type	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> CHROM NAME cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> POS cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> ID cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> DBSNP cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> REF cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> ALT cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> QUAL cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> FILTER cell
	<input type="checkbox"/> X	>-property?->	<input type="checkbox"/> INFO cell

[Add another root node](#)

[Load](#) [Share](#) [Save](#)

# DEFINE SEMANTIC MODEL

Edit the base URI



### RDF Schema Alignment

The Semantic model below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the model. Configure the model by specifying which column to substitute into which node.

**Base URI:** <http://www.mydomain.org/variants/> [edit](#)

**Semantic model** | [RDF Preview](#)

Available Prefixes: [rdf](#) [owl](#) [rdfs](#) [foaf](#) [+ add prefix](#) [manage prefixes](#)

<input type="checkbox"/> (row index) URI	<input type="checkbox"/> >-property?->	<input type="checkbox"/> #CHROM cell
<input type="checkbox"/> add rdfs:type	<input type="checkbox"/> >-property?->	<input type="checkbox"/> CHROM NAME cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> POS cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> ID cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> DBSNP cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> REF cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> ALT cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> QUAL cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> FILTER cell
	<input type="checkbox"/> >-property?->	<input type="checkbox"/> INFO cell

[Add another root node](#) [Load](#) [Share](#) [Save](#)

2/2/2014



## DEFINE SEMANTIC MODEL

Edit the base URI

### RDF Schema Alignment

The Semantic model below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the model. Configure the model by specifying which column to substitute into which node.

Base URI: <http://www.mydomain.org/variants/> [edit](#)

Semantic model [RDF Preview](#)

Available Prefixes: [rdf](#) [owl](#) [rdfs](#) [foaf](#) [+ add prefix](#) [manage prefixes](#)

<a href="#">(row index) LIP</a>	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">#CHROM cell</a>
<a href="#">add rdf:type</a>	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">CHROM NAME cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">POS cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">ID cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">DBSNP cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">REF cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">ALT cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">QUAL cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">FILTER cell</a>
	<a href="#">X</a>	<a href="#">&gt;</a>	<a href="#">-property?-&gt;</a>	<a href="#">INFO cell</a>

[Add another root node](#)

[Load](#) [Share](#) [Save](#)

[OK](#) [Cancel](#)

Add SNPO prefix:  
<http://data.bioontology.org/ontology/SNPO/submission/4/download?apikey=8b5b7825-538d-40e0-9e9e-5ab9274a9aeb>

2/2/2014

## DEFINE SEMANTIC MODEL

**RDF Schema Alignment**

The Semantic model below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the model. Configure the model by specifying which column to substitute into which node.

**Base URI:** <http://www.mydomain.org/variants/> [edit](#)

**Semantic model** | **RDF Preview**

Available Prefixes: [rdf](#) [owl](#) [rdfs](#) [foaf](#) [SNPO](#) [+add prefix](#) [manage prefixes](#)

(row index) URI [add property](#)  
[add rdf:type](#)

Search for class:

Select an item from the list:

- [SNPO:reference\\_variant](http://www.loria.fr/~coulet/)
- [SNPO:variant](http://www.loria.fr/~coulet/)**
- [SNPO:generic\\_variant\\_db](http://www.loria.fr/~coulet/)
- [SNPO:locus\\_specific\\_variant\\_c](http://www.loria.fr/~coulet/)
- [SNPO:variant\\_database](http://www.loria.fr/~coulet/)

Your item not in the list?  
[Add it](#) (Shift+Enter)

[http://www.loria.fr/~coulet/ontology/ontology/version1.6/ontology\\_full.owl#variant](http://www.loria.fr/~coulet/ontology/ontology/version1.6/ontology_full.owl#variant)  
label:  
description: A variant is a genomic variation description at the nucleotide or protein level. It is defined by a position on a sequence. This sequence could be either a nucleotide sequence if the observed variation is a nucleotide variation, either a residue change if the observed variation is an residue variation. At the predicted position is observed a variation (the observed variation) in the sequence respectively to a "reference" (or "known") sequence.

[Add another root node](#) [Load](#) [Share](#) [Save](#)

[OK](#) [Cancel](#)

Add snpo:variant  
as rdf:type



## DEFINE SEMANTIC MODEL

Add *snpo:hasPosition* as the predicate relating variant and CHROMID

The screenshot shows a web-based interface for defining a semantic model. At the top, there are two tabs: "Semantic model" and "RDF Preview". Below the tabs, a header bar displays "Available Prefixes:" followed by a list of prefixes: "rdf owl rdfs foaf snpo". To the right of this list are two buttons: "+ add prefix" and "manage prefixes".

The main area is a table with columns for "(row index) URI", a list of existing properties, and a search input field. The existing properties listed are:

- >-property?->
- >-snpo:hasPosition->
- >-property?->
- >-property?->
- >-property?->
- >-property?->

The search input field contains the text "hasPosition". A dropdown menu is open below the search field, showing a list of items to select from:

- snpo:hasPosition (with the URL <http://www.loria.fr/~coulet/c>)

Below the list, there is a message "Your item not in the list?" and a button labeled "Add it (Shift+Enter)".

A red arrow points from the text on the left to the "snpo:hasPosition" entry in the table.

## DEFINE SEMANTIC MODEL

Add Dublin Core prefix

org/variants/ edit

iew

owl rdfs foaf

snpo:hasPos

property

**New Prefix**

Note: please make sure base uri in the rdf resources and the URI are the same.

prefix: dc

URI: http://purl.org/dc/elements/1.1/

(suggested by [prefix.cc](http://prefix.cc))

OK Cancel Advanced...

# DEFINE SEMANTIC MODEL

Semantic model RDF Preview

Available Prefixes: rdf owl rdfs foaf snpo dc [+add prefix](#) [manage prefixes](#)

(row index) URI   > snpo:hasPosition →  CHROMID cell

snpo:variant  
[add rdf.type](#)

> property? → Search for property: identifier

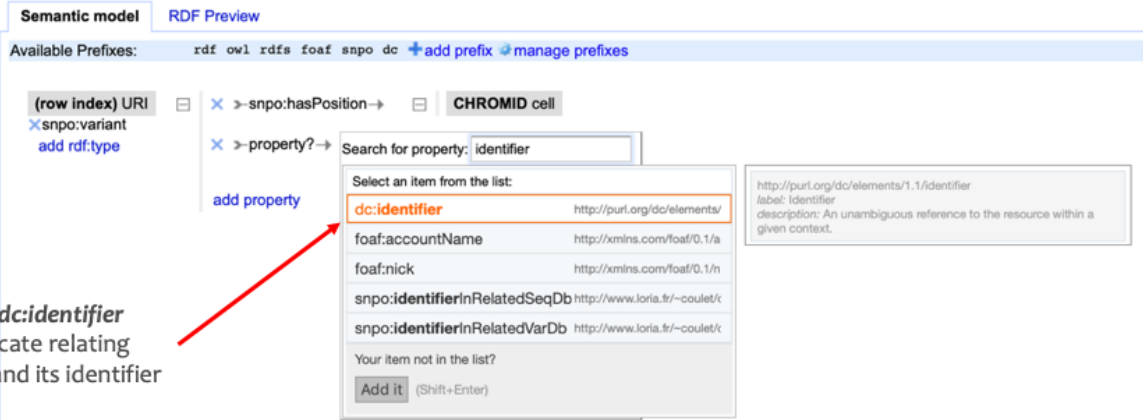
[add property](#)

Select an item from the list:

dc:identifier	http://purl.org/dc/elements/
foaf:accountName	http://xmlns.com/foaf/0.1/a
foaf:nick	http://xmlns.com/foaf/0.1/n
snpo:identifierInRelatedSeqDb	http://www.loria.fr/~coulet/c
snpo:identifierInRelatedVarDb	http://www.loria.fr/~coulet/c

Your item not in the list?  
 (Shift+Enter)

<http://purl.org/dc/elements/1.1/identifier>  
label: Identifier  
description: An unambiguous reference to the resource within a given context.



Add *dc:identifier* predicate relating variant and its identifier

## DEFINE SEMANTIC MODEL

### RDF Node

#### Use content from cell...

- (row index)
- #CHROM
- CHROMID
- POS
- ID
- DBSNP
- REF
- ALT
- QUAL
- FILTER
- INFO 1
- INFO 2
- INFO 3
- INFO 4
- INFO 5
- INFO 6
- INFO 7

Add *dc:identifier*  
predicate relating  
variant and its identifier

#### The cell's content is used ...

- as a URI
- as text
- as language-tagged text
- as integer number
- as non-integer number
- as date (YYYY-MM-DD)
- as dateTime (YYYY-MM-DD HH:MM:SS)
- as boolean
- as custom datatype (specify type URI)
- as a blank node

#### Use custom expression...

value  
[preview/edit](#)

OK Cancel

2/2/2014

## DEFINE SEMANTIC MODEL

Semantic model **RDF Preview**

This is a sample `Turtle` representation of (up-to) the *first 10 rows*

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix snpo: <http://data.bioontology.org/ontologies/SNPO/submissions/4/download?apikey=8b5b7825-538d-40e0-9e9e-5ab9274a9aeb420> .

<http://www.mydomain.org/variants/0> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos<

<http://www.mydomain.org/variants/1> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos<
  dc:identifier <http://identifiers.org/dbsnp/rs201003251> .

<http://www.mydomain.org/variants/2> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos<

<http://www.mydomain.org/variants/3> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos<
  dc:identifier <http://identifiers.org/dbsnp/rs34521632> .

<http://www.mydomain.org/variants/4> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos<

<http://www.mydomain.org/variants/5> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos<

<http://www.mydomain.org/variants/6> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
```

3/20/20

## DEFINE SEMANTIC MODEL

Semantic model **RDF Preview**

This is a sample Turtle representation of (up-to) the first 10 rows

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix snpo: <http://data.bioontology.org/ontologies/SNPO/submissions/4/download?apikey=8b5b7825-538d-40e0-9e9e-5ab9274a9aeb420> .

<http://www.mydomain.org/variants/0> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos

<http://www.mydomain.org/variants/1> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos
  dc:identifier <http://identifiers.org/dbsnp/rs201003251> .

<http://www.mydomain.org/variants/2> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos

<http://www.mydomain.org/variants/3> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos
  dc:identifier <http://identifiers.org/dbsnp/rs34521632> .

<http://www.mydomain.org/variants/4> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos

<http://www.mydomain.org/variants/5> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#hasPosition> <http://www.mydomain.org/variants/Chromos

<http://www.mydomain.org/variants/6> a <http://www.loria.fr/~coulet/ontology/snponology/version1.6/snponology_full.owl#variant> ;
```

We have an issue. The chromosome URI is not the one from NCIT but it is using our domain. We need to fix it.



## DEFINE SEMANTIC MODEL

### RDF Node

#### Use content from cell...

- (row index)
- #CHROM
- CHROMID
- POS
- ID
- DBSNP
- REF
- ALT
- QUAL
- FILTER
- INFO
- Constant Value

#### The cell's content is used ...

- as a URI
- as text
- as language-tagged text
- as integer number
- as non-integer number
- as date (YYYY-MM-DD)
- as dateTime (YYYY-MM-DD HH:MM:SS)
- as boolean
- as custom datatype (specify type URI)
- as a blank node

#### Use custom expression...

value  
[preview/edit](#)

OK

Cancel

2/2/2014

## DEFINE SEMANTIC MODEL

### Preview URI values

General Refine Expression Language (GREL) 

`cell.recon.match.id`

No syntax error.

row	value	cell.recon.match.id	resolved against the base URI
1.	Chromosome 1	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>
2.	Chromosome 1	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>
3.	Chromosome 1	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>
4.	Chromosome 1	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>
5.	Chromosome 1	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>
6.	Chromosome 1	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>
7.	Chromosome 1	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaur</a>

## DEFINE SEMANTIC MODEL

Semantic model | **RDF Preview**

This is a sample Turtle representation of (up-to) the first 10 rows

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix snpo: <http://data.bioontology.org/ontologies/SNPO/submissions/4/download?apikey=8b5b7825-538d-40e0-9e9e-5ab9274a9aeb&20> .

<http://www.mydomain.org/variants/0> a <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#hasPosition> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204> .

<http://www.mydomain.org/variants/1> a <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#hasPosition> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204> ;
  dc:identifier <http://identifiers.org/dbsnp/rs201003251> .

<http://www.mydomain.org/variants/2> a <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#hasPosition> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204> .

<http://www.mydomain.org/variants/3> a <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#hasPosition> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204> ;
  dc:identifier <http://identifiers.org/dbsnp/rs34521632> .

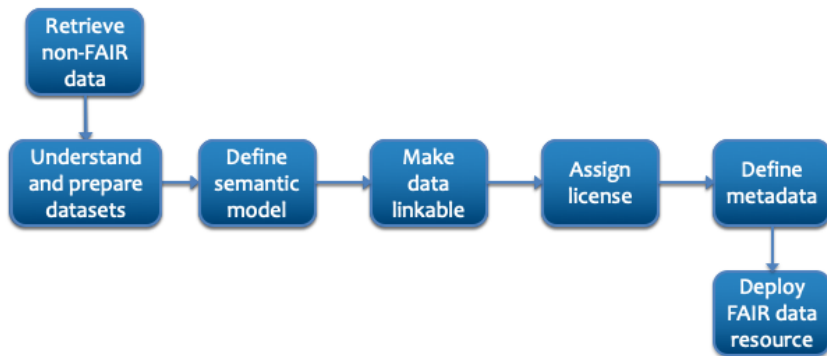
<http://www.mydomain.org/variants/4> a <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#hasPosition> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204> .

<http://www.mydomain.org/variants/5> a <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#hasPosition> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204> .

<http://www.mydomain.org/variants/6> a <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#variant> ;
  <http://www.loria.fr/~coulet/ontology/snponontology/version1.6/snponontology_full.owl#hasPosition> <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204> .
```

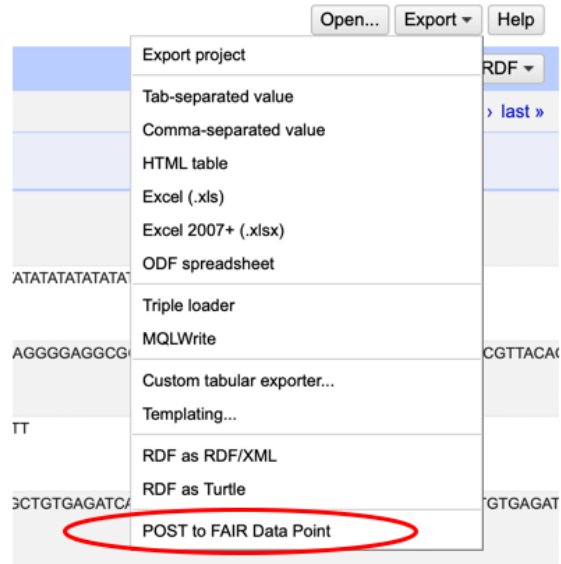
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C13204>

## FAIRIFICATION WORKFLOW



© 2024

## PUBLISH TO FAIR DATA POINT



## PUBLISH TO FAIR DATA POINT

**POST to FAIR Data Point**

The created RDF schema provided can now be uploaded to a FAIR Data Point.

**Base URI** [http:// edit](#)

## PUBLISH TO FAIR DATA POINT – ADD CATALOG

### POST to FAIR Data Point

The created RDF schema provided can now be uploaded to a FAIR Data Point.

**Base URI** <https://fairtraining.fair-dtls.surf-hosted.nl/fdp> [edit](#)

#### catalogs

[+ add catalog](#)



## PUBLISH TO FAIR DATA POINT – ADD CATALOG

4.	1	<b>POST to FAIR Data Point</b>	TTTTTTTTTTTT
5.	1	The created RDF schema provided can now be uploaded to a FAIR Data Point.	AGTGCATGGTGCTGTGAGATCAGCATG

### Add new catalog to FAIR Data Point [close]

[Show optional fields](#)

Title	<input type="text"/>
Has version	<input type="text"/>
Publisher	<input type="text" value="http://"/>
Publisher Name	<input type="text"/>
Theme taxonomy	<input type="text" value="http://"/>



## PUBLISH TO FAIR DATA POINT – ADD DATASET

### POST to FAIR Data Point

The created RDF schema provided can now be uploaded to a FAIR Data Point.

**Base URI** <https://fairtraining.fair-dtls.surf-hosted.nl/fdp> [edit](#)

#### catalogs

+ add catalog

[object Object] - FDS Course preparation

#### datasets

+ add dataset

## PUBLISH TO FAIR DATA POINT – ADD DATASET

### POST to FAIR Data Point

The created RDF schema provided can now be uploaded to a FAIR Data Point.

**Base URI** <https://fairtraining.fair-dtls.surf-hosted.nl/fdp> [edit](#)

#### catalogs

+ add catalog

[object Object] - FDS Course preparation

#### datasets

+ add dataset

## PUBLISH TO FAIR DATA POINT – ADD DATASET

4.	1	<b>POST to FAIR Data Point</b>	TTTTTTTTTTTT
5.	1	The created RDF schema provided can now be uploaded to a FAIR Data Point.	AGTGCATGGTGTGTGAGATCAGCAT

### Add new dataset to FAIR Data Point

[close]

[Show optional fields](#)

Title	<input type="text" value="GO NL 5 VCF - prep"/>
Has version	<input type="text" value="0.1"/>
Publisher	<input type="text" value="http://fairdata.tech/#/bonino"/>
Publisher Name	<input type="text" value="Luiz Bonino"/>
Theme	<input type="text" value="http://edamontology.org/format_3016"/>

3/2/2024

## PUBLISH TO FAIR DATA POINT – ADD DISTRIBUTION

### POST to FAIR Data Point

The created RDF schema provided can now be uploaded to a FAIR Data Point.

**Base URI** <https://fairtraining.fair-dtls.surf-hosted.nl/fdp> [edit](#)

#### catalogs

+ add catalog

[object Object] - FDS Course preparation

#### datasets

+ add dataset

undefined - GO NL 5 VCF - prep

push FAIRified data to triplestore

#### distribution

+ add distribution

OK Cancel



**Luiz Bonino**

International Technology Coordinator – GO FAIR  
Associate Professor BioSemantics – LUMC

E-mail: [luiz.bonino@go-fair.org](mailto:luiz.bonino@go-fair.org)

Skype: [luizolavobonino](https://www.skype.com/people/luizolavobonino)

Web: [www.go-fair.org](http://www.go-fair.org)