Transformation of the «Porphyry Cu Geochemistry Module» dataset to a database «sqlite»

1. Abstract

It's good practice to use the data in relational database, because it's the way to use SQL and python math libraries in any combinations. The source dataset Porphyry Cu Geochemistry Module.xls was transformed to an SQLite database – «geo3.db»

Source data is in the table «src». Each table has integer primary key "id" with autoincrement. Foreign key constraints are virtual (no SQLite engine supports)

2. Table description
   - 2.1. Table «holes» contains 14 rows with drill holes
   - 2.2. Table «samples» contains data of each sample with link to the hole by sapmles.hole_ref=holes.id
   - 2.3. Table «samples» contains data of each sample with link to the hole by sapmles.hole_ref=holes.id
   - 2.4. Table «elements» contains data rows of 48 elements
   - 2.5. Table «data» contains 183291 values of concentration to each chemical element in each sample.
     - Link to the «sample» «data.sample_ref»= «sample.id»
     - Link to the «element» «data. element _ref» = «elements.id»
     - Expo = -2 for «ppm» and -6 for «pct»
     - Mant – is the value of concentration, the «concentration» = «mant» * 10^ «expo»
     - Link to the data exclusion reason is «data. exclusion_ref»= «exclusion_ reasons.id». This field is used to exclude wrong data due the statistical ambiguity. If «data. exclusion_ref»=0 it means no exclusion normal data. If 1,2,3 etc… so look to «exclusion_reasons» table about the reasons list.

Additional segment of data model contains next tables:

2.6. Table «exclusion_reasons» contains list of reasons.

id	name
0	normal
1	apparent outlier
2	contradicts to base physics
3	results below the limit of detection

2.7. Table «geochem_methods» contains list of methods 4 acid Digestion

id	name
0	Current method
1	1F2 – Near Total Digestion – ICP
2	UT-4M – Near Total Digestion ICP-MS
3	Ultratrace 4 – Near Total Digestion – ICP and ICP-MS
4	Ultratrace 6 – Near Total Digestion – ICP and ICP-MS
5	UT-6M – Near Total Digestion – ICP and ICP-MS
6	8 - 4 Acid ICP Assay

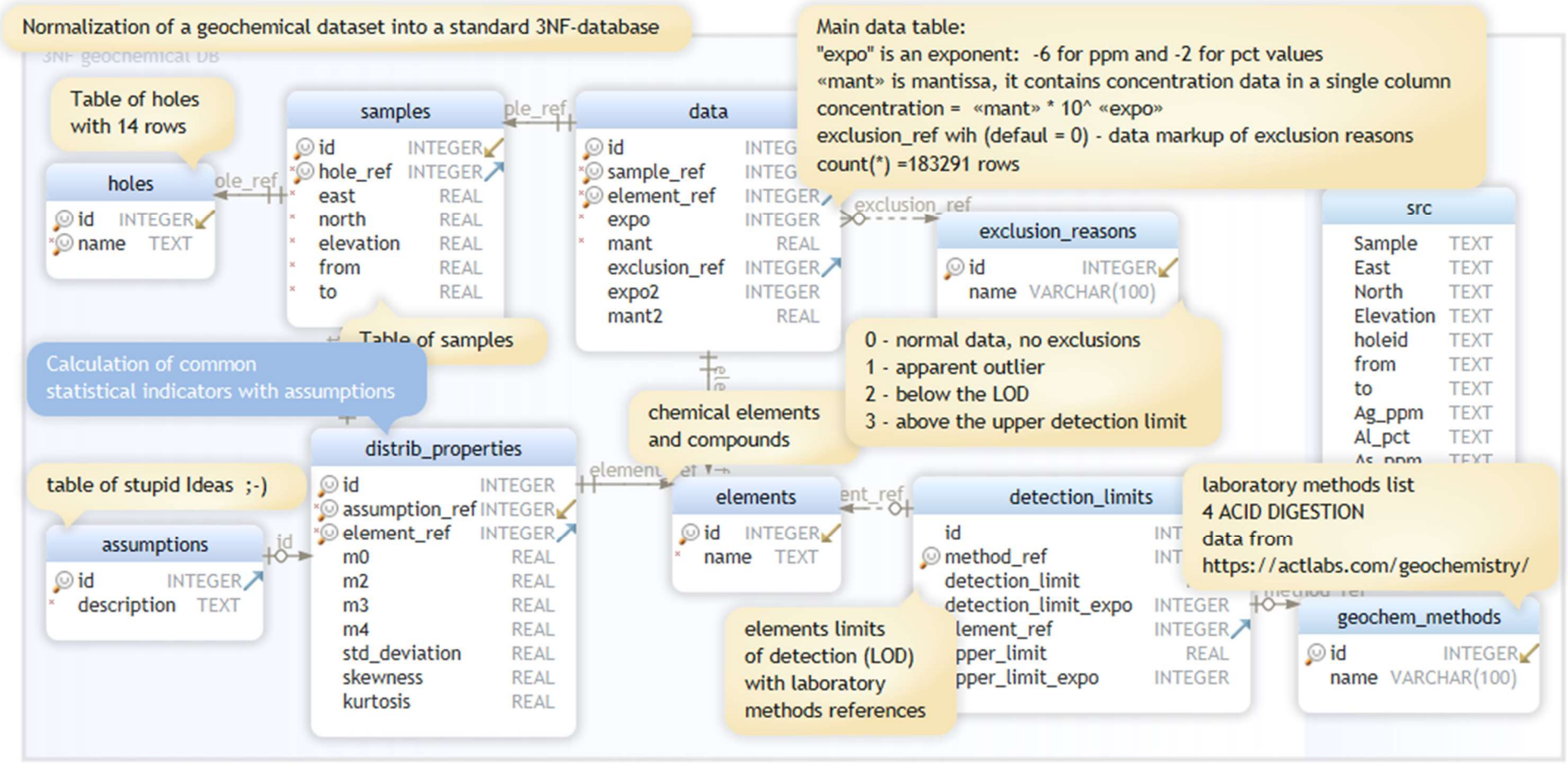
2.8. Table « detection_limits» contains limits of detection with lower and upper bounds for each method and each element contains 228 rows with laboratory instruments detection limits From https://actlabs.com/geochemistry/exploration-geochemistry/4-acid-near-total-digestion/

- Link to the «sample» « detection_limits. element_ref »= «elements.id»
- Link to the «element» « detection_limits. method_ref» = « geochem_methods.id»
- Detection limit is «detection_limit»*10^ «detection_limit_expo»

2.9. Table «assumptions» contains rows with calculation hypothesis

2.10. Table « distrib_properties » contains rows with calculations of statistical parameters for each hypothesis.

- M1 is the Mean.
- M2 is the Variance.
- M3 and M4 is the 3rd and 4th order moment
- Skew reflects symmetry of distribution: M3 / (M2**(3/2))
- Kurtosis reflects flatness of distribution: M4 / Sqr(M2)


Normalization of a geochemical dataset into a standard 3NF-database
Table of holes with 14 rows
holes
id INTEGER
name TEXT
samples
id INTEGER
hole_ref INTEGER
east REAL
north REAL
elevation REAL
from REAL
to REAL
Table of samples
data
id INTEG
sample_ref INTEG
element_ref INTEGER
expo INTEGER
mant REAL
exclusion_ref INTEGER
expo2 INTEGER
mant2 REAL
Main data table:
"expo" is an exponent: -6 for ppm and -2 for pct values
«mant» is mantissa, it contains concentration data in a single column
concentration = «mant» * 10^ «expo»
exclusion_ref wih (defaul = 0) - data markup of exclusion reasons
count(*) =183291 rows
exclusion_reasons
id INTEGER
name VARCHAR(100)
0 - normal data, no exclusions
1 - apparent outlier
2 - below the LOD
3 - above the upper detection limit
src
Sample TEXT
East TEXT
North TEXT
Elevation TEXT
holeid TEXT
from TEXT
to TEXT
Ag_ppm TEXT
Al_pct TEXT
Calculation of common statistical indicators with assumptions
chemical elements and compounds
table of stupid Ideas ;-)
assumptions
id INTEGER
description TEXT
distrib_properties
id INTEGER
assumption_ref INTEGER
element_ref INTEGER
m0 REAL
m2 REAL
m3 REAL
m4 REAL
std_deviation REAL
skewness REAL
kurtosis REAL
elements
id INTEGER
name TEXT
detection_limits
id INT
method_ref INT
detection_limit
detection_limit_expo INTEGER
lement_ref INTEGER
pper_limit REAL
pper_limit_expo INTEGER
elements limits of detection (LOD) with laboratory methods references
laboratory methods list
4 ACID DIGESTION
data from
https://actlabs.com/geochemistry/
geochem_methods
id INTEGER
name VARCHAR(100)