

# Data Quality



## Inconsistencies

Formats: Time Stamps, Date, Numericals, ...

Scales: (kg, g) (m, mm)

System: (metric, imperial)



## Wrong Entries:

Impossible Values -> set to NULL for analysis

Outliers -> exclude from analysis

NEVER overwrite original set



## Duplicates:

In DBS: Primary Keys

Consistent IDs should be used

# Data Quality @ ChemBioSys



## Inconsistencies

Formats: Machine-created Data

Scales: Machine-created Data e.g. NMR-Machine, Mass-Spectrometry, ...

System: metric



## Wrong Entries:

Impossible values: Machine Errors

Outlier Detection:

Checkup Routine done by individual Scientists



## Duplicates:

Data Management Platform (SEEK)

ID-consistency in each spreadsheet

Overlooked by each individual scientist