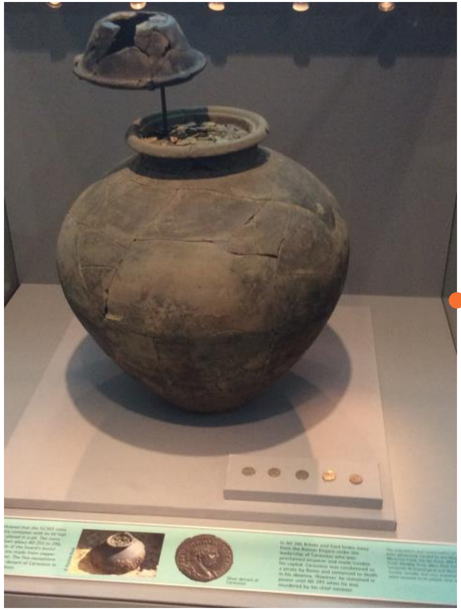
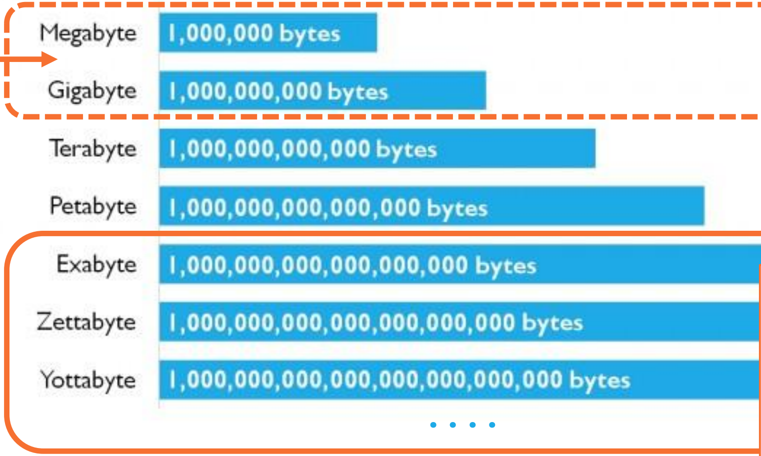
Applying Statistics and Computer Science to the study of big coin finds: an engineering approach

Luca Gianazza – https://www.sibrium.org/

Congress "Too big to Study?"
Trieste, May 17th – 19th 2018

How big is «BIG»?





«BIG» for Numismatists

«BIG» for Computer Scientists

Every day
we have …

150,000,000 VISA transactions

735,000,000 comments
posted on Facebook

4,000+ Petabytes
(= 4E18 Bytes) of Internet traffic

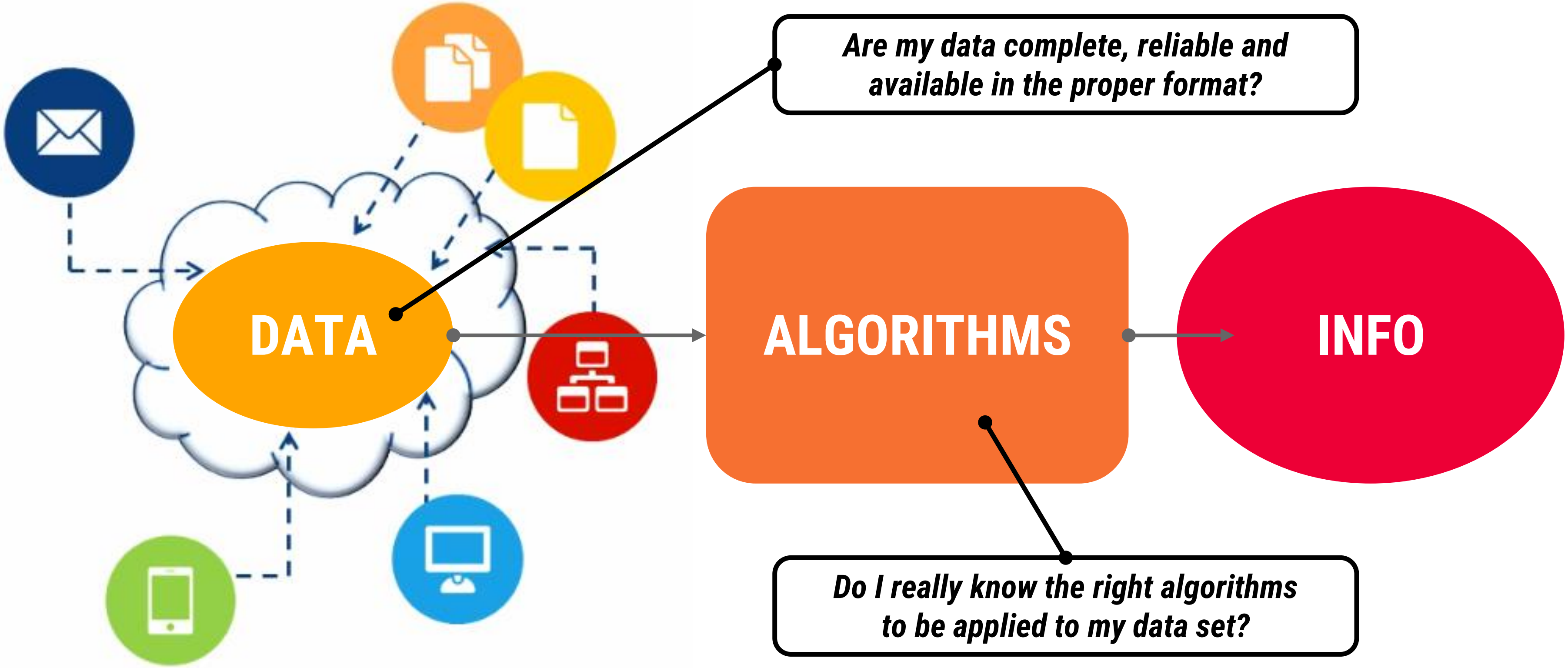
… whereas …

Portable Antiquity Scheme: 590,000 records

Coin Hoards of the Roman Empire project: 11,000 records

My personal archive of coin finds: 9,500 records

The challenges of Data Availability and Data Processing



The challenges of Data Availability and Data Processing

- **Synthetic descriptions of the coins**
- **Usually lack of archaeological details**
- **Attributions not always made by numismatists**
- **Obsolete/misleading bibliography**

Je puis vous signaler quelques trouvailles du dernier temps, qui cependant n'ont pas reçu une certaine publicité.

La première date du commencement d'Août de cette année et eut lieu dans les environs de Ornovaseo, petit bourg aux frontières du Piémont.

Quelques paysans en faisant sauter en l'air des vieilles ruines y trouvèrent un coffret, qui, immédiatement mis en pièces montra à leurs yeux ébahis une certaine quantité de monnaies d'un aspect douteux. Trompées dans leur espérance d'y voir de l'or, ces bonnes gens se prirent d'abord à les réduire dans un état tout à fait méconnaissable, puis découvrant qu'elles étaient cependant

E. TAHEN, *Lettre critique à Mons. F. Schweitzer touchant la première décade*, in *Mittheilungen aus dem Gebiete der Numismatik und Archaeologie. Notizie peregrine di Numismatica e d'Archeologia*, II, Trieste 1854, pp. 81-96 (82-84)

a)

b)

Coins in the name of Berengar king of Italy, usually both attributed to Berengar I (888-924), mint of Milan.
Attribution recently changed to:
a) Berengar I (888-924), mint of Verona (or Venice?)
b) Berengar II (950-961), mint of Venice

Studying a coin find: a budgetary perspective

general assumptions:
- 8 working hours per day
- 250 working days per year
- 300 €/MWD

MWD: man working days

Completeness and accuracy have a cost

The number of entries is affected by the time dedicated to each entry, whatever they are high level descriptions of hoards or single coins

case 1: 15 min. per entry

- 4 entries per hour
- 32 entries per day
- 8,000 entries per year

1,000 entries = 31.25 MWD

↓↓↓

1,000 entries = 9,375 €

case 2: 30 min. per entry

- 2 entries per hour
- 16 entries per day
- 4,000 entries per year

1,000 entries = 62.5 MWD

↓↓↓

1,000 entries = 18,750 €

case 3: 60 min. per entry

- 1 entries per hour
- 8 entries per day
- 2,000 entries per year

1,000 entries = 125 MWD

↓↓↓

1,000 entries = 37,500 €

What if I want to publish the Reka Devnia (BG) hoard?

81,096 republican and imperial roman coins found in 1929 on the site of ancient Marcianopolis (Moesia Inferior), preliminarily publicated by N. Mouchmov in 1934, after a further 20,000+ coins had been dispersed. **The total number of coins therefore exceeds 101,096 specimen**

(*source*: http://chre.ashmus.ox.ac.uk/hoard/3406)

846 €

Time to count the coins ≈ 2.82 MWD

1 count per second
lack of errors

506,850 €

Time to record the coins into a database (full description and pictures)

≈ 1,689.5 MWD (≈ 6.76 MWY)

10 minutes per coin

???

Number of pages needed to publish the coins (full description and pictures)

≈ 2,704 pages and 4,055 plates

description: 30 coins per page
pictures: 20 coins per plate

general assumptions:
- 8 working hours per day
- 250 working days per year
- 300 €/MWD

MWD: man working days
MWY: man working years

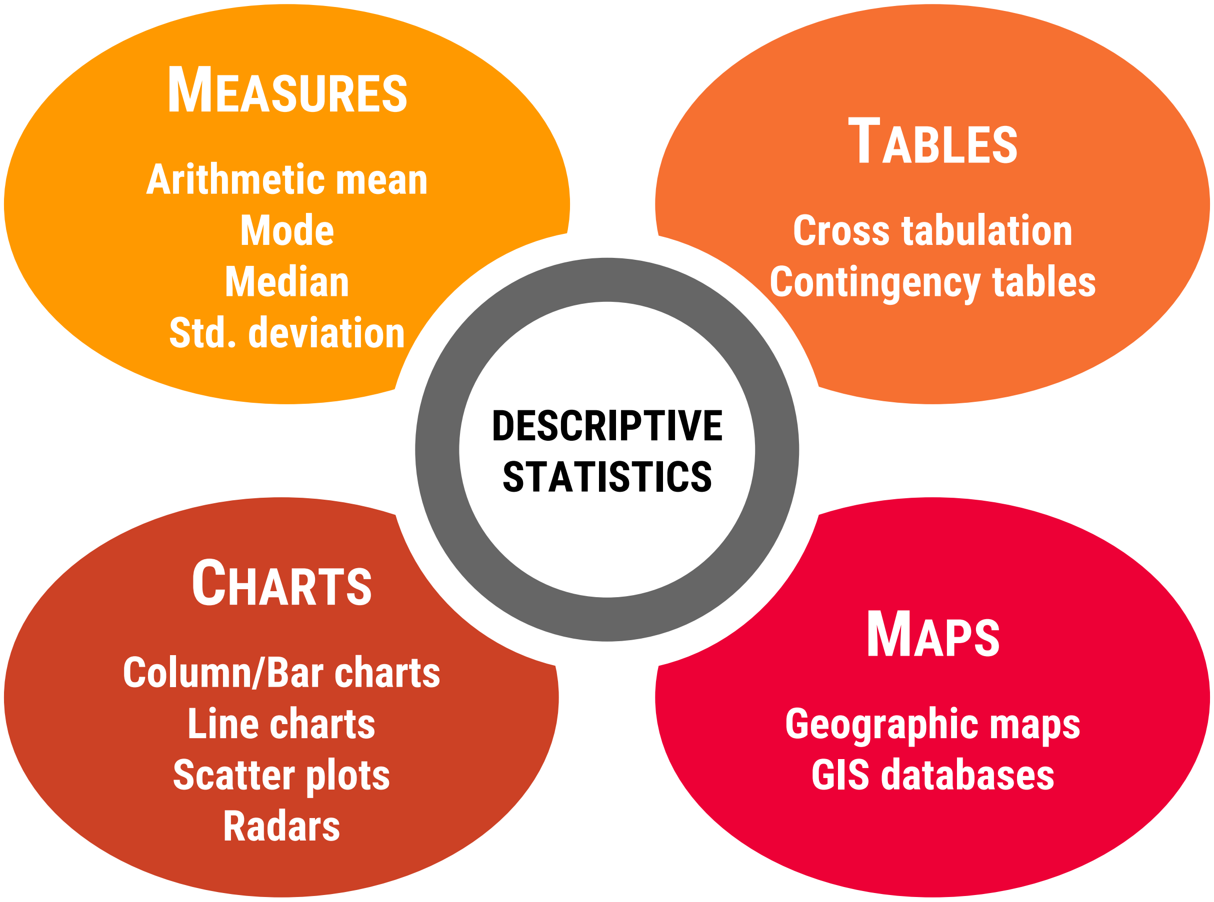
Let’s now talk about Statistics…

How many STATISTICS?

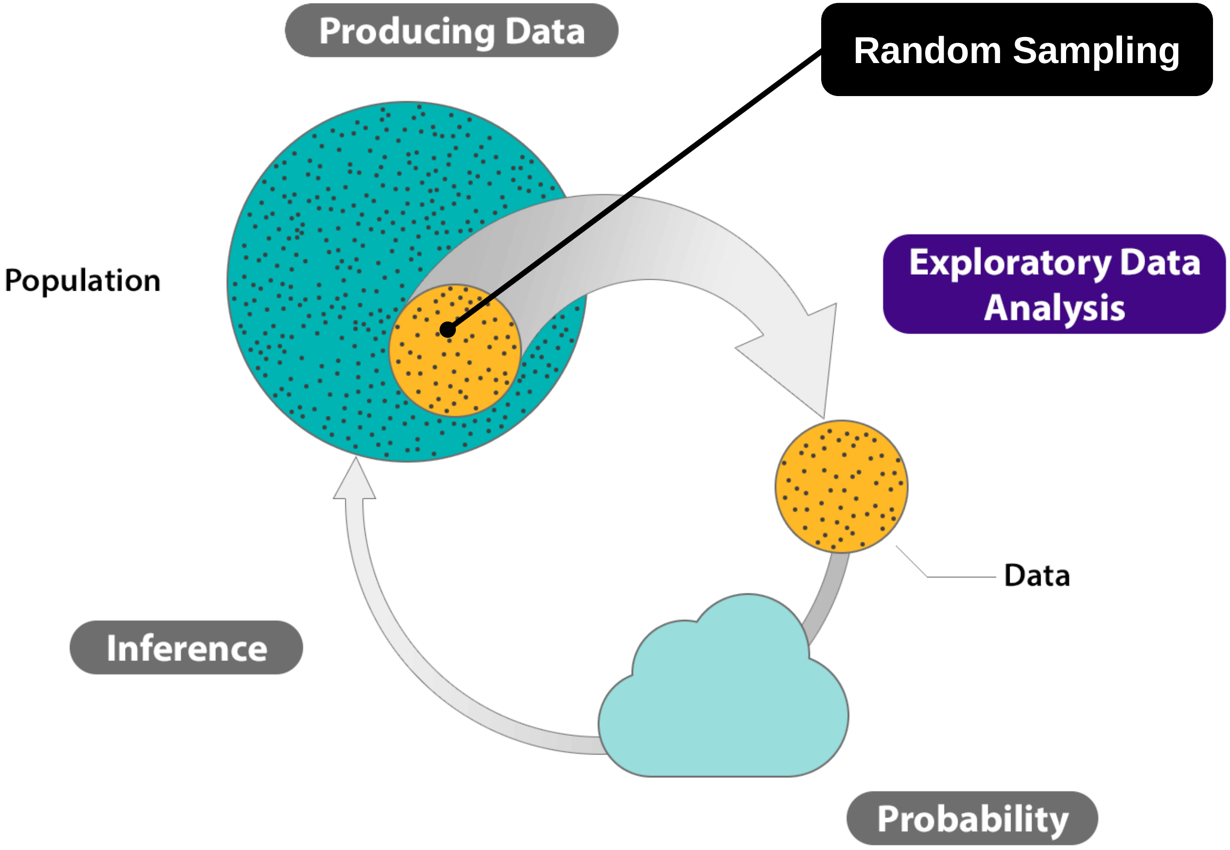
DESCRIPTIVE STATISTICS is the process of using and analyzing a summary statistic that quantitatively describes or summarizes features of a collection of data

STATISTICAL INFERENCE is the process of using data analysis to deduce properties of an underlying probability distribution, inferring properties of a population from a data set sampled from a larger population

Descriptive Statistics



Statistical Inference



“

*Unfortunately, hoard data are not the ideal “experimental” data treated in statistics texts. Numismatic analyses are often complicated by **small sample sizes** and **non-randomness**, which may invalidate statistical conclusions.*

(Warren W. ESTY, *Statistical analysis of hoard data in ancient numismatics*)

The limits of Statistics applied to the analysis of coin hoards

- ***small sample sizes***

  The larger the observed sample is, the better the quality of an estimate is

- ***non-randomness***

  A hoard is not necessarily the result of a random sampling of the coins circulating in a region, but rather of a selection of them

From a statistical perspective, a hoard can be hardly considered a reliable estimator of a larger population (e.g. mint production, coin circulation)

Modelling a hoard as an array of independent random variables

We can model a hoard of ***N*** coins as an **array of *independent random variables***

$$\hat{\mathcal{H}} = (X_1, X_2, \dots, X_N)$$

Each random variable X_i **is a single coin/object** having specific **qualitative** (e.g., mint, types, inscriptions) **and quantitative** (e.g., module, weight, axis orientation) **properties**.

Not all the coins have comparable properties
(e.g., a golden *florin* was minted with a theoretical weight different from that of a bullion *penny*)

$$\Downarrow$$

Not all the random variables are *identically distributed*

The problem of sampling

Working with ***M<N* coins *randomly extracted from the hoard*** implies that the subset $\widehat{\mathcal{H}}'$ of coins

$$\widehat{\mathcal{H}}' = (X_1', X_2', \dots, X_M')$$

where $\widehat{\mathcal{H}}' \subseteq \widehat{\mathcal{H}}$

- **is considering coins all with comparable properties** (i.e., *identically distributed random variables*);
- **is informative enough** (i.e., *M* is «big» enough to give validity to the theorems of Statistics and accuracy to the used estimators).

The previous assumptions drastically reduce the application scenarios of Statistical Inference to the study of «big» coin hoards

The case of the Reka Devnia (BG) hoard

- Julia Maesa (Augusta), Denarius, Rome (218/22 CE), RIC 268: **547 pieces**
- Faustina I (Diva), Denarius, Rome (141 CE), RIC 351a: **498 pieces**
- Julia Mamaea (Augusta), Denarius, Rome (225/35 CE), RIC 343: **467 pieces**
- Maximinus I Thrax (Augustus), Denarius, Rome (235/6 CE), RIC 14: **418 pieces**
- Faustina I (Diva), Denarius, Rome (141 CE), RIC 344a: **390 pieces**
- Marcus Aurelius (Caesar), Denarius, Rome (145/60 CE), RIC 429a: **338 pieces**
- Julia Mamaea (Augusta), Denarius, Rome (225/35 CE), RIC 360: **319 pieces**
- Julia Maesa (Augusta), Denarius, Rome (218/22 CE), RIC 271 or 272: **316 pieces**
- Faustina II (Augusta), Denarius, Rome (161/75 CE), RIC 677: **311 pieces**
- Faustina I (Diva), Denarius, Rome (141 CE), RIC 362: **309 pieces**
- Julia Domna (Augusta), Denarius, Rome (196/211 CE), RIC 574: **300 pieces**

- **28 more types with 200 to 299 pieces each**
- **146 more types with 100 to 199 pieces each**
- **1,322 more types with 10 to 99 pieces each**
- **1,156 more types with 2 to 9 pieces each**
- **689 more types with 1 piece each only**

(*source*: http://chre.ashmus.ox.ac.uk/hoard/3406)

The multinomial distribution

Evaluation of **qualitative information** (e.g., distribution analysis of types, mints, metal)

$$f(x_1, \dots, x_m; n, p_1, \dots, p_m) = \mathrm{P}(X_1 = x_1, \dots, X_m = x_m) =$$

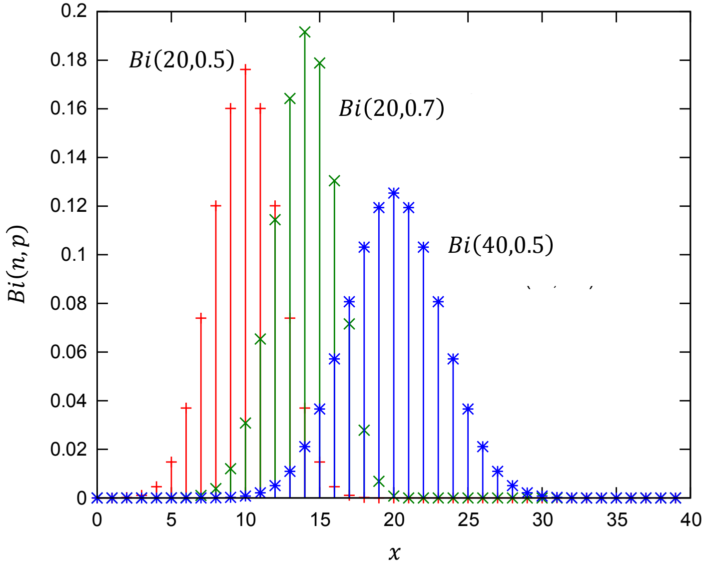
$$= \begin{cases} \dfrac{n!}{x_1! \dots x_m!} \cdot p_1^{x_1} \cdot \dots \cdot p_m^{x_m}, & \text{when } \sum_{i=1}^{m} x_1 = n \\ 0, & \text{otherwise} \end{cases}$$

- **Multinomial distribution**: $\boldsymbol{n}$ mutually independent trials, each with $\boldsymbol{m}$ possible outcomes
- **Binominal distribution**: $\boldsymbol{n}$ mutually independent trials, each with $\mathbf{2}$ possible outcomes
- **Bernoulli distribution**: $\mathbf{1}$ trial, with $\mathbf{2}$ possible outcomes

The binomial distribution $\mathcal{B}in(n, p)$

Evaluation of **qualitative information** (e.g., distribution analysis of types, mints, metal)

$$Bin(n,p):\ f(k,n,p) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 0,1,2,\dots n$$



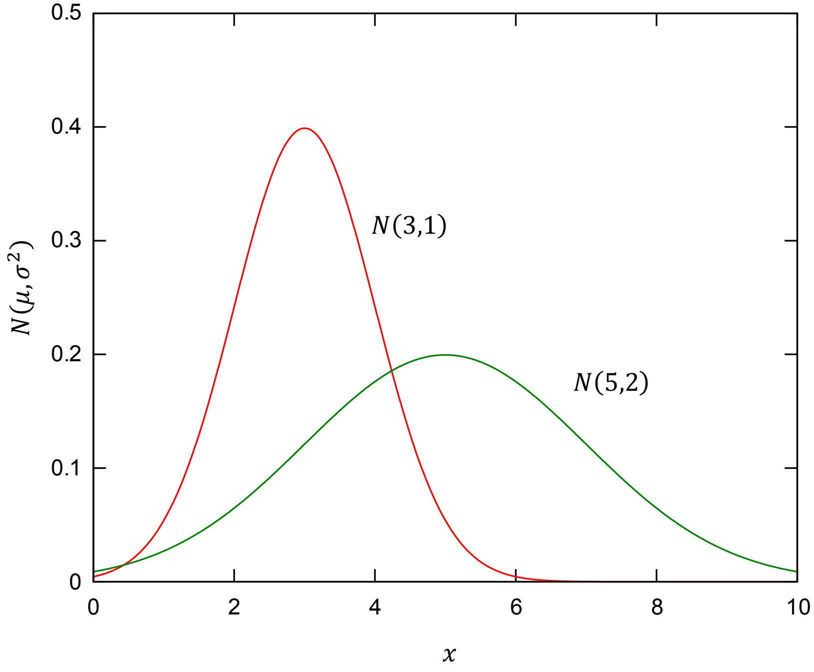
For $n = 1$ the binomial distribution is a **Bernoulli distribution**

$$Bernoulli(p):\ f(k,p) = p^k (1-p)^{1-k} \text{ for } k \in \{0,1\}$$

The normal (gaussian) distribution $N(\mu, \sigma^2)$

Evaluation of **quantitative information** (e.g., weight, diameter)

$$N(\mu, \sigma^2): f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$



The need to have a large sample size

Weak law of large numbers

Suppose $X_1, X_2, \ldots$ is a sequence of ***independent and identically distributed random variables*** with finite expected value $\mathbb{E}(X_i) = \mu$.
The ***sample average***

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + \cdots + X_n)$$

converges to μ as $n \to \infty$.

The need to have a large sample size

Weak law of large numbers (case with $\sigma^2 < \infty$)

Suppose X_1, X_2, ... is a sequence of ***independent and identically distributed random variables*** with finite expected value $\mathbb{E}(X_i) = \mu$ and finite variance $\mathrm{var}(X_i) = \sigma^2$. From the **Chebyshev's inequality** it follows that

$$P(|\bar{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

where $n > 0$ and $\varepsilon > 0$.

The need to have a large sample size

Example

Suppose a distribution with finite average value μ unknown and variance $\sigma^2 = 1$.
How big should be the sample to have a probability at least of 95% to have the sample average $\bar{X}_n$ distant less than 0.5 from the average value μ?

With $\sigma^2 = 1, \varepsilon = 0.5$ and $P(|\bar{X}_n - \mu| < \varepsilon) \geq 0.95$, it follows that

$$1 - \frac{\sigma^2}{n\varepsilon^2} \geq 0.95 \Rightarrow \boldsymbol{n \geq 80}$$

The need to have a large sample size

Central Limit Theorem

Suppose X_1, X_2, ... is a sequence of ***independent and identically distributed random variables*** with finite expected value $\mathbb{E}(X_i) = \mu$ and finite variance $\mathrm{var}(X_i) = \sigma^2$. The random variable

$$Z_n = \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sqrt{\mathrm{var}(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

converges in distribution to a normal distribution $N(0,1)$ as $n \to \infty$.

Estimation

A set of random samples $X_1, X_2, \dots, X_n$ extracted from a population with probability distribution function $f(\cdot;\theta)$ known except for the parameter $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ is used to determine an estimator $\widehat{\theta}$ for θ.

POINT ESTIMATION: calculation of a ***single value*** which is to serve as a «best estimate» of an unknown parameter $\boldsymbol{\theta_i}$ $(1 \le i \le k)$

INTERVAL ESTIMATION: calculation of an ***interval of plausible values*** of an unknown parameter $\boldsymbol{\theta_i}$ $(1 \le i \le k)$

Point estimators for a Bernoulli distribution $Bernoulli(p)$: an example

$Bernoulli(p)$: $f(k,p) = p^k(1-p)^{1-k}$ for $k \in \{0,1\}$

In this case $\theta = (\theta_1) = p$.

The estimator $\hat{\theta} = (\hat{\theta}_1) = \hat{p}$ is a function of $X_1, X_2, \ldots, X_n$.

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

Point estimators for a normal distribution $N(\mu, \sigma^2)$: an example

$$N(\mu, \sigma^2):\ f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$

In this case $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$.

The estimator $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2\right) = (\hat{\mu}, \hat{\sigma}^2)$ is a function of $X_1, X_2, \ldots, X_n$.

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

Interval estimators for a binomial distribution $Bin(n, p)$: an example

Under the assumptions that

- **$\widehat{\boldsymbol{p}}$ and $\mathbf{1} - \widehat{\boldsymbol{p}}$ are not close either to 0, or to 1**
- **$\boldsymbol{n}(\mathbf{1} - \widehat{\boldsymbol{p}}) > \mathbf{5}$ and $\boldsymbol{n}\widehat{\boldsymbol{p}} > \mathbf{5}$**

the interval of confidence for p is

$$p \sim \left[\bar{X}_n - z_{1-\frac{r}{2}} \cdot \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} ; \bar{X}_n + z_{1-\frac{r}{2}} \cdot \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \right]$$

where:

n is the number of samples

$\bar{X}_n$ is the sample average over n samples

z_b is the b-th quantile of the normal distribution $N(0,1)$

Interval estimators for a normal distribution $N(\mu, \sigma^2)$: an example

Since the variance of the sample average is not corresponding to σ^2 ($\rightarrow$ see *Central limit theorem*), we define a ***sample variance*** as

$$\boldsymbol{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

for which

$$\mathbb{E}(\boldsymbol{S}_n^2) = \sigma^2$$

Interval estimators for a normal distribution $N(\mu, \sigma^2)$: an example

To estimate the average value μ, it worth noting that the random variable T

$$T = \frac{\bar{X}_n - \mu}{\boldsymbol{S}_n/\sqrt{n}} \sim t_{n-1}$$

i.e., it follows the **Student's *t*-distribution** with $n - 1$ degrees of freedom.

Interval estimators for a normal distribution $N(\mu, \sigma^2)$: an example

To estimate the average value σ^2, it worth noting that the random variable V

$$V = \frac{(n-1)\boldsymbol{S}_n^2}{\sigma} \sim \chi_{n-1}^2$$

i.e., it follows the **chi-squared χ^2 distribution** with $n-1$ degrees of freedom.

Interval estimators for a normal distribution $N(\mu, \sigma^2)$: an example

The intervals of confidence for μ and σ^2 are

$$\mu \sim \left[\bar{X}_n - t_{n-1,1-\frac{r}{2}} \cdot \frac{\boldsymbol{S}_n}{\sqrt{n}} ; \bar{X}_n + t_{n-1,1-\frac{r}{2}} \cdot \frac{\boldsymbol{S}_n}{\sqrt{n}}\right]$$

$$\sigma^2 \sim \left[(n-1)\frac{\boldsymbol{S}_n^2}{v_{n-1,1-\frac{r}{2}}} ; (n-1)\frac{\boldsymbol{S}_n^2}{v_{n-1,\frac{r}{2}}}\right]$$

where:

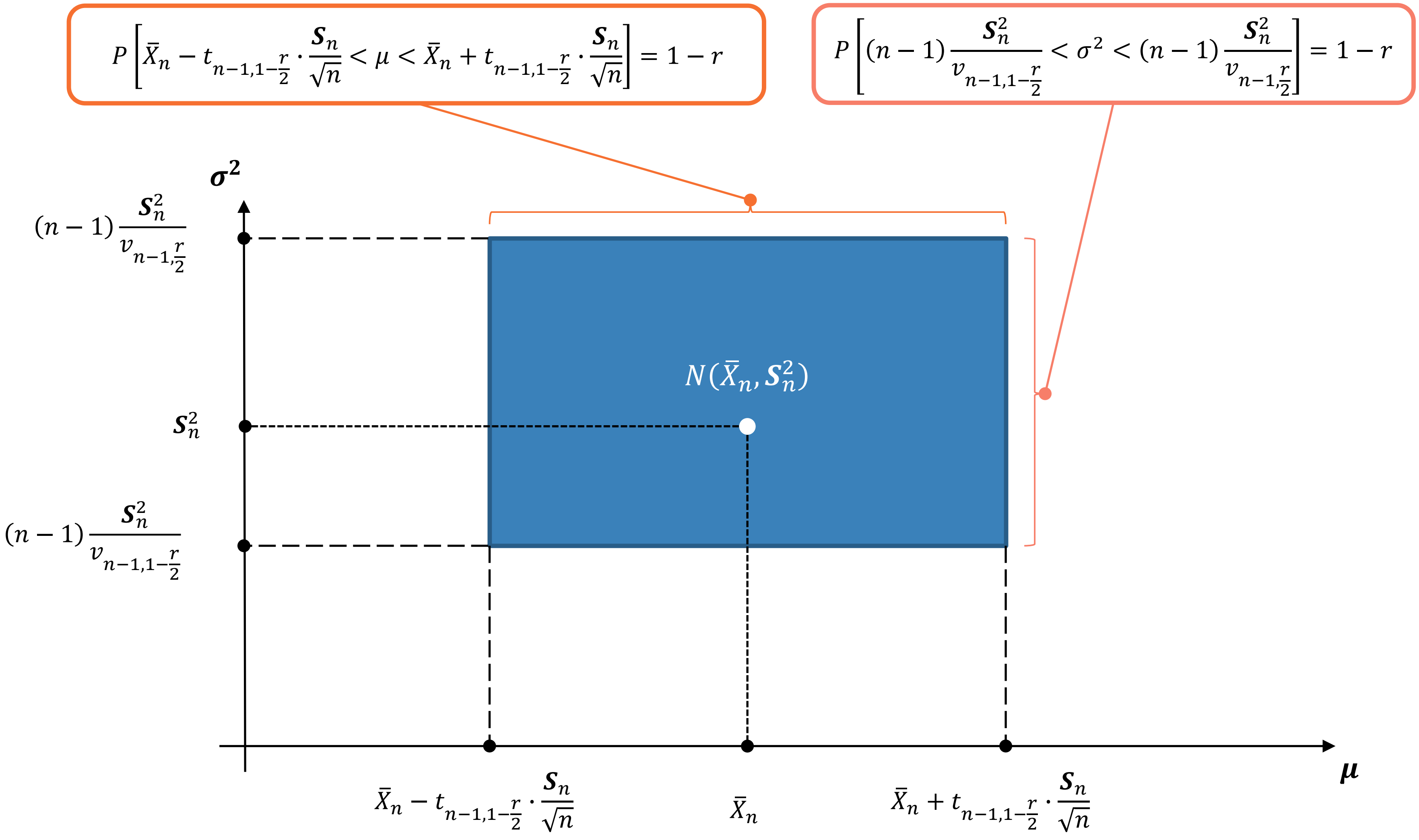
n is the number of samples

$\bar{X}_n$ is the sample average over n samples

$\boldsymbol{S}_n^2$ is the sample variance over n samples

$t_{a,b}$ is the b-th quantile of the *t* distribution with a degrees of freedom

$v_{a,b}$ is the b-th quantile of the χ^2 distribution with a degrees of freedom



$P\left[\bar{X}_n - t_{n-1,1-\frac{r}{2}} \cdot \frac{S_n}{\sqrt{n}} < \mu < \bar{X}_n + t_{n-1,1-\frac{r}{2}} \cdot \frac{S_n}{\sqrt{n}}\right] = 1 - r$
$P\left[(n-1)\frac{S_n^2}{v_{n-1,1-\frac{r}{2}}} < \sigma^2 < (n-1)\frac{S_n^2}{v_{n-1,\frac{r}{2}}}\right] = 1 - r$
σ^2
$(n-1)\frac{S_n^2}{v_{n-1,\frac{r}{2}}}$
S_n^2
$(n-1)\frac{S_n^2}{v_{n-1,1-\frac{r}{2}}}$
$N(\bar{X}_n, S_n^2)$
μ
$\bar{X}_n - t_{n-1,1-\frac{r}{2}} \cdot \frac{S_n}{\sqrt{n}}$
$\bar{X}_n$
$\bar{X}_n + t_{n-1,1-\frac{r}{2}} \cdot \frac{S_n}{\sqrt{n}}$

Use case: estimation of a **QUALITATIVE** property

Let's have a subset of **M coins** $\widehat{\mathcal{H}}' = (X_1', X_2', \dots, X_M')$ ***randomly extracted from the hoard*** of **N coins** $\widehat{\mathcal{H}} = (X_1, X_2, \dots, X_N)$, with M<N.

To estimate the frequency occurrence e.g. of specific mints, instead of using a multinomial distribution we might apply the binomial distribution $\mathcal{B}in(N, p_j)$ for each possible mint j (or – better – ***for the most represented mints***)

- p_j= percentage of coins from the mint j in the whole hoard $\widehat{\mathcal{H}}$ (i.e, probability of extracting a coin of the mint j from the N coins of the hoard $\widehat{\mathcal{H}}$);
- $\hat{p}_{j,M}$= percentage of coins from the mint j in the data subset $\widehat{\mathcal{H}}'$.

$p_{j,N} \cong \hat{p}_{j,M} = \frac{1}{N}\sum_{i=1}^{N} X_i'$ (± delta from the interval of confidence)

Use case: estimation of a **QUALITATIVE** property

example: estimation of mint distribution

Need to group the poorly represented mints to satisfy the assumptions for interval estimation

Example *(M = 100 coins, out of N = 100,000 coins)*

41 coins from **mint A** $\rightarrow \hat{p}_{A,100} = 0.41$
23 coins from **mint B** $\rightarrow \hat{p}_{B,100} = 0.23$
14 coins from **mint C** $\rightarrow \hat{p}_{C,100} = 0.14$
12 coins from **mint D** $\rightarrow \hat{p}_{D,100} = 0.12$
10 coins from **other mints** $\rightarrow \hat{p}_{oth,100} = 0.10$

This means that in the hoard $\widehat{\mathcal{H}}$ we should expect

- ***41,000 ± 9,640*** coins from **mint A** with a probability of 95%
- ***23,000 ± 8,248*** coins from **mint B** with a probability of 95%
- ***14,000 ± 6,801*** coins from **mint C** with a probability of 95%
- ***12,000 ± 6,369*** coins from **mint D** with a probability of 95%
- ***10,000 ± 5,880*** coins from **other mints** with a probability of 95%

Use case: estimation of a **QUALITATIVE** property

In other words:

- ***41,000 ± 24%*** coins from **mint A** with a probability of 95%
- ***23,000 ± 36%*** coins from **mint B** with a probability of 95%
- ***14,000 ± 49%*** coins from **mint C** with a probability of 95%
- ***12,000 ± 53%*** coins from **mint D** with a probability of 95%
- ***10,000 ± 59%*** coins from **other mints** with a probability of 95%

The probability of having exactly 41,000 + 23,000 + 14,000 + 12,000 + 10,000 coins is less than 10E-11 (1 out of 100,000,000,000; approximation via normal distribution).

The probability of having a distribution of the mints in the whole $\widehat{\mathcal{H}}$ hoard matching the intervals of confidence for the five mints is approximately around 77 %.

Use case: estimation of a **QUANTITATIVE** property

example: DUCATONI of Vincenzo I Gonzaga (1587-1612), mint of Casale Monferrato

Number of samples: 10

sample average: **31.316000 grams**
sample variance: **0.192893 grams^2**

t-Student quantile (95%): **2.262157**
chi-squared quantiles (95%): **19.022768 / 2.700389**

estimated weight (95%): 31.00 / 31.63
estimated variance (95%): 0.09 / 0.64

Number of samples: 20

sample average: **30.939500 grams**
sample variance: **1.778489 grams^2**

t-Student quantile (95%): **2.093024**
chi-squared quantiles (95%): **32.852327 / 8.906516**

estimated weight (95%): 30.32 / 31.56
estimated variance (95%): 1.03 / 3.79

Number of samples: 50

sample average: **30.855800 grams**
sample variance: **2.389776 grams^2**

t-Student quantile (95%): **2.009575**
chi-squared quantiles (95%): **70.222414 / 31.554916**

estimated weight (95%): 30.42 / 31.30
estimated variance (95%): 1.67 / 3.71

Number of samples: 100

sample average: **30.745521 grams**
sample variance: **2.757318 grams^2**

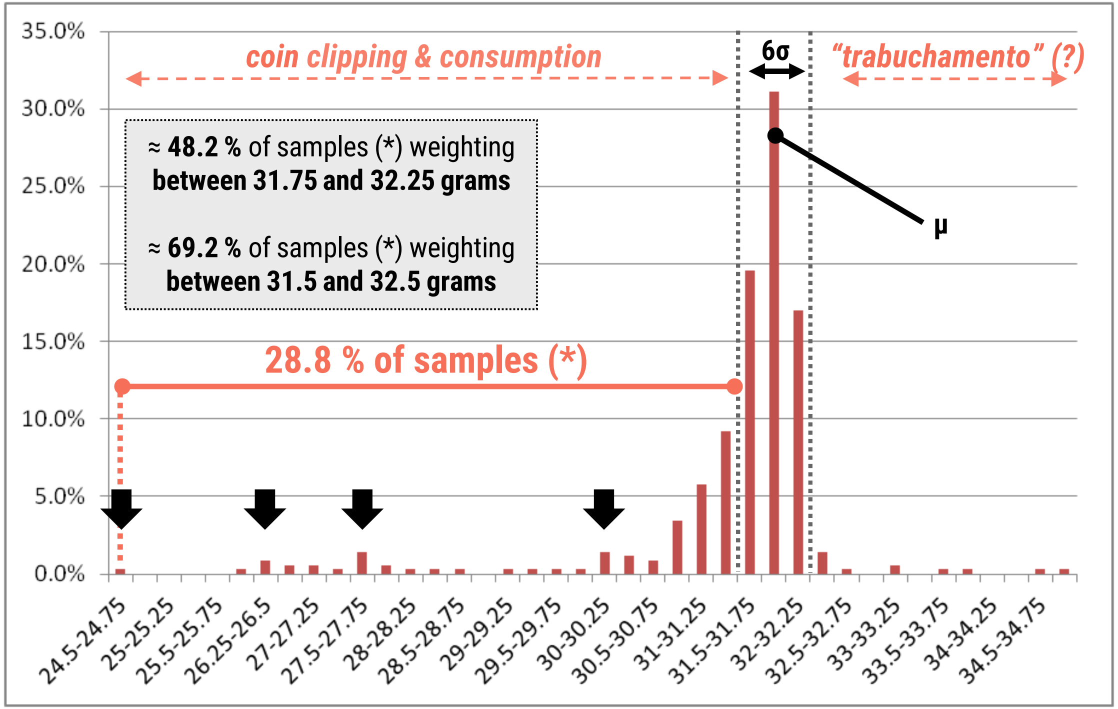
t-Student quantile (95%): **1.984217**
chi-squared quantiles (95%): **128.421989 / 73.361080**

estimated weight (95%): 30.42 / 31.08
estimated variance (95%): 2.13 / 3.72

Use case: estimation of a quantitative property

μ ≈ theorethical weight
3σ ≈ *remedium in pondere*

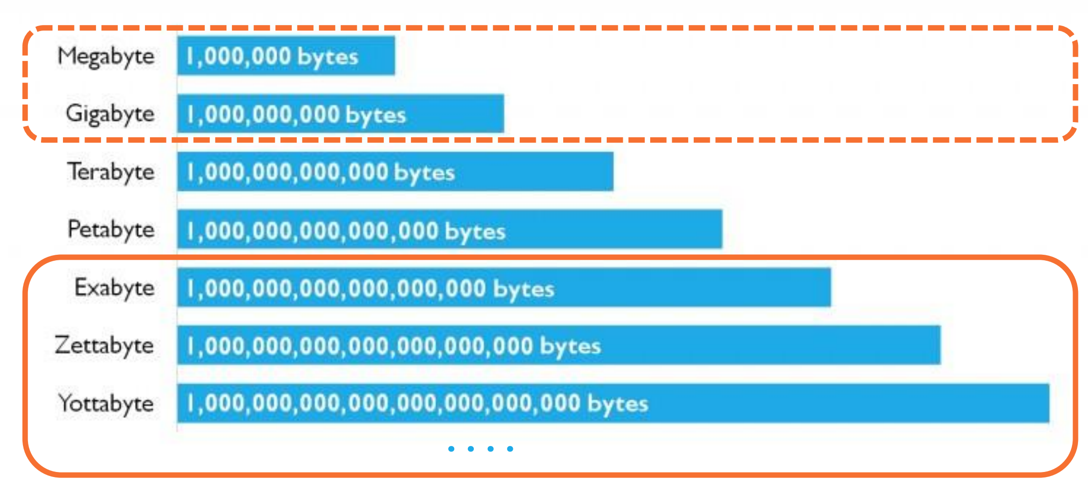
Such a large amount of **samples weighting less than μ – 3σ grams** leads the statistic **not to pass statistical hypothesis tests** (e.g., χ^2 **test**) **if the data were used to estimate the theoretical weight**.



(*) analysis conducted on 1,200+ DUCATONI from Northern Italy

… and now about Computer Science

“Small” data vs. “big” data: different IT architecture paradigms



- Local file systems
- Relational databases (RDBMS)
- Local storage solutions (2.5” HDD)
- Structured data

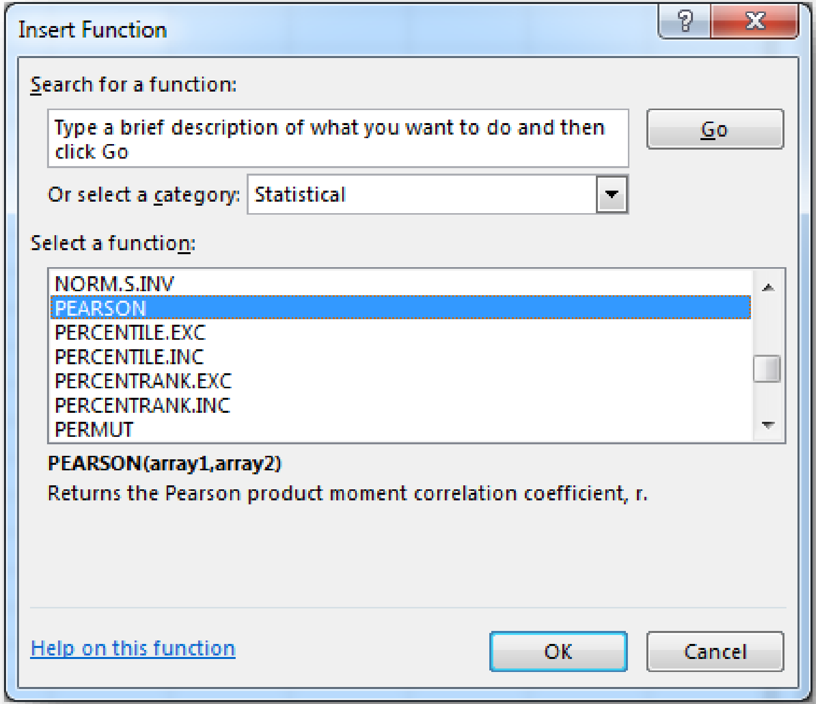
- Distributed file systems
- NoSQL databases
- Distributed data store
- Unstructured / semi-structured data

The amount of data to be managed in coin find recording is «small» from the IT point of view, so there are convenient solutions for DATA STORING AND PROCESSING that are also accessible in home computing and/or small academic networks

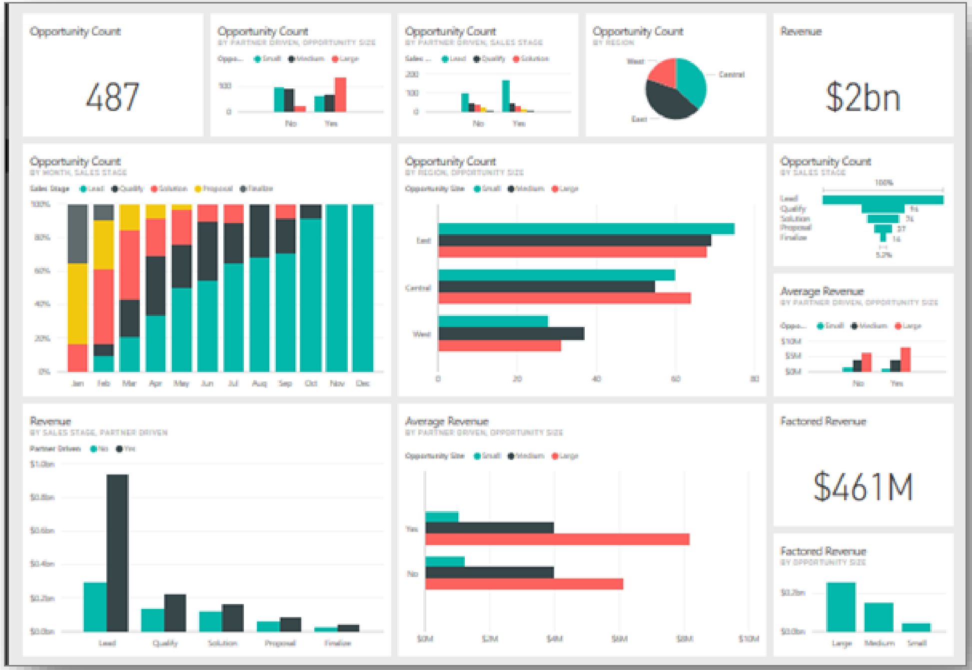
Data storing and processing

Relational databases RDBMS (e.g. Microsoft Access, MySQL, PostgreSQL) allow the storing of a data set much larger than that determined by all known coin finds.

Spreadsheets (e.g. Microsoft Excel) offer a complete set of statistical functions.
The enhanced capacity in terms of rows and columns allows to easily manage millions of data – both quantitative and qualitative – in a tabular form.



Data storing and processing



Data analytics platforms (e.g. Qlik) offers self-service visualization, guided and embedded analytics and reporting capabilities.

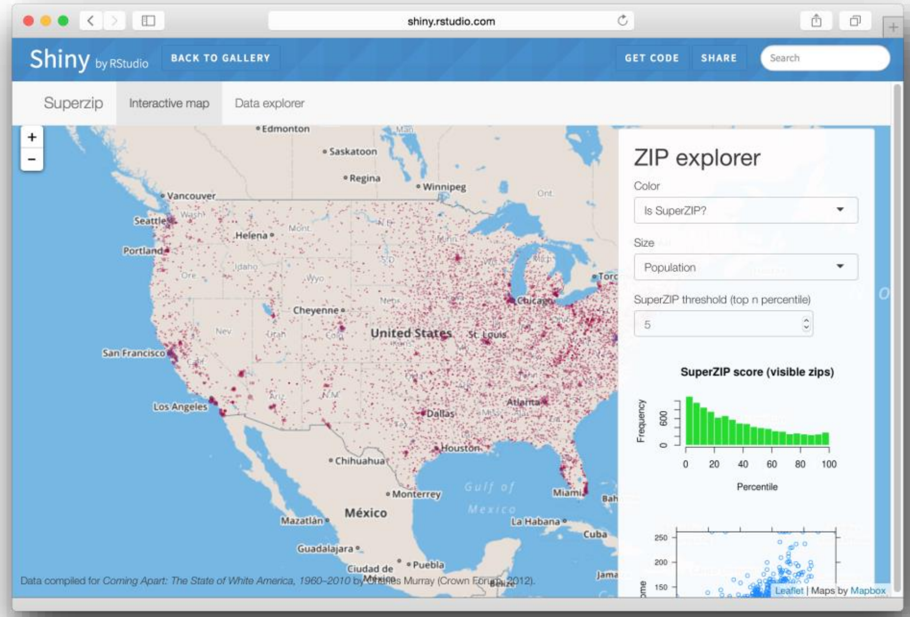
They usually provide a dynamic, highly customizable dashboard connected to a data set (e.g., ODBC database, OLE DB database , local or network file folders, web URLs), with plenty of pre-defined templates.

Data storing and processing

Dozens of free **APIs, plugins and widgets** are available on the Internet for spatial visualization and data processing.

Use case: ***Leaflet*** (one of the most popular open-source JavaScript libraries for interactive maps) combined with ***Shiny*** (web framework for R, a language and environment for statistical computing and graphics applications) to achieve enhanced spatial visualization of the data set.

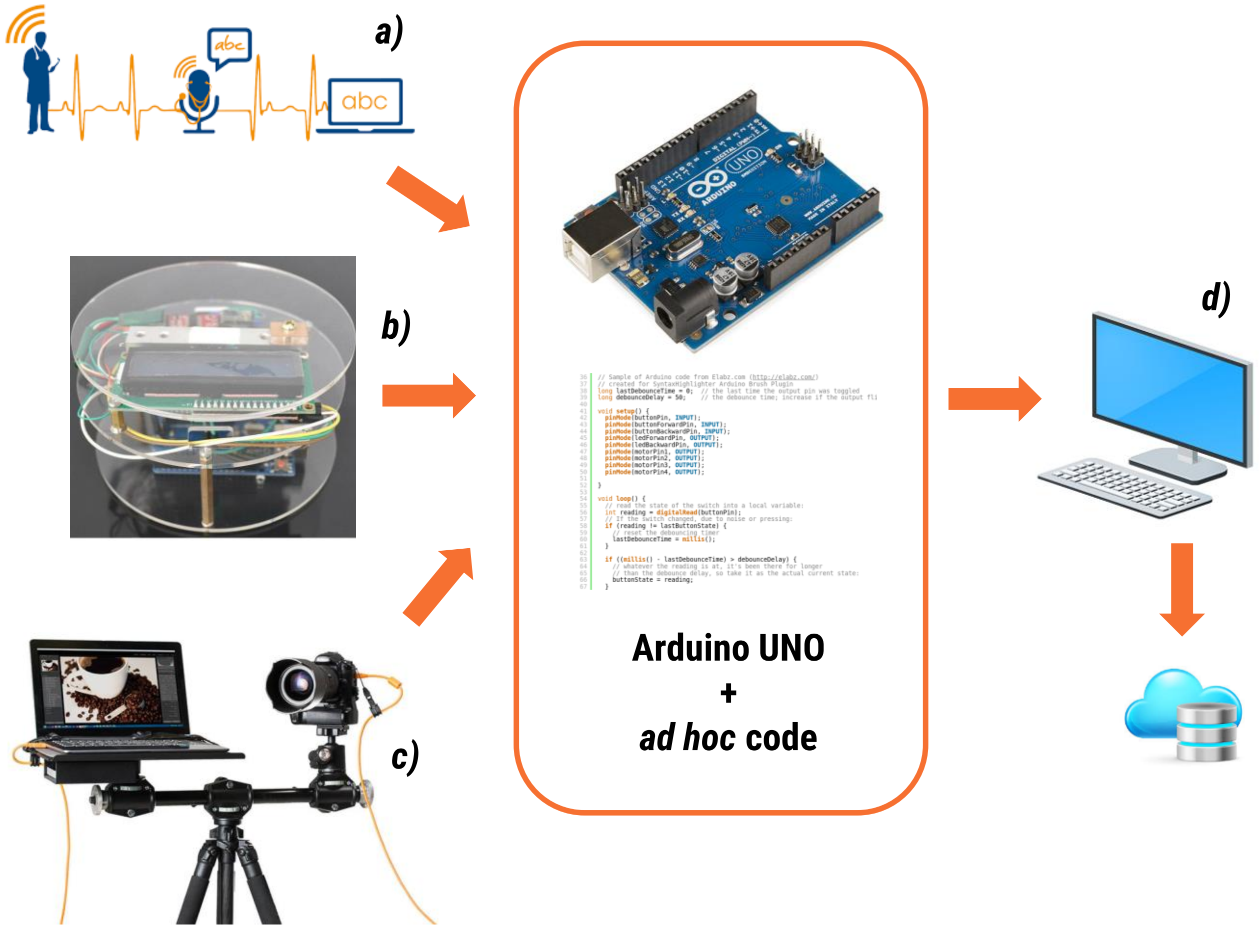
(*source*: https://rstudio.github.io/leaflet/)



Looking for smart ways to digitalize data

Simultaneous data acquisition based on the Arduino UNO board:

a) speech recognition data entry
b) weight sensor
c) tethered shooting
d) open source APIs



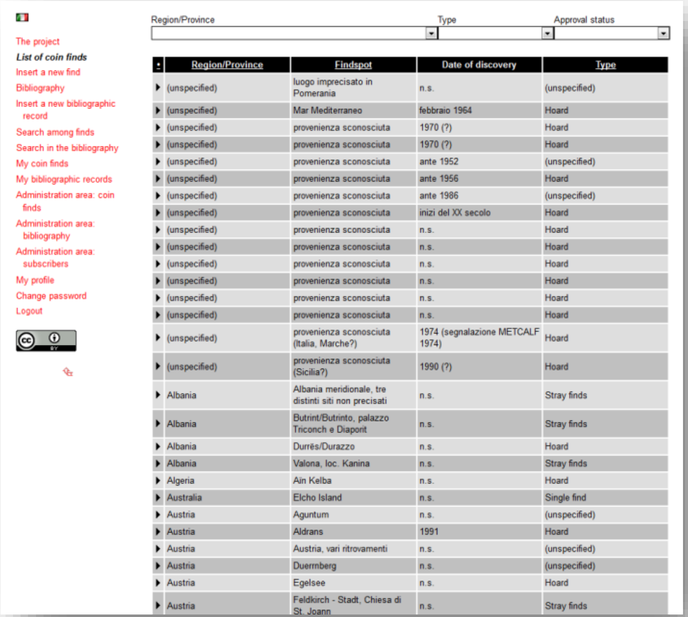
Looking for smart ways to digitalize data

A “wiki” approach to the recording of coin finds

Portable Antiquities Scheme

www.finds.org.uk

https://www.sibrium.org/CoinFinds/

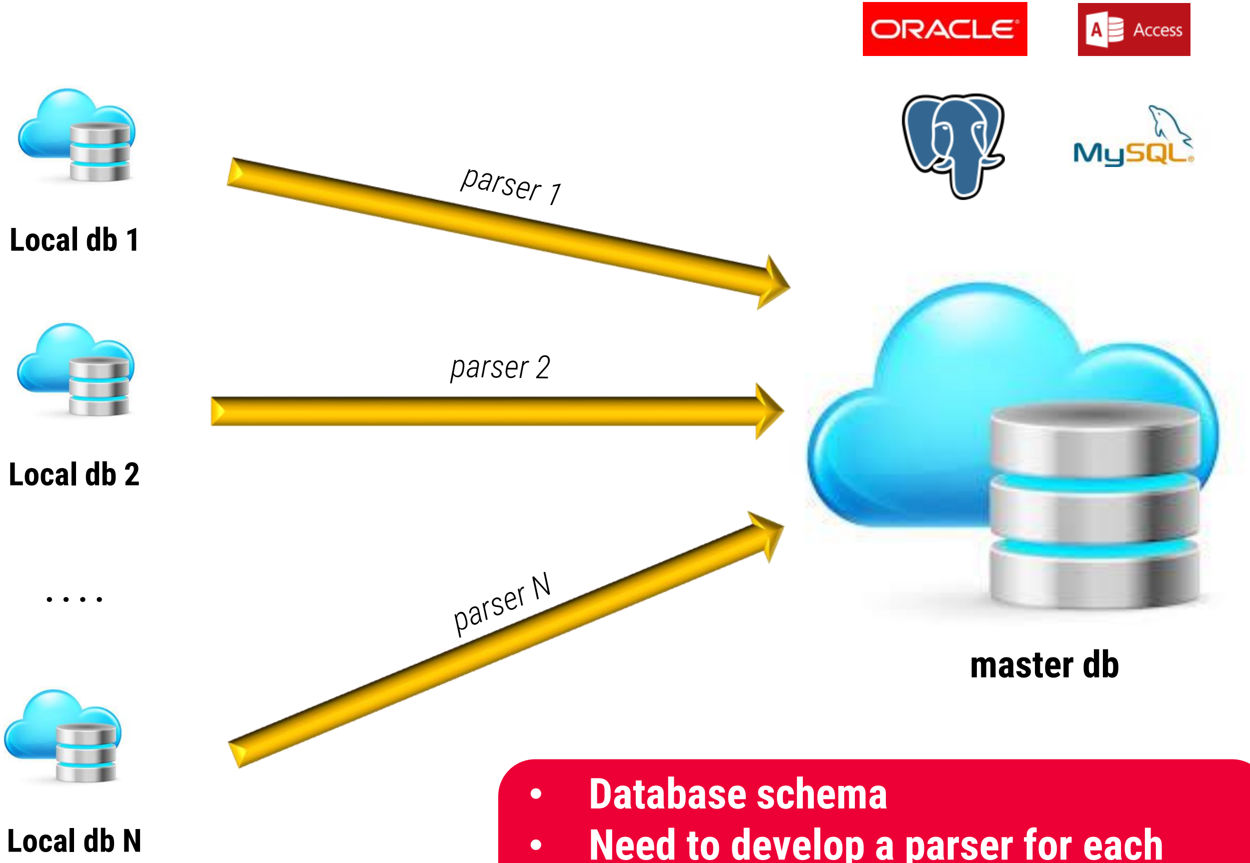

The project
List of coin finds
Insert a new find
Bibliography
Insert a new bibliographic record
Search among finds
Search in the bibliography
My coin finds
My bibliographic records
Administration area: coin finds
Administration area: bibliography
Administration area: subscribers
My profile
Change password
Logout

Region/Province
Type
Approval status

Region/Province	Findspot	Date of discovery	Type
(unspecified)	luogo imprecisato in Pomerania	n.s.	(unspecified)
(unspecified)	Mar Mediterraneo	febbraio 1964	Hoard
(unspecified)	provenienza sconosciuta	1970 (?)	Hoard
(unspecified)	provenienza sconosciuta	1970 (?)	Hoard
(unspecified)	provenienza sconosciuta	ante 1952	(unspecified)
(unspecified)	provenienza sconosciuta	ante 1956	Hoard
(unspecified)	provenienza sconosciuta	ante 1986	(unspecified)
(unspecified)	provenienza sconosciuta	inizi del XX secolo	Hoard
(unspecified)	provenienza sconosciuta	n.s.	Hoard
(unspecified)	provenienza sconosciuta	n.s.	Hoard
(unspecified)	provenienza sconosciuta	n.s.	Hoard
(unspecified)	provenienza sconosciuta	n.s.	Hoard
(unspecified)	provenienza sconosciuta	n.s.	Hoard
(unspecified)	provenienza sconosciuta	n.s.	Hoard
(unspecified)	provenienza sconosciuta (Italia, Marche?)	1974 (segnalazione METCALF 1974)	Hoard
(unspecified)	provenienza sconosciuta (Sicilia?)	1990 (?)	Hoard
Albania	Albania meridionale, tre distinti siti non precisati	n.s.	Stray finds
Albania	Butrint/Butrinto, palazzo Triconch e Diaporit	n.s.	Stray finds
Albania	Durrës/Durazzo	n.s.	Hoard
Albania	Valona, loc. Kanina	n.s.	Stray finds
Algeria	Aïn Kelba	n.s.	Hoard
Australia	Elcho Island	n.s.	Single find
Austria	Aguntum	n.s.	(unspecified)
Austria	Aldrans	1991	Hoard
Austria	Austria, vari ritrovamenti	n.s.	(unspecified)
Austria	Duerrnberg	n.s.	(unspecified)
Austria	Egelsee	n.s.	Hoard
Austria	Feldkirch - Stadt, Chiesa di St. Joann	n.s.	Stray finds

Data reuse and data convergence

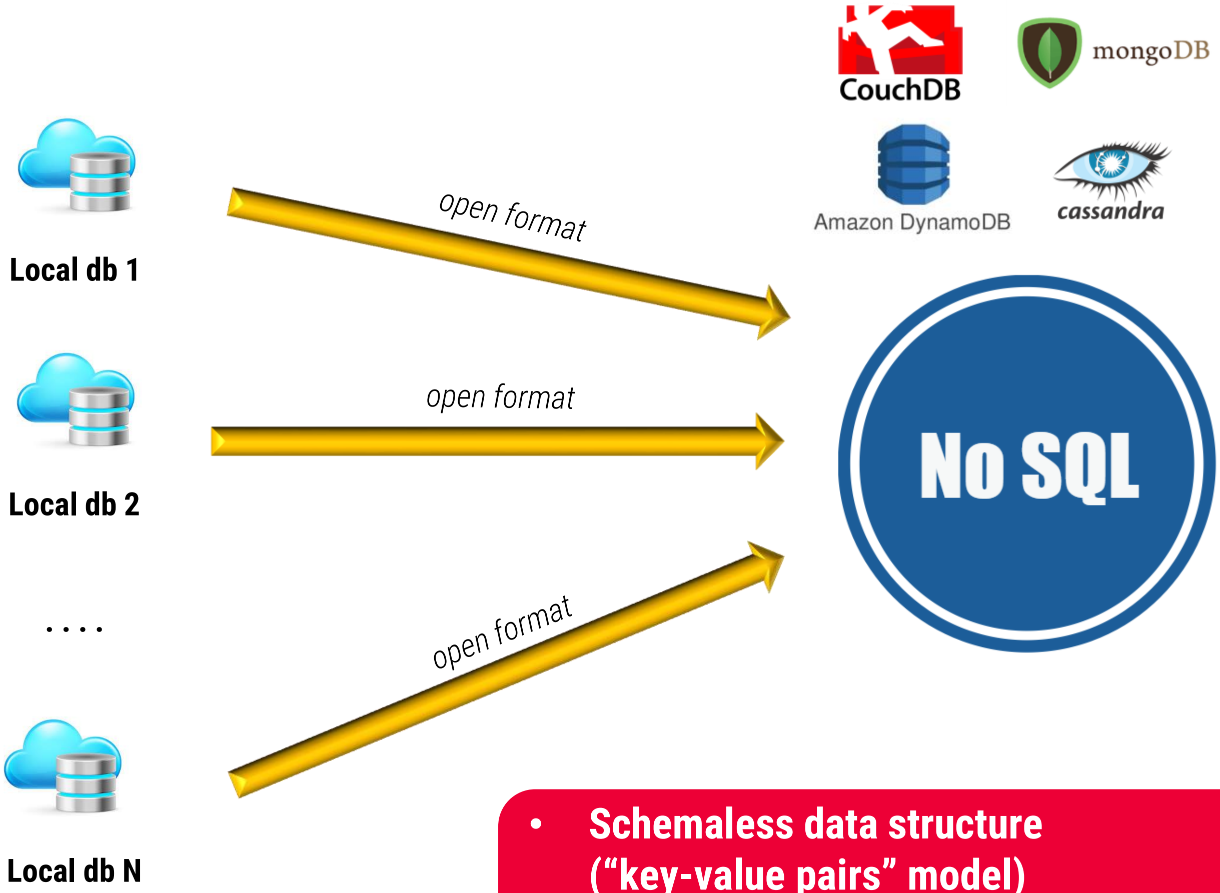
Integrating existing RDBMSs into a "master" RDBMS



- Database schema
- Need to develop a parser for each RDBMS, to adapt the local schema to a reference schema

Data reuse and data convergence

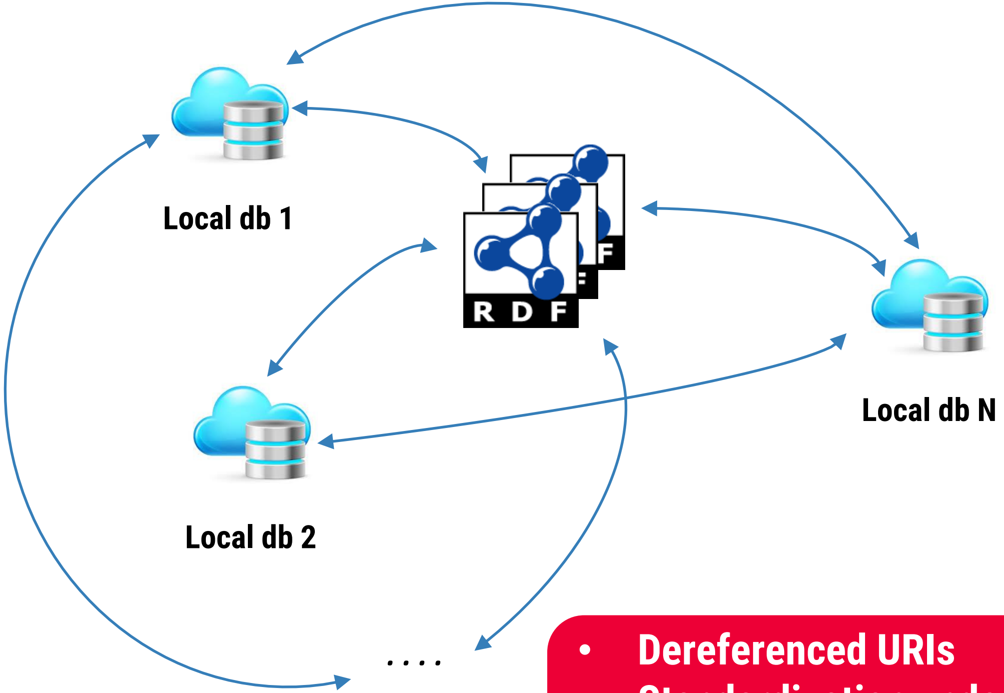
Integrating existing RDBMSs into a noSQL database



- Schemaless data structure ("key-value pairs" model)
- No need to develop local parsers
- Queries on JavaScript

Data reuse and data convergence

Linked open data and semantic web



- Dereferenced URIs
- Standardization under the care of World Wide Web Consortium (W3C)
- Queries on e.g. SPARQL

The risks from IT Obsolescence and Short Term Projects

Computer Science is evolving at the speed of light, **IT solutions may become obsolete** in a flash.

Projects of digitalization and data recording can shut down due to lack of funds or people leaving.

Need to have OPEN DATA, in non-proprietary file format (e.g. CSV, XML, JSON), accessible in a universally defined location

JSON

CHRE-3406.json


```
{
  "id": 3406,
  "findSpotName": "REKA DEVNIA 1929",
  "coinCount": 101567,
  "findSpotOtherNames": "Marcianopolis I(I), Reka Devyna, Réka-Devnia; Девня",
  "address": "",
  "city": "Devnya",
  "county": "Varna",
  "region": "",
  "lattitude": 43.225,
  "longitude": 27.585,
  "altitude": 25.307,
  "comment": "Coins run from Mark Antony to Trajan Decius, although it has been argued that the last coins could be
  intrusive. In that case, the hoard would end with issues of Gordian III. 81,096 (and not 81,044!) coins were published in
  1934 by Mouchmov, after a further 20,000+ coins had been dispersed. The total number of coins therefore exceeds 101,096
  specimens. The difference between our comput and the 81,044 coins mentioned in the original publication resides in an error
  in counting the number of coins of Commodus (3146, whereas 3191 is correct) and reporting the correct number of coins of
  Caracalla in the table p. 6 of Mouchmov's publication. Note that further errors have occurred for Trajan (p. 21, last three
  entries, where numbers do not add up).",
  "findSpotComment": "",
  "discoveryDay1": 10,
  "discoveryMonth1": 11,
  "discoveryYear1": 1929,
  "discoveryDay2": 10,
  "discoveryMonth2": 11,
  "discoveryYear2": 1929,
  "openingYear1": -32,
  "openingYear2": -31,
  "terminalYear1": 243,
  "terminalYear2": 251,
  "reference_string": "",
  "references": "Moushmov, N.A. - 1934 - Le trésor numismatique de Réka Devnia (Marcianopolis); Fitz, J. - 1978 - Der
  Geldumlauf der römischen Provinzen im Donaugebiet Mitte des 3. Jahrhunderts: pp. 119-121 and 254-255; Găzdac, C. - 2002 -
  Monetary circulation in Dacia and the provinces from the Middle and Lower Danube from Trajan to Constantine I (AD 306-337):
  pp. 544-5; Paunov, E. and Prokopov, I. - 2002 - An Inventory of Roman Republican Coin Hoards and coins from Bulgaria
  (IRRCHBulg): pp. 48-50, no. 75; Depeyrot, Georges - 2004 - La propagande monétaire (64-235) et le trésor de Marcianopolis
  (251); Metcalf, W.E. - 2002 - The Reka Devnia hoard re-examined.; Szaivert, W. - 1985 - Einige Bemerkungen zum Fund von
  Réka-Devnia.; Moushmov, N. - 1930 - Une trouvaille de monnaies antiques près du village de Reka-Devnia (Marcianopolis)",
  "toEnterAtCoinLevel": "",
  "discoveryDepth": "2 m",
  "owner": "National Museum, Sofia (ca. 68,783 coins). Archaeology Museum, Varna, Bulgaria: inv.nos. 2807-2829, 4301-16363
  (12,261 coins: Nero - Gordian III)",
  "finder": "Pavel Todorov",
  "discoveryComment": "The hoard was put into 7 crates. 6 of these were sent to the Museum in Sofia (236 kg) and one remained
  in Varna (50 kg).",
  "archaeologyStartYear": "",
```

XML

rpc-3-2834.xml


```xml
<?xml version="1.0" encoding="utf-8"?><nuds xmlns="http://nomisma.org/nuds" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xs="
http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" recordType="conceptual">
            <control>
            <recordId>rpc-3-2834</recordId>
            <publicationStatus><publicationStatus>approved</publicationStatus></publicationStatus>
            <maintenanceAgency><agencyName>University of Oxford</agencyName></maintenanceAgency>
            <maintenanceStatus>new</maintenanceStatus>
            <maintenanceHistory>
                 <maintenanceEvent>
                    <eventType>derived</eventType>
                    <eventDateTime standardDateTime="2017-5-2T11:11:11">Tue, 2 May 2017 11:11:11</eventDateTime>
                    <agentType>machine</agentType>
                    <agent>PHP</agent>
                    <eventDescription>Generated from MySQL database</eventDescription>
                </maintenanceEvent>
            </maintenanceHistory>
            <rightsStmt>
                <copyrightHolder>Ashmolean Museum, University of Oxford</copyrightHolder>
                <license xlink:type="simple" xlink:href="http://rpc.ashmus.ox.ac.uk/copyright/"/>
            </rightsStmt>
            <semanticDeclaration>
                <prefix>dcterms</prefix>
                <namespace>http://purl.org/dc/terms/</namespace>
            </semanticDeclaration>
        </control>
        <descMeta>
            <title xml:lang="en">RPC III, 2834</title>
            <typeDesc>    <objectType xlink:type="simple" xlink:href="http://nomisma.org/id/coin">Coin</objectType>
                <manufacture xlink:type="simple" xlink:href="http://nomisma.org/id/struck">Struck
                </manufacture><authority><persname xlink:type="simple" xlink:role="authority" xlink:href="
                http://nomisma.org/id/nerva">Nerva</persname></authority><geographic><geogname xlink:type="simple" xlink:role=
                "mint" xlink:href="http://nomisma.org/id/ancyra">Ancyra</geogname><geogname xlink:type="simple" xlink:role=
                "region" xlink:href="http://nomisma.org/id/galatia">Galatia</geogname></geographic>     <obverse>
                    <legend>ΑΥΤΟ ΝΕΡΟΥΑΣ ΚΑΙΣΑΡ ΣΕΒΑΣΤΟΣ</legend>
                    <type>
                        <description xml:lang="en">laureate head of Nerva, r.</description>
                    </type><persname xlink:type="simple" xlink:role="portrait" xlink:href="http://nomisma.org/id/nerva">Nerva
                    </persname>     </obverse>
                <reverse>
                    <legend>ΕΠΙ (Τ)(ΙΤ) ΠΟΜΠΩΝΙΟΥ ΒΑΣΣΟΥ (ΠΡ)</legend>
                    <type>
                        <description xml:lang="en">Mên standing l., with chiton, cloak and oriental tiara, crescent on shoulder,
                        holding patera in his extended r. hand</description>
                    </type>
                </reverse>
            </typeDesc>
        </descMeta>
    </nuds>
```

RDF

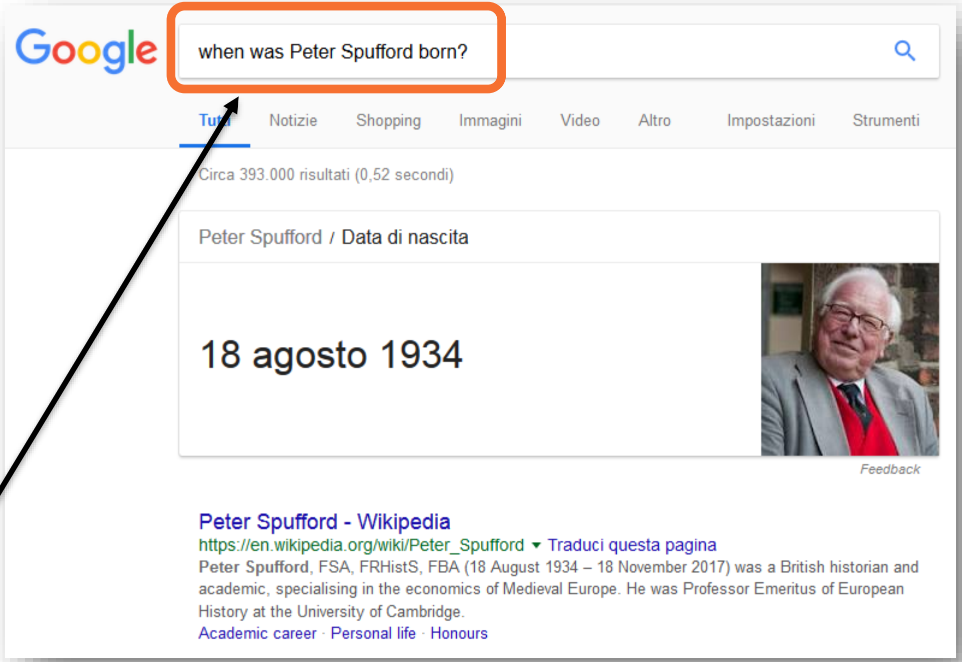
```xml
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:nmo="http://nomisma.org/ontology#"
         xmlns:crm="http://www.cidoc-crm.org/cidoc-crm/"
         xmlns:dcterms="http://purl.org/dc/terms/"
         xmlns:oa="http://www.w3.org/ns/oa#"
         xmlns:pelagios="http://pelagios.github.io/vocab/terms#"
         xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
         xmlns:void="http://rdfs.org/ns/void#"
         xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
         xmlns:dcmitype="http://purl.org/dc/dcmitype/"
         xmlns:skos="http://www.w3.org/2004/02/skos/core#"
         xmlns:relations="http://pelagios.github.io/vocab/relations#"
         xmlns:nm="http://nomisma.org/id/"
         xmlns:foaf="http://xmlns.com/foaf/0.1/"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:xsd="http://www.w3.org/2001/XMLSchema#">

    <nmo:TypeSeriesItem rdf:about="http://rpc.ashmus.ox.ac.uk/coins/3/2834">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:prefLabel xml:lang="en">RPC III, 2834</skos:prefLabel>
    <skos:definition xml:lang="en">RPC III, 2834</skos:definition>
    <nmo:representsObjectType rdf:resource="http://nomisma.org/id/coin"/>
    <nmo:hasManufacture rdf:resource="http://nomisma.org/id/struck"/>
    <nmo:hasAuthority rdf:resource="http://nomisma.org/id/nerva"/>
    <nmo:hasMint rdf:resource="http://nomisma.org/id/ancyra"/>
    <nmo:hasRegion rdf:resource="http://nomisma.org/id/galatia"/>
    <nmo:hasObverse rdf:resource="http://rpc.ashmus.ox.ac.uk/coins/3/2834#obverse"/>
    <nmo:hasReverse rdf:resource="http://rpc.ashmus.ox.ac.uk/coins/3/2834#reverse"/>
    <void:inDataset rdf:resource="http://rpc.ashmus.ox.ac.uk"/>
    </nmo:TypeSeriesItem>

    <rdf:Description rdf:about="http://rpc.ashmus.ox.ac.uk/coins/3/2834#obverse">
    <nmo:hasLegend>ΑΥΤΟ ΝΕΡΟΥΑΣ ΚΑΙΣΑΡ ΣΕΒΑΣΤΟΣ</nmo:hasLegend>
    <dcterms:description xml:lang="en">laureate head of Nerva, r.</dcterms:description>
    <nmo:hasPortrait rdf:resource="http://nomisma.org/id/nerva"/>
    </rdf:Description>

    <rdf:Description rdf:about="http://rpc.ashmus.ox.ac.uk/coins/3/2834#reverse">
    <nmo:hasLegend>ΕΠΙ (Τ)(ΙΤ) ΠΟΜΠΩΝΙΟΥ ΒΑΣΣΟΥ (ΠΡ)</nmo:hasLegend>
    <dcterms:description xml:lang="en">Mên standing l., with chiton, cloak and oriental tiara, crescent on shoulder, holding
    in his extended r. hand</dcterms:description>
    </rdf:Description>

</rdf:RDF>
```

rpc-3-2834.rdf

A (hopefully) great future… but not for us?

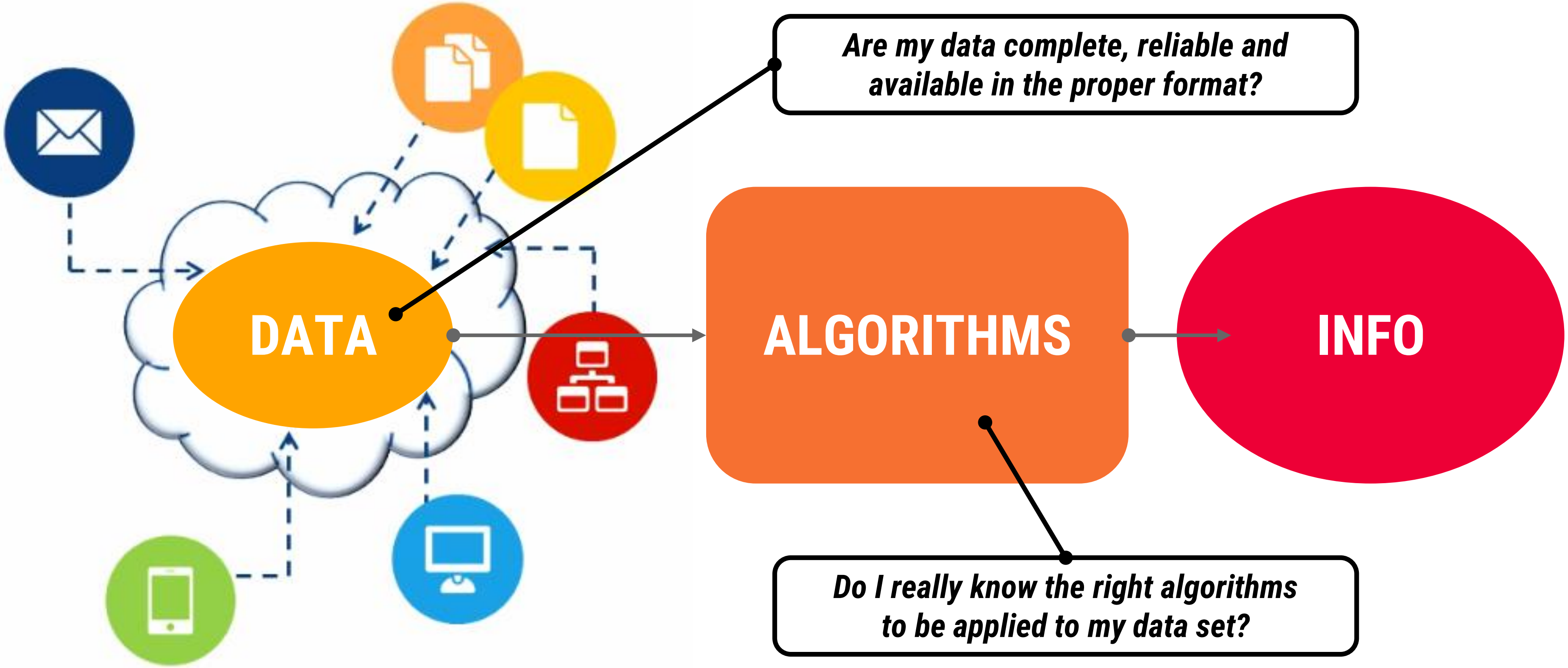
Not all the mature technologies of daily use will find an application in Numismatics in a short term (***no return on investments***):

- **PATTERN RECOGNITION**
  (→ automatic identification of a picture)
- **ARTIFICIAL NEURAL NETWORKS**
  (→ automatic production of information from structured data)
- **MACHINE LEARNING**
  (→ learning by use cases, problem solving)
- **QUESTION ANSWERING**
  (→ natural language queries on data sets)



Let’s move onto the conclusions…

The challenges of Data Availability and Data Processing



«Are my data complete, reliable and available in the proper format?»

There are so many tools for processing data!

The real challenge is to have an OPEN data set that is as COMPLETE and RELIABLE as possible.

Data should be in a digital and – preferably – indexed format (e.g., schema database, key-value pairs, ontology).

«Do I really know the right algorithms to be applied to my data set?»

ALGORITHMS

Knowing what to do is fundamental! Data processing is not just Statistics or Computer Science, ***it is above all great sensitivity dictated by experience and intuition***.

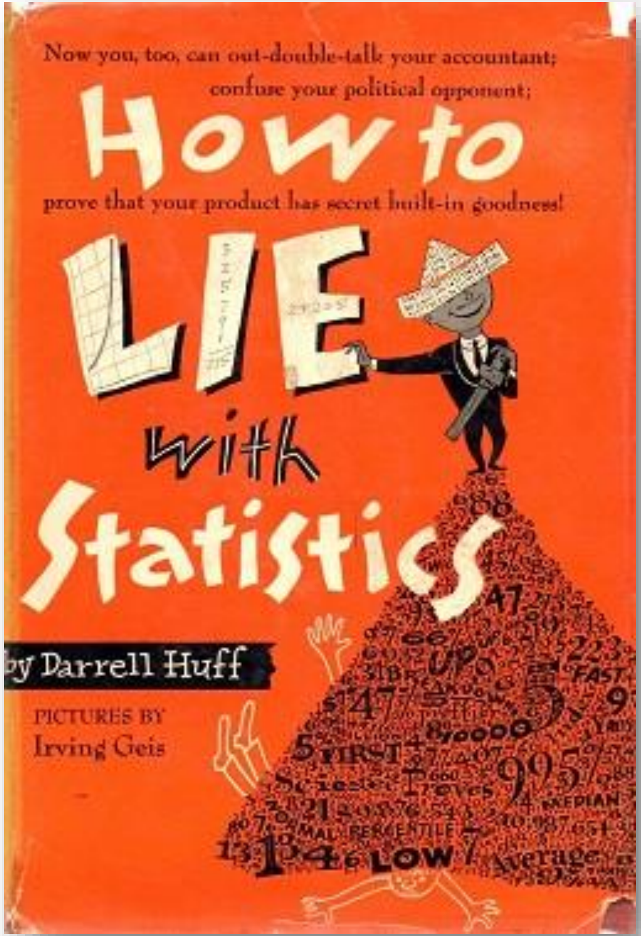
- Descriptive Statistics
- Statistical Inference
- Statistical Modelling
- Multivariate Statistics
- Spatial Analysis

Whatever your imagination can invent!

“

There are three kinds of lies: lies, damned lies, and statistics

(Benjamin DISRAELI, *attr.*)



WORK SMARTER, NOT HARDER!

It all comes to an end…

If you have any questions about this document, please don't hesitate to contact me at:

- ***https://www.sibrium.org/***
- ***mail@sibrium.org***