# Data validation

Automated:

Data field to data type validation (fx: CPRnummer)

Number of data points

Source vs hvad der er gemt på database.

Data Quality Metrics:

Conforming and nonconforming

Row passed, row failed (Is the row a real country or not?) (Is the email an email?)

Could only accept passing rows or destroy the entire data set.

Data validation techniques.

* Sampling, spot checks and audits.
* Cross validation
* Reasonable expectations (Data analysis judge data)
* Data profiling

# Cyber Security

## Confidentiality

Protects information and systems from unauthorized access.

Disclosure attacks seek to undermine confidentiality

## Availability

Ensure that information and systems are available for authorized users when needed.

Denial attacks seek to undermine availability.

## Integrity

Protects information and systems from unauthorized modification

Alteration attacks seek to undermine integrity

## Preservering data confidentiality

Access control is the primary way to limit this.

Access controls restrict users from accessing sensitivie information without permission.

Encryption is also an important security..

Encryption protects information from unauthorized access.

Encryption protects data at rest.

Encryption protects both data in motion and data in rest.

Data sanitization techniques:

* **Clearing** overwrites sensitive information to frustrate central analysis
* Purging uses more advanced techniques to frustrate laboratory analysis.
  + Might use cryptography to frustrate.
* **Destroying** completely obliterates the media through shredding, pulverization, melting and burning (Cant ever reuse the data)

