



The presentation

- The “(de)evolving” database life cycle:

I will talk about the NULL data problem, especially for older databases and try to make awareness and cover some of the issues regarding the NULL data problem. This will be showcased with an example: the (old) Danmap database.

Content

- Database
- Life cycle
- NULL data
- (de)evolving
- Example: Danmap

| | | | |
|-------------------------|------|-----|-------|
| Virgin Daisy | 2006 | 179 | 29.99 |
| Uncut Suicides | 2006 | 172 | 29.99 |
| Tracy Cider | 2006 | 142 | 29.99 |
| Song Hedwig | 2006 | 165 | 29.99 |
| Slacker Liaisons | 2006 | 179 | 29.99 |
| Sassy Packer | 2006 | 154 | 29.99 |
| River Outlaw | 2006 | 149 | 29.99 |
| Right Cranes | 2006 | 153 | 29.99 |
| Quest Mussolini | 2006 | 177 | 29.99 |
| Poseidon Forever | 2006 | 159 | 29.99 |
| Loathing Legally | 2006 | 140 | 29.99 |
| Lawless Vision | 2006 | 181 | 29.99 |
| Jingle Sagebrush | 2006 | 124 | 29.99 |
| Jericho Mulan | 2006 | 171 | 29.99 |
| Japanese Run | 2006 | 135 | 29.99 |
| Gilmore Boiled | 2006 | 163 | 29.99 |
| Floats Garden | 2006 | 145 | 29.99 |
| Fantasia Park | 2006 | 131 | 29.99 |
| Extraordinary Conquerer | 2006 | 122 | 29.99 |
| Everyone Craft | 2006 | 167 | 29.99 |



What is a Database?

| Name | Size | Last commit | Message |
|-------------------------|-----------|-------------|---|
| gltignore | 37 B | 2018-12-14 | Add install script to install database for KMA Inc. |
| CHECK-entries.sh | 2.33 KB | 2019-01-23 | CHECK-entries: make sure to escape regex chars |
| INSTALL.py | 3.79 KB | 2020-04-24 | Fixed version of KMA |
| README.md | 5.37 KB | 2021-04-20 | Added history file to content overview |
| aminoglycoside.fasta | 199.41 KB | 2021-03-11 | added gar1_fosM1_ermS0.qnrB89.cat.qnrB91.aac6.qnrB90.mcr126.mcr127 |
| antibiotic_classes.txt | 2.54 KB | 2021-12-01 | disinfectant separated, multiclassed fixed |
| beta-lactam.fasta | 1.79 MB | 2021-08-04 | Added blaVMB-2 |
| colistin.fasta | 91.6 KB | 2021-03-11 | added gar1_fosM1_ermS0.qnrB89.cat.qnrB91.aac6.qnrB90.mcr126.mcr127 |
| config | 873 B | 2021-12-01 | disinfectant separated, multiclassed fixed |
| disinfectant.fasta | 24.15 KB | 2021-02-19 | added disinfectant db |
| fosfomycin.fasta | 18.68 KB | 2021-03-11 | added gar1_fosM1_ermS0.qnrB89.cat.qnrB91.aac6.qnrB90.mcr126.mcr127 |
| fusidicacid.fasta | 1.96 KB | 2019-02-20 | Update fusidic acid db |
| glycopeptide.fasta | 94.34 KB | 2018-06-01 | Modify glycopeptide db Remove accessory genes not conferring glycopeptide resistance Provide the concatenated sequence of |
| history.txt | 39.46 KB | 2021-08-16 | Changed tabs in history.txt file |
| macrolide.fasta | 173.03 KB | 2021-03-11 | added gar1_fosM1_ermS0.qnrB89.cat.qnrB91.aac6.qnrB90.mcr126.mcr127 |
| nitroimidazole.fasta | 6.81 KB | 2018-05-24 | Reformat fosfomycin, fusidic acid and nitroimidazole db |
| notes.txt | 88.46 KB | 2021-09-04 | Added blaVMB-2 |
| oxazolidinone.fasta | 44.3 KB | 2018-06-25 | Correct gene names The names of a few genes repeated across databases (as they confer resistance to more than one antimicrobial class) were corrected to ensure |
| phenicol.fasta | 43.52 KB | 2021-03-11 | added gar1_fosM1_ermS0.qnrB89.cat.qnrB91.aac6.qnrB90.mcr126.mcr127 |
| phenotype_panels.txt | 2.55 KB | 2021-10-06 | changed rifampin for rifampicin phenotype_panels |
| phenotypes.txt | 509.09 KB | 2021-12-29 | Update classes in phenotypes.txt |
| pseudomonadicacid.fasta | 9.21 KB | 2021-03-09 | added aac(3)-IIa_6_CP023555_blaCMY1-150_2_NG_060513_blaCARB-4_L1J14749_mupA_LX75439_mupA_2_GU23736_mupB_LJQ231224_blaBEL1_LKF745070_blaBEL2_L |
| quinolone.fasta | 91.83 KB | 2021-06-12 | Fixed oqxA and oqxB gene names |
| rifampicin.fasta | 4.77 KB | 2020-08-28 | Updated with genes ARR-8_blaIMI-4_ant(3)-Ih_aac(6)-IId_aph(4)-Ib_aac(5)Ib-cr |
| sulphonamide.fasta | 45.76 KB | 2018-05-24 | Reformat phenicol, rifampicin and sulphonamide db |
| tetracycline.fasta | 226.64 KB | 2020-06-02 | Update db with blaVMB-1_blaLMB-1_blaCMY1-150_dfrA36/35_mcr-10_tetM6_blaPAU-1_blaKPC-41/36/40/37/38/39_blaKLC-3/4/2 |
| trimethoprim.fasta | 51.95 KB | 2021-04-13 | fixed dfrA19_2 bug |



Life cycle

- The life span of the software/data base
- Cycle since the database can be re-iterated/readjusted multiple times

Null data

- The data we use to represent nothing
- Problematic when "nothing" is not the same

Fun facts about "NULL":

- Null is typically unique for each instance of Null
- Does not specify an explicit value thus:
 - if $x = \text{null}$, will always be false
 - (many dbms's have made workaround to allow such a query)
- Meant to represent "unknown" or "not applicable"

Null Data: types

- Zero / nothing (😞)
- Not present
- Non-defined
- Not relevant

Example: Null data

| Vehicle name | Vehicle type |
|--------------|--------------|
| Boing 737 | Air plane |
| Fiat panda | Car |

Example: Null data 2

| Vehicle name | Vehicle type | wings |
|--------------|--------------|-------|
| Boing 737 | Air plane | yes |
| Fiat panda | Car | null |



Example: Null data 2 cont.

| Vehicle name | Vehicle type | wings |
|--------------|--------------|-------|
| Boing 737 | Air plane | yes |
| Fiat panda | Car | null |

Multiple correct responses:

- No
- What do you mean?
- That question is silly
- I don't know



Example: Null data 3

| Vehicle name | vehicle_type | wings | Amount_of_wings |
|--------------|--------------|-------|-----------------|
| Boing 737 | Airplane | Yes | 2 |
| Fiat panda | Car | null | null |
| Sco Classic | bike | no | null |

Example: Null data 4

| Vehicle name | vehicle_type | Amount_of_wings |
|--------------|--------------|-----------------|
| Boing 737 | Airplane | 2 |
| Fiat panda | Car | null |
| Sco Classic | bike | null |

Example: Null data 4 cont.

| Vehicle name | vehicle_type | Amount_of_wings |
|--------------|--------------|-----------------|
| Boing 737 | Airplane | 2 |
| Fiat panda | Car | null |
| Sco Classic | bike | null |

Again multiple interpretations:
How many wings do you vehicle have?

- 2
- 0
- None?
- Makes no sense?
- idk

Null data: potential consequences

- Increased redundancy
- Inconsistencies
- Decreased interpretability
- Increase number of assumptions

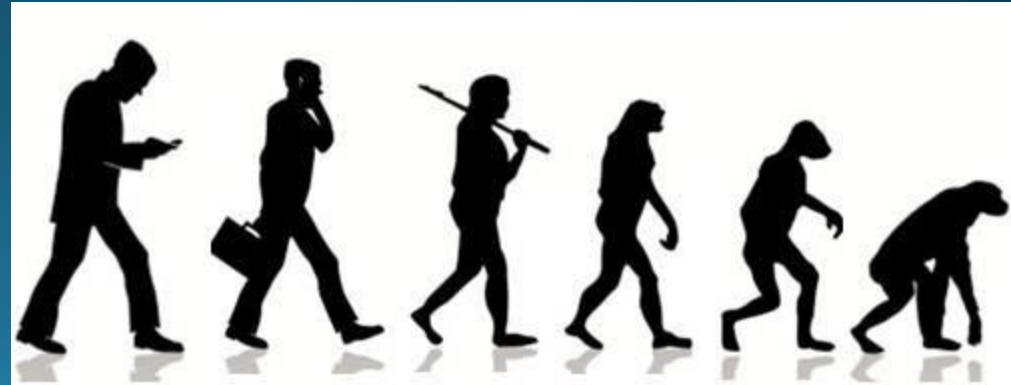
- Storing of non-informative information is for a long lived database quite high and may clutter and in some cases also decrease functionality in the long run
 - (waste sorting)

Null data – few remarks

- Is not bad nor good, is a functionality
- Can be miss used (unintentional)
- This is not a “DO NOT” presentation, more to create awareness

(de)evolving

- The database evolves and makes schematical changes
- The schematical change add/removes attributes for all rows
- Might add internal functional dependencies, which increases redundancy and lowers interpretability
- Often increases null values present in the database



Example: Danmap 1

| FUND | UNDERTYPE | DWH_ANALY | syntestCTX | syntestCAZ | phenoType | Fund_select |
|-----------------|-----------|-----------|------------|------------|-----------|-------------|
| Dublin | | 0 | - | - | - | 3 |
| Derby | | 0 | - | - | - | 2 |
| 4,[5],12:i- | | 0 | - | - | - | 2 |
| Typhimurium | | 0 | - | - | - | 2 |
| 4,[5],12:i- | | 0 | - | - | - | 2 |
| Typhimurium | | 0 | - | - | - | 2 |
| Agona | | 0 | - | - | - | 2 |
| Infantis | | 0 | - | - | - | 2 |
| 4,[5],12:i- | | 0 | - | - | - | 2 |
| O:4,5,12; H:b:- | | 0 | - | - | - | 2 |
| Bareilly | | 0 | - | - | - | 2 |
| Corvallis | | 0 | - | - | - | 2 |

Example: Danmap 2

- 602 of 75804 has a non-null value in the column in "syntestCAZ"

| UNDERTYPE | DWH_ANALY | syntestCTX | syntestCAZ | phenoType | Fund_select | Click to A |
|-----------|-----------|------------|------------|-----------|-------------|------------|
| + | | 21797 NEG | NEG | AmpC | | 1 |
| + | | 1325 NEG | NEG | AmpC | | 1 |
| + | | 1924 NEG | NEG | AmpC | | 1 |
| + | | 1387 NEG | NEG | AmpC | | 1 |
| + | | 21852 NEG | NEG | AmpC | | 1 |
| + | | 1767 NEG | NEG | AmpC | | 1 |
| + | | 1561 NEG | NEG | AmpC | | 1 |
| + | | 1576 NEG | NEG | AmpC | | 1 |
| + | | 21714 NEG | NEG | Other | | 1 |
| + | | 1707 NEG | NEG | AmpC | | 1 |
| + | | 1711 NEG | NEG | AmpC | | 1 |
| + | | 1735 NEG | NEG | AmpC | | 1 |
| + | | 21787 NEG | NEG | AmpC | | 1 |
| + | | 33 NEG | NEG | AmpC | | 1 |
| + | | 22625 NEG | NEG | AmpC | | 1 |
| + | | 263 NEG | NEG | AmpC | | 1 |
| + | | 22363 NEG | NEG | AmpC | | 1 |
| + | | 273 NEG | NEG | AmpC | | 1 |
| + | | 22442 NEG | NEG | AmpC | | 1 |
| + | | 22600 NEG | NEG | AmpC | | 1 |
| + | | 1919 NEG | NEG | AmpC | | 1 |
| + | | 155 NEG | NEG | AmpC | | 1 |
| + | | 100 NEG | NEG | AmpC | | 1 |
| + | | 23372 NEG | NEG | AmpC | | 1 |
| + | | 0 _ | - | - | | 4 |
| + | DT 41 | 0 _ | - | - | | 2 |
| + | DT 135 | 0 _ | - | - | | 2 |
| + | U292 | 0 _ | - | - | | 2 |
| + | U292 | 0 _ | - | - | | 2 |
| + | | 0 _ | - | - | | 4 |
| + | | 0 _ | - | - | | 4 |
| + | | 0 _ | - | - | | 4 |
| + | | 0 _ | - | - | | 4 |
| + | | 0 _ | - | - | | 4 |
| + | | 0 _ | - | - | | 4 |
| + | | 0 _ | - | - | | 4 |
| + | | 0 _ | - | - | | 4 |

Example: Danmap 3

- 10 out of 192 entries has a non-null value for "comment"
- 2 out of 192 entries has 2 comments in single field

| Breakpoint | Comment | ANTIBIOTIKA |
|------------|---|-------------|
| 4 | | 52 |
| 4 | | 52 |
| 4 | | 52 |
| 4 | | 52 |
| 8 | | 76 |
| 8 | | 76 |
| 8 | | 76 |
| 8 | | 76 |
| 8 | | 76 |
| 8 | | 76 |
| 8 | | 76 |
| 2 | | |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| | | 28 |
| 4 | | |
| 0,125 | | 33 |
| | Skal ikke tolkes i DANMAP - tolket med 1 til EFSA | |
| | Skal ikke tolkes i DANMAP - tolket med 1 til EFSA | |
| | Tolket med 1 til EFSA | |
| | Tolket med 1 til EFSA | |
| 0,5 | | |
| | | 38 |
| | | 38 |
| | | 38 |
| | | 38 |
| 4 | | |

Questions

