FAIR data or a fairy tale?

How to support data reliability for their Re-use
(in archaeology, but not only)

Franco Niccolucci

PIN

Who am I

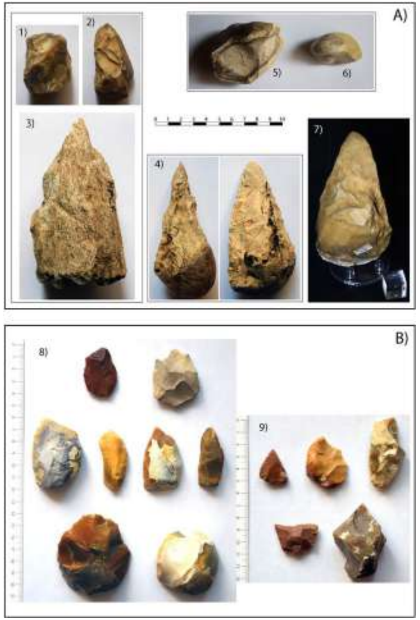
- Degree in Mathematics (1970)
- Professor, University of Pisa, then University of Florence (1970-2007)
- Founding Director of STARC the Science and Technology in Archaeology Research Centre at the Cyprus Institute, Nicosia, Cyprus (2008-2013) and Director of the PhD in Digital Cultural Heritage
- Director of the VAST Lab at PIN (2004-2023), a research and educational agency of U. Florence
- Scientific Coordinator of several EU-funded projects, the latest being ARIADNE (2013-2022) an infrastructure for accessing archaeological datasets, cataloguing more than 3 500 000 items
- Editor-in-Chief (2018-present) of the ACM Journal on Computing and Cultural Heritage (JOCCH)
- Currently, President of the ARIADNE Research Infrastructure AISBL, a research association of 40 European archaeological and research institutions

ARIADNE
Research Infrastructure AISBL

THE PREHISTORY: DATABASES, OR THE ACCOUNTANT’S COMPUTING

*Iurare in verba magistris**

- This happens when one must rely on an expert's evaluation
- Flint tools typology regarded as the main "objective" tool in establishing chronological frameworks and comparisons between sites. It relies on the "*fossile directeur*" concept
- Typology based on the characteristic tool "retouch" with "objective" parameters
- However it was stated** that «...le bon typologiste perçoit avec l'expérience après avoir analysé quelques milliers de pièces...», spending a good deal of time «...au course de longs tête-à-tête avec les outillages lithiques...»



An assemblage of flint tools

* Horace (*Epist*. I, 1, 14) "To swear in the words of a master"

** Demars P.Y., Laurent P. 1992, *Types d'outils lithiques du Paléolithique Supérieur en Europe*, Paris, CNRS

An experiment

- Experiment (2001)*: classification of lithic tools
- Assemblage of 50 flint tools from a proto-historic site in Southern Israel (all from the same occupation phase)
- 5 experts (“referees”) from the Ben-Gurion University of the Negev were asked to classify the tools and indicate which type is prevalent in the assemblage
- Different “schools” and approaches for all the experts
- Classification made with attention and with no apparent mistakes

* S. Hermon & F. Niccolucci (2002) “Estimating subjectivity of typologists and typological classification with fuzzy logic“ *Archeologia e Calcolatori*, 13, 2002, 217-232, *Proceedings of the XXIV UISPP Congress*, Liege 2001

Results

Type / Referee	Scraper	Tabular Scraper	Borer	Burin	Truncation	Retouched Flake	Notch	Denticulat ed	Retouched Blade	Sickle Blade	Retouched Bladelet	Bifacial	Varia
A	**20%**	0%	14%	2%	10%	14%	**20%**	4%	10%	2%	2%	2%	0%
B	**32%**	0%	6%	4%	0%	**18%**	6%	12%	12%	6%	2%	2%	0%
C	**28%**	2%	**20%**	0%	8%	8%	2%	16%	0%	16%	0%	0%	0%
D	6%	2%	12%	0%	6%	18%	**22%**	**20%**	6%	2%	0%	0%	6%
E	6%	0%	10%	2%	0%	**24%**	12%	**22%**	12%	4%	0%	0%	8%

- The two prevalent types are different for the five referees. The conclusions by one of them might be very different from those of another one
- This happens because the obligation to assign a specific type forces the referee to choose a type even if in doubt, thus compromising the final count
- Solution: make this uncertainty apparent, allowing referees to express their confidence in the type attribution with a value – we may call it *R*, the *reliability coefficient* – between 0 and 1 ➜ 0 = does not belong to type, 1 = surely belongs to type (but might be of another type as well)

Comparison

Assignments made by one of the referees for items 1-7. The table on the left refers to yes/no type assignments. The one on the right allows uncertain type assignments

Item No. / ***Type***	1	2	3	4	5	6	7
Scraper							
Tabular Scraper							
Borer						1	
Burin		1					
Truncation							
Retouched Flake			1		1		
Notch				1			1
Denticulated	1						
Retouched Blade							
Sickle Blade							
Retouched Bladelet							
Bifacial							
Varia							

Item No. / ***Type***	1	2	3	4	5	6	7
Scraper	0.7						
Tabular Scraper							
Borer						0.5	
Burin		1.0					
Truncation							
Retouched Flake			1.0		0.8		0.8
Notch			1.0	1.0	0.8		1.0
Denticulated	0.9		1.0		0.5	1.0	
Retouched Blade							
Sickle Blade							
Retouched Bladelet							
Bifacial							
Varia							

Comparison

0 or 1

Type Referee	Scraper	Tabular Scraper	Borer	Burin	Truncation	Retouched Flake	Notch	Denticulated	Retouched Blade	Sickle Blade	Retouched Bladelet	Bifacial	Varia
A	**20%**	0%	14%	2%	10%	14%	**20%**	4%	10%	2%	2%	2%	0%
B	**32%**	0%	6%	4%	0%	**18%**	6%	12%	12%	6%	2%	2%	0%
C	**28%**	2%	**20%**	0%	8%	8%	2%	16%	0%	16%	0%	0%	0%
D	6%	2%	12%	0%	6%	18%	**22%**	**20%**	6%	2%	0%	0%	6%
E	6%	0%	10%	2%	0%	**24%**	12%	**22%**	12%	4%	0%	0%	8%

Value between 0 and 1

Type Referee	Scraper	Tabular Scraper	Borer	Burin	Truncation	Retouched Flake	Notch	Denticulated	Retouched Blade	Sickle Blade	Retouched Bladelet	Bifacial	Varia
A	**14%**	0%	12%	1%	8%	12%	**18%**	14%	9%	6%	1%	3%	0%
B	**26%**	0%	7%	4%	0%	**17%**	12%	14%	10%	5%	1%	4%	0%
C	**24%**	2%	**20%**	1%	5%	15%	5%	12%	3%	11%	0%	2%	0%
D	5%	2%	13%	8%	8%	14%	**24%**	**20%**	5%	4%	0%	0%	6%
E	6%	0%	7%	2%	0%	**23%**	14%	**27%**	9%	3%	0%	7%	9%

Conclusions

- Referees felt more comfortable when they were allowed to indicate also alternate type choices
- The method indicates difficult artefacts – or difficult referees!
- Knowing that the type assignment includes a “reliability coefficient” helps a re-user to understand to which degree the data are trustworthy
- Even if the aggregated results may not show a substantial difference, knowing the uncertainty of each attribution by the reliability index helps re-users to further investigate on their own before relying on potentially unreliable data

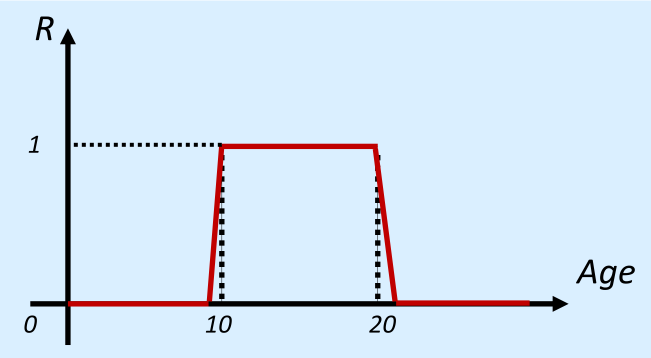
*There are three kinds of lies: lies, damned lies and statistics**

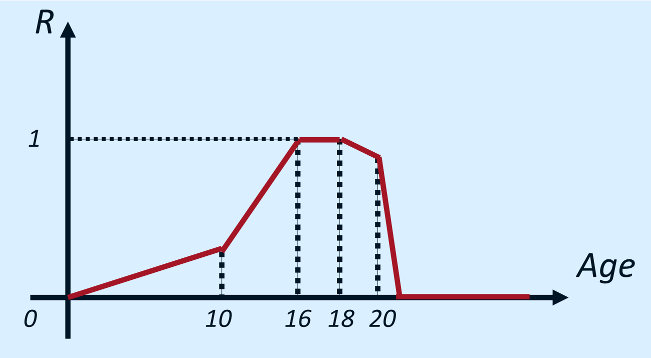
- Many archaeological investigations rely on statistics. One example is the study of cemeteries to understand the society they represent
- An excavation campaign on the Etruscan site of Pontecagnano (near Naples, Italy) in the ‘90s analysed more than 8,000 tombs and for each one it assessed the sex and age of the deceased. This was done using anthropometric methods (aDNA was not available at the time). As a result, in many cases the attribution was uncertain and the sex was recorded as “Male?” or “Female?”, while the age was recorded as ”Young?”, “Adult?” etc. , with question marks denoting the attribution uncertainty
- Are statistics made on such uncertain data reliable? What can be done to improve them?

*attributed to Mark Twain

A step forward

- It has been suggested* to store with the age range – i.e. “Child”, “Young”, “Adult” etc. – a reliability coefficient *R* in the form of a diagram relating age to confidence in age assignment
- In the cemetery case without any other information the *R* assignment is straightforward although uncertain (see top diagram). But there may be additional information, for example from the grave goods, that privileges some age range (see bottom diagram, with the range 16-20 indicated as more likely and 16-18 as almost sure)
- Then to infer knowledge about the population, when computing statistics on the deceased’s age in cemetery the values of *R* are kept into account,





Examples of fuzzy age diagrams

* F. Niccolucci, A. D’Andrea & M. Crescioli (2001) “Archaeological applications of fuzzy databases“, in Z. Stancic & T Velianovski (eds) *Proceedings of CAA2000*, Archaeopress 2001.

PRETTY PICTURES: COMPUTER VISUALIZATION (OR DECEPTION?)

*We prefer seeing to everything else**

- The visual appeal of computer data became fashionable and feasible at the beginning of this century, in archaeology especially after a 1998 CAA Conference**
- Producing visual content involves many processing steps so the relationship between the original and the digital copy may become very thin
- Starting from the VAST Euroconference in the year 2000, there are appeals to support the reliability of the visual product with additional documentation. They culminated in the manifesto for a critical approach “From CVR to CVRO”***, advocating a philological approach to computer visualization for cultural heritage
- Since then, guaranteeing credibility has become a requirement. But still…
  - ➢ Most current visual products do not offer the necessary information for the above

*Aristotle, *Metaphysics, 980 a*

** J.A. Barcelo, M. Forte & D.H. Saunders (eds.) (2000) *Virtual Reality in Archaeology*, Archeopress

*** B. Frisher, F. Niccolucci, N. Ryan & J.A. Barcelo (2001) “From CVR to CVRO: the past, present, and future of Cultural Virtual Reality” in F. Niccolucci (ed.) *Proceedings of the VAST Conference* 2000. Archaeopress

The London Charter for Computer Visualization

- The ***London Charter for Computer Visualization**** (2005) establishes principles for the computer visualization of cultural heritage. For today's discourse the most relevant ones are:
- Principle 3. Research Sources. *In order to ensure the intellectual integrity of computer-based visualisation methods and outcomes, relevant research sources should be identified and evaluated in a structured and documented way*.
- Principle 4. Documentation. *Sufficient information should be documented and disseminated to allow computer-based visualisation methods and outcomes to be understood and evaluated in relation to the contexts and purposes for which they are deployed*. This includes:
  - Documentation of Knowledge Claims: to what such visualization is aimed
  - Documentation of Research Sources: on which research the results rely
  - Documentation of Process (Paradata): in which environment the visualization was produced
  - Documentation of Methods: why computer visualization was preferred to other methods
  - Documentation of Dependency Relationships: which relationships exist between elements
  - Documentation of Formats and Standards: which formats and standards where chosen and why

*https://www.london-charter.org/

3D technology for data acquisition

- **3D scanning**, using a device (a 3D scanner) based on various technologies to acquire the position of points on the surface of the object being scanned. The result, produced by the device and its software, is a *point cloud*. Points are then interpolated (*meshing*) to obtain a surface. Critical points:
  - The device and its software are proprietary and undisclosed to users
  - The object surface (dark, shiny, transparent) may affect scan quality
  - The interpolation may be too approximative or introduce errors
- **Photogrammetry** automatically analyses a large number of digital images, combining them using specialised software to obtain the point cloud on the surface. Critical points:
  - The position of the images may be interpreted incorrectly
  - The software is undisclosed to the user
  - The interpolation may introduce errors

3D-scanning an archaeological site

An example of good practice

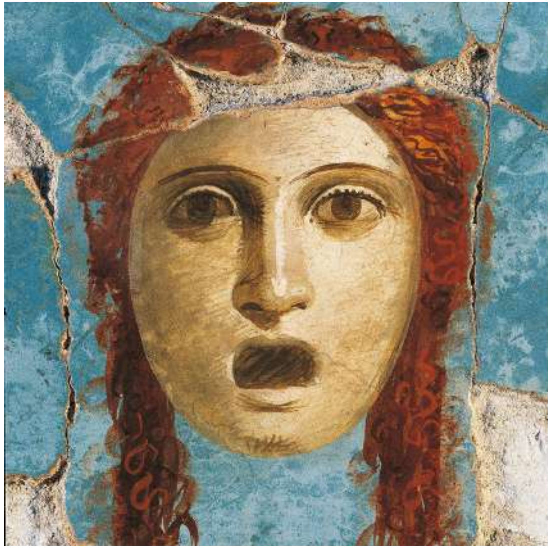
- The *Uffizi Digitization Project**: 3D digitization (2016) of the complete collection of Greek and Roman sculpture in the Uffizi Gallery, Pitti Palace, and Boboli Garden in Florence, including some 1250 works of art
- The table lists the metadata and the paradata of the digitization process, which used photogrammetry

Object Metadata	Photo Paradata	Model Paradata
Identifiers	***Camera Body***	***Photogrammetry***
Unique record #	Model	Aligned photos
Title/Name	Serial#	Tie points
Uffizi inventory #	Resolution	Reprojection error
Description	White balance	Matching mode
Basic description	Color temperature	Dense cloud size
Characteristics	ISO	Meshing mode
Format	***Lens***	Raw mesh size
Authors	Model	Texture size
Date	Serial#	Targets STD (mm)
Materials	Aperture	Phogrammetrist
Inscription	***Photo acquisition***	Processing date
Dimensions	Average GSD (mm)	***Modeling***
Provenience	File Range	SW program(s)
Surface appearance	Number of Photo	Interventions
References	Photographer name	Mesh size
Physical location	Shooting date	Major errata
Link to bibliography	***Photo processing***	Name of modeler
Rights	Software used	Date of modeling
Recording	Major edits	
Note taker	Link to edits PDF	
Date of document.	Lens Correction	
Access constraints	Processor name	
Transcriber name	Processing date	
Transcription date		

* G. Guidi, U. S. Malik, B. Frischer, C. Barandoni & F. Paolucci (2017) "The Indiana University-Uffizi project: Metrologica! challenges and workflow for massive 3D digitization of sculptures ", *23rd Int. Conf. on Virtual System & Multimedia (VSMM), Dublin, Ireland, 2017, pp. 1-8*

*O quanta species, cerebrum non habet**

- A 3D model is just a pretty picture if its paradata and metadata are not properly recorded
- Even if it is properly documented, additional information must be attached to it and to its parts (*annotations*) unless its aim is just visualization
  - It may be necessary to identify its parts, for example in a building the roof, a beam, a pillar etc. to which annotations are attached
  - The point to which an annotation is attached may be removed in further visual processing
  - The 3D model is only a way to represent the object appearance and material shape. The digital counterpart of the real asset includes also global information, such as for example relationships to other assets, immaterial components of the asset, general documentation and so on, which are attached nowhere on the object



Roman mask from a Pompeii fresco

* Phaedrus *The fox and the tragic mask* (*Fabulae*. I, 7) “O rare the headpiece, if but brains were there“

WHEN AND WHERE

*Quid est ergo tempus?**

Si nemo ex me quaerat, scio; si quaerenti explicare velim, nescio

- According to Colin Renfrew** «Dating is crucial to archaeology. Without a reliable chronology the past is chaotic». But it is often impossible to get exact dating, so comparing two different sources may become impossible
- Comparing two different time periods is based on the so-called “Allen algebra” which lists all the possible overlapping relationships between the two periods
- Dating heavily depends on the granularity of time measurement, which can vary from the hour to the day, month, year, decade, century and so on. Below the minimum granularity, there is no way of further analysing the time relationship
- In conclusion, time periods with sharp beginning and end are possible only with reference to *fiat**** time boundaries, i.e. those starting and ending according to human-made rules. Otherwise the time boundaries are *bona fide* ones, subject to uncertainty and imprecision

*Augustinus, *Confessiones*, XI, 14, 17 ”*What is then time? If nobody asks me, I know. If I want to explain to whom is asking, I ignore*”

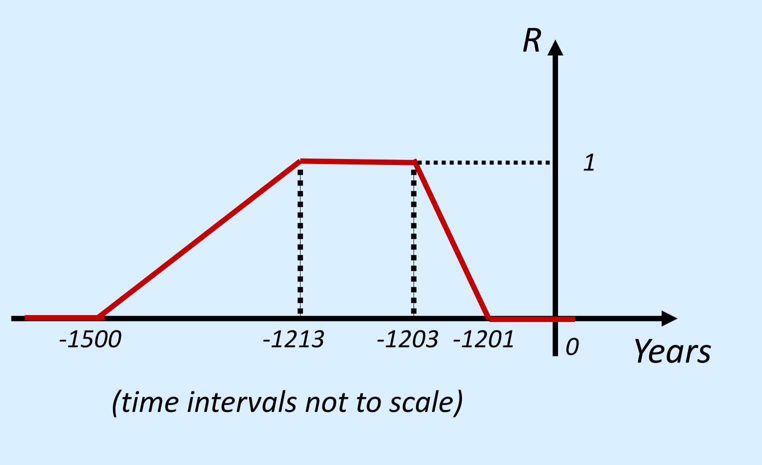
**C. Renfrew (1973) *The Explanation of Culture Change*, 20. Duckworth

***B. Smith & A.C. Varzi (2000) ”Fiat and Bona Fide Boundaries”, *Philosophy and Phenomenological Research* 60(2), 401-420. It introduces the concept for space rather than for time.

An archaeological example

The following example* concerns the time span of existence and seafaring of the so-called ”Sea Peoples”

- According to inscriptions, they were surely active during the reign of Merneptah (1213–1203 BC)
- Their existence took place during the Late Bronze Age (15th to 13th century BC)
- Even without any written evidence, it is not excluded that they existed before or after Merneptah’s reign
- The confidence *R* in the time span of their existence may therefore be represented by this diagram

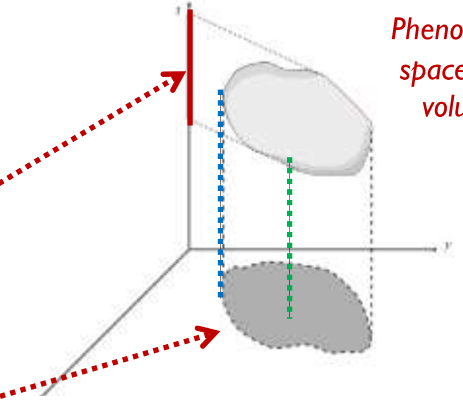


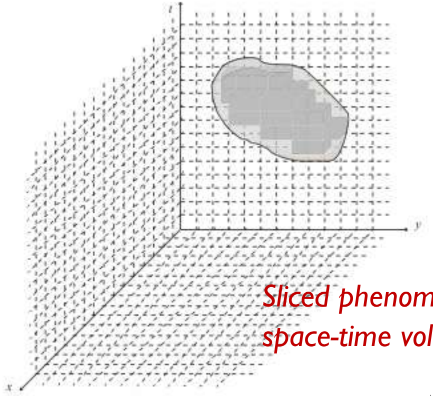
The existence time span of Sea Peoples

* F. Niccolucci & S. Hermon (2015) “Time, Chronology and Classification”, in J. A. Barcelo & I. Bogdanovic (eds.) *Mathematics and Archaeology*, CRC Press 2015, 257- 271

Space-time gazetteers

- Gazetteers are list of places associated to placenames. Places may also be associated to events, for example the place of the battle of Cannae (216 BC)
- The place corresponding to a placename changes in time so the combination of place and time is better represented by a *phenomenal space-time volume* in the 4D space (*x, y, z, t*)
- The projection on the *t* axis gives the time span of existence
- The projection on the (*x, y, z*) plane gives the overall space occupation along all the time of existence
- Since the time reference is not continuous (e.g. as referenced in historical documents) the representation of a determined space-time volume consists in a combination of cylinders (*slices*) indicating the space occupied during the validity of a historical attestation. They approximate the real (but inaccessible) space-time volume, as a potato made with Lego bricks approximates the shape of a real potato. No consideration can be reliably made on time periods below the available granularity





Fiat and bona fide boundaries

- *Fiat boundaries** – and the resulting objects defined by them – are artificial borders defined by social conventions such as, for example, a law, an official act, etc.
- *Bona fide boundaries* – and the resulting objects defined by them – are natural borders based on physical features, which exist independently of human perception
- An archaeological site is a (bona fide) geographic object defined in a conceptual way. It is a human construct as well as a portion of space, where the distinction of what is part of it and what is not, is blurred, and may change in time due to new discoveries
- In this case indeterminacy is caused by imperfect knowledge
- This indeterminacy applies also to the phenomenal space-time volume as it concerns the site occupation in time

Map of the Pompeii archaeological site at 79 BC, with a possible buffer zone

*Firstly introduced in B. Smith & A. C. Varzi (2000) "Fiat and Bona Fide Boundaries", *Philosophy and Phenomenological Research* 60, 401–420. See also: B. Smith (2001) "Fiat Objects", *Topoi* 20. The topic is addressed in many other papers and is a the subject of lively philosophical discussion.

THE FALSE CONFIDENCE IN SCIENCE

*Provando e Riprovando**

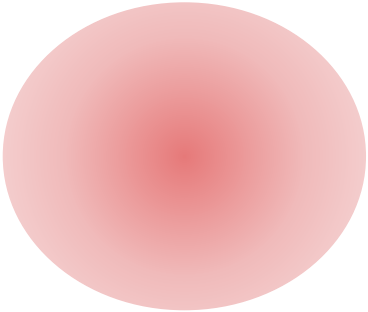
- Archaeological sciences – the use of scientific analyses to answer to an archaeological research question – seem to provide an undisputable evidence
- A counterexample: the dating and the analyses on the Turin Shroud (the *Sindone*) were carried out with all possible techniques and with the participation of an incredible number of competent scientists, with inconclusive results
- So, why should we unconditionally accept the result of one experiment carried out by one scientist or by a single team? Also in this case the confidence in the scientific results should be expressed in a clear way
- This confidence index should combine the confidence in the analyses and the reliability of the analyst

**Dante, Divina Commedia, Paradiso*, c. III, v. 3. "*proving and disproving*" but also literally *"trying and trying again"*
Motto of the *Accademia del Cimento,* a learned society founded in Florence in 1657 which established the rules of the scientific method

SO WHAT?
A POSSIBLE SOLUTION TO ASSESS RELIABILITY

Acknowledging the fuzziness of world description

- Fuzzy logic is based on the concept of degree of truth of a statement, expressed with a number between 0 and 1, as opposed to binary logic for which a statement can only be true or false
- The degree of truth is NOT a probability: it models vagueness while probability models ignorance
- Fuzzy logic is the counterpart of the fuzzy set theory with the degree of truth corresponding to the fuzzy membership in fuzzy sets
- Fuzzy statements can be composed according to precise rules for AND and OR logical operators
- This degree of truth corresponds to the reliability coefficient as defined in the previous examples



Representing the membership in a fuzzy set with the intensity of colour

Quantifying reliability: the two thieves' rule

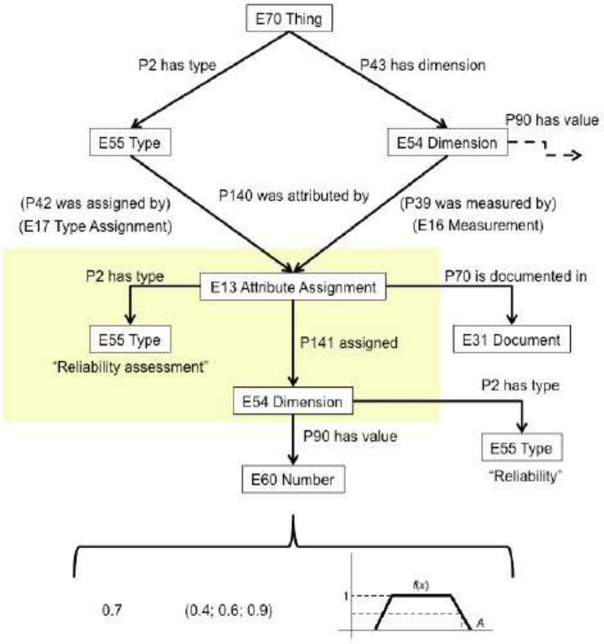
- Bruno De Finetti introduced* (1970) the subjective theory of probability as the price you would pay to receive 1 if an event happens, 0 otherwise, with the clause that the opponent may swap the roles, i.e. become the gambler. This theory was further developed by Leonard Savage and Dempster
- The method is subjective, not arbitrary: it is the two thieves' rule one slices the cake, the other one chooses the preferred slice
- This subjective approach may be applied also to evaluate reliability
- Archaeologists may be interested** in applying this approach

Slicing a cake. As the second thief, which slice would you choose?

* B. De Finetti (1970) *Teoria delle probabilità*, Einaudi. English translation: (1974) *Theory of Probability*, Wiley & sons.
** As regards archaeological sources, an example of application is published in M. Figuera (2021) "A fuzzy approach to evaluate the attributions reliability in the archaeological sources" *Int. J. Digit. Libr.* 22, 289–296

A CRM-compliant inference model

- In a 2011 paper* Doerr et al. have proposed an inference model for reasoning. This model assumed that ”the belief value system is restricted to TRUE, FALSE, and UNKNOWN ”.
- In 2015 the CRM SIG has accepted as draft CRMinf**, a CRM compliant model forr inference
- CRMinf introduced the class I6 Belief Value as a subclass of E59 Primitive Value on a par with E60 Number, with possible values (at least) “True; False; Unknown” but also envisaging that “It may be expressed in terms of discrete logic, modal logic, probability, fuzziness or other representational system”
- A fuzzy case has been proposed*** for scientific results



Modelling reliability assignment

*M. Doerr, A. Krutsotaki & K. Boutsika (2011) “Factual argumentation—a core model for assertions making” ACM JOCCH, 3, 3, Article 8 (March 2011), 34 pages
** CRMinf: the Argumentation Model, version 0.7 (February 2015), http://old.cidoc-crm.org/CRM-Extensions/CRMinf/0.7/CRMinf-0.7.pdf
*** F. Niccolucci & S. Hermon (2017) “Expressing reliability with CIDOC CRM” *Int. J. Digit. Libr.* 18, 281–287

*Hanc marginis exiguitas non caperet**

- Reliability results from several factors related to the individual logical components of an argumentation. Examining each atomic component enables to identify (in a *subjective* way) the corresponding reliability as a "matter of fact", for example the error range of an experiment, the unclear nature of artefacts, the imprecise chronology and so on. Paradata help in assessing such values
- There are then computationally feasible methods to compose such individual reliability coefficients into a global coefficient applied to the result. This aggregation is used in many applicative fields, for example decision making based on imprecise or vague sources, pattern recognition and artificial intelligence
- A re-user may take the result as is, or combine it with an own evaluation of the proponent's arguments, and proceed so forth in a chain of trust
- But, unfortunately, my space-time remaining to explain the details is too small…

Time over!

clock clipart PNG Designed By cornecoba from https://pngtree.com/freepng/retro-alarm-clock-ringing_4185384.html?sol=downref&id=bef

*"The narrow margin cannot contain it". This is the famous note concerning Pierre de Fermat's demonstration of his *Last Theorem*, handwritten by Fermat using the margin of a copy of Diophantus' *Arithmetica*