Augmented Scope and Didactics for Initiation to the Theory of Sampling (TOS): Three Domains Behind Valid Data Quality

By Kim H. Esbensen[1]

DOI: 10.62178/sst.003.007

ABSTRACT

Professional comprehension and competence of sampling of particulate and aggregate materials, mixtures, and slurries depend on a minimum set of basic concepts, terms and definitions with derived procedures, equipment design, and practical skills as stipulated in the Theory of Sampling (TOS). Valid analytical data quality assurance involves acknowledgement of three interconnected domains along the lot-to-aliquot-to analysis-to-decision making pathway: i) sampling, ii) analysis, and iii) data analysis/modelling/decision making. This fundamental three-fold domain scope presented here for the first time allows establishment of a new axiomatic 'simplest possible, self-contained' introduction to representative sampling of heterogeneous materials under delineated conditions (TOS).

Introduction - Background

Complete data quality assurance necessitates acknowledgement of three interconnected domains along the full lot-to-aliquot-to analysis-to-decision making pathway: i) sampling, ii) analysis, and iii) data analysis/modelling/decision making. The data analysis/decision making domain is where use of analytical results takes place; this may range from simple data analysis/statistical treatment of analytical data, complex analytical signal calibration (multivariate calibration'), modeling, prediction, and validation to higher level considerations, for example as input to risk assessment. It is counterproductive to view any single, or just two of these three domains in isolation; professional overview is needed for all three.

Professional comprehension and competence of sampling of particulate and aggregate materials, mixtures, and slurries depend on a minimum set of basic concepts, terms and definitions with derived procedures, equipment and practical skills as stipulated in the Theory of Sampling (TOS).

This can be accomplished by comprehension of a set of focus points, constituting the simplest possible' initiation into the complex field of representative sampling:

1. The objective of sampling
2. Physical vs. statistical sampling – a critical distinction
3. All material lots of sampling interest are heterogeneous – the sampling bias
4. Practical sampling follows a universal 'lot-to-aliquot' pathway
5. "Everything" begins in the domain of sampling
6. Theory of Sampling (TOS) at a glance
7. Three necessary-and-sufficient domains behind valid data quality and use of analytical results
8. A new, augmented scope for the Theory of Sampling (TOS)
9. The representative analytical aliquot – the only valid creator of information
10. Error vs. uncertainty – clearing up monumental terminology confusions
11. Global sampling standard, DS 3077:2024 (3rd ed.)
12. Sampling – Historical timeline
13. Full professional competence

1 KHE Consulting, Copenhagen, Denmark

This contribution presents a complete three-domain background necessary for fully professional endeavours in the analytical and data analytical domains, amounting to a new, augmented didactic scope for initiation to the Theory of Sampling (TOS); see also [7].

Preamble

The body of concepts, definitions and terms necessary to master a professional competence regarding sampling is not trivial. Despite many claims to the contrary to found in the marketplace or online, for example 'sampling made simple' (a hook meant to lure customers to buy sampling equipment and solutions from OEMs on trust) trust us, we are sampling experts' a.o. But in science, technology, industry, and commerce there are insights and skills that can only be acquired at the expense of a minimum investment of intellectual work. The present new scope for the Theory of Sampling (TOS) intends to provide the holy grail of outreach from the sampling community, i.e., the 'simplest possible, self-contained' introduction to sampling of heterogeneous materials and processes under delineated conditions (TOS). This can best be accomplished by gradually developing a set of focus points enabling interested parties (at any level) as well as new practical samplers to acquire the theoretical overview and the practical skills necessary for representative sampling. Below an overview is presented of the theory of sampling as a system's framework introducing all elements and relationships necessary for full comprehension and practical competence. This article also contains an authoritative glossary of TOS definitions and terms, a curated list of introductions to TOS [1-10], and recommended further in-depth documentation and literature [11-23].

Where and how to start?

How to sample in a manner that will always guarantee a **representative sample**[2] from any lot, be it stationery, or a dynamic moving lot? Enter the **Theory and Practice of Sampling (TOS)**. It is essential to be able to communicate the complex issue of sampling of heterogeneous materials unambiguously, with absolute clarity. This requires a minimum, gradually developing set of definitions and terms.

A very first definition of sampling could be:

Sampling is the process of physical extraction and mass-reduction of a **composite sample** *counteracting* lot/material **heterogeneity** according to conditions as stipulated in the Theory of Sampling (TOS).

Credit: KHE Consulting; used with permission.

Figure 1: Despite lots having infinitely many, widely different manifestations, with infinitely many sizes and grainsize distributions – from TOS' point of view of they are all but heterogeneous materials with a smaller, intermediate or high degree of heterogeneity (never zero) - which allows them to be sampled with one universal sampling approach: composite sampling.

2 Terms in **boldface** are defined in the glossary (Appendix)

Focus #1 – the objective of sampling

The objective of sampling heterogeneous aggregate materials, mixtures, slurries a.o. is to produce a guaranteed **representative analytical aliquot**.

Primary samples are extracted from heterogeneous lots, sub-sampled (where needed in several stages) with the resulting aliquot analysed to estimate one or more properties of interest (quantification of 'the analyte') with which to characterise the lot adequately according to defined objectives e.g., data analytical, statistical, decision-making, business, or regulatory use of analytical results.

A **lot** is characterized by its size (from a miniscule to an extremely large mass) and its inherent material features. A **sample**, S, is a (very) small part of a larger **lot** (L), realized with a sampling rate *r* = sample weight / lot weight (for example 1:1,000 or 0.1%). While it is no practical challenge to extract a small portion from <u>any</u> lot of <u>any</u> size, using a practical mechanical tool e.g., a spatula, spoon, shovel, spade, corer, cross-stream sampler, mechanical or automated sampler, this is <u>not</u> sampling, only blind bulk mass-reduction. What is needed is r**epresentative sampling**, $\text{sampling}_{\text{REPR}}$.

A first set of fundamental terms and definitions includes:

Sampling (verb): a practical, mechanical process (or a virtual equivalent, see **PAT: Process Analytical Technologies**) extracting a physical sample (or intangible representation of a sample in the form of sensor spectroscopic information) from a lot. For the present initiation purpose 'sampling' denotes sampling from a physical lot made up of particulate, aggregate material.

Sample (noun): A portion of a larger lot produced by a documentable representative sampling process under specified conditions.

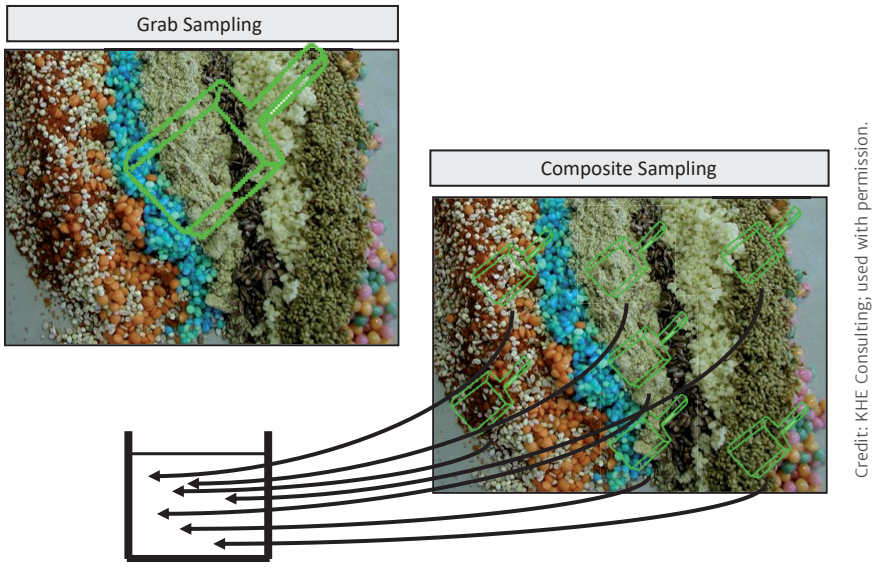
Specimen (noun): Portions extracted from a lot that <u>cannot</u> be documented to result from a representative sampling process are termed **specimens**. Specimens are *worthless* lumps of material because they do not carry valid information regarding their relationship with the original lot; specimens have no useful provenance.

Increment: fundamental unit of practical sampling, defined by a specific mass or volume extracted by a specified procedure using a specific sampling tool.

Grab sampling: process of extracting a *singular* increment. For heterogeneous materials, grab sampling <u>cannot</u> ensure representativity [1–7, 9,11,12].

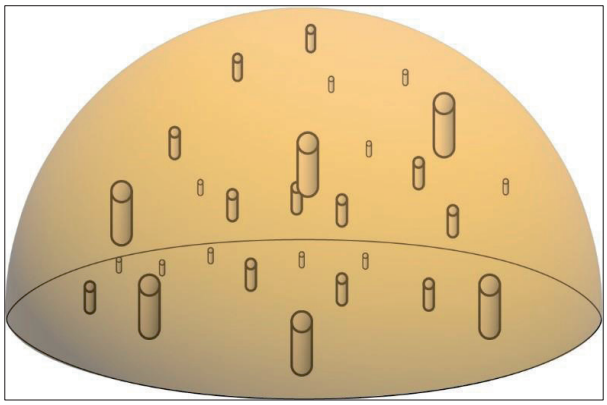
Composite sampling: process resulting in a compound sample made by aggregating a set of Q increments subject to the **Fundamental Sampling Principle (FSP)**. Q can be focused to make sampling *fit-for-purpose*.

Representativity (noun): prime objective of all proper sampling processes. Representativity refers to intrinsic material features, e.g., composition, grain size distribution, physical properties (e.g., intrinsic moisture). The representativity status of an individual sample cannot be defined nor ascertained if removed from the context of its generating sampling-and-analysis pathway. The attribute 'representative' can only be accorded a sampling process in compliance with all relevant demands specified by TOS.



Credit: KHE Consulting; used with permission.

Figure 2: Grab sampling vs. **composite sampling**



Credit: KHE Consulting; used with permission.

Figure 3: Fundamental Sampling Principle (FSP)

Fundamental Sampling Principle (FSP): mandated principle for sampling processes ensuring all increments an identical, non-zero extraction probability while covering the full material lot (volume/mass). Sampling of a lot in which certain areas, volumes, parts are not physically accessible cannot ensure representativity.

Process Analytical Technologies (PAT): Sampling performed using a suitable sensor technology allowing acquisition of spectral characterization of a delineated volume of a ducted flux of matter (a 'process sample') by way of an appropriate sampling interface [22]. In the realm of process sampling, PAT is aka 'sensor sampling'.

Focus #2: physical vs. statistical sampling – a critical distinction

It is essential to distinguish between:

Sampling (*from a physical lot*), sampling$_{TOS}$

versus

Statistical sampling (*from a population of units*), sampling$_{STAT}$

Long lasting terminology ambiguity between sampling$_{TOS}$ and sampling$_{STAT}$ has caused significant confusion across and between many scientific disciplines and technological/ industrial application fields, see [8,9] for in-depth treatment. This distraction can first be fully resolved after all terms and definitions pertaining to TOS have been properly comprehended. For the present initiation purpose sampling shall always denote sampling$_{TOS}$ unless otherwise specified.

Focus #3: All material lots are heterogeneous – the sampling bias

Heterogeneity is one of two key influential factors that must be counteracted by all practical sampling processes, lest these will be compromised by a fatal **sampling bias**. A sampling bias will also be incurred by an incorrect[3] sampling procedure, e.g., **grab sampling**. A sampling bias is fundamentally different from an analytical bias. While the latter can be subjected to a conventional analytical laboratory bias-correction, the sampling bias cannot be corrected by any means (data analytical, statistical, other). Instead, TOS stipulates that all sampling operations must be designed to eliminate the so-called Incorrect Sampling Errors (ISE), which, when unmitigated, are unavoidable hidden sampling bias generators, see Focus #6 and [1-10].

Focus #4: Universal sampling$_{TOS}$ 'lot-to-aliquot' pathway

"What is the meaning of analysing, with ultimate analytical accuracy and precision, the concentration of an **aliquot** that represents only a *miniscule* $1/10^3$ - $1/10^9$ mass-reduced fraction of the original lot mass - if the process by which it is obtained is compromised, not representative?" None, there is no meaning! The resulting analytical results carry no reliable information about the original lot. Non-representative samples, sub-samples and aliquots unavoidably lead to non-representative analytical results, *regardless* of the quality of analysis. All costs incurred in sampling 'from-lot-to-aliquot-to-analysis' are therefore lost and cannot ever be recouped. Therefore, focus must be exclusively on *how* to guarantee extraction of representative primary samples, followed (equally important) by a number of representative mass-reducing sub-sampling stages [6,12] until having produced the representative aliquot – to be delivered to the domain of analysis.

Focus #5: "Everything" begins in the domain of sampling

It is necessary to step back from the traditional preoccupation with analytical accuracy, analytical precision, which resides in the domain of analysis to the 'before analysis' domain. This is the sampling domain (*verb*) – not the sample (*noun*) domain. The latter designation would imply that 'samples' are already existing, ready to be selected and extracted *in toto*. However, the reality concerning how to sample heterogeneous materials, lots and processes is very different.

3 Incorrect vs. correct sampling errors a.o., see Focus #6 and references [1-10]

Representative sampling must follow the universal pathway 'from-lot-to-aliquot' that demands appropriate scientific and technological competence, enter the **Theory of Sampling (TOS)**. Everything is critically dependent upon the competence and skills needed for the extraction of a representative primary sample from the lot (regardless of the world's very many, very different lot manifestations).

Presciently TOS makes provisions allowing for a *unified* approach, Fig. 4. By focusing on the *common* characteristics of material heterogeneity [1-7,11], it is possible to address sampling of all types of material using a singular *generic* sampling pathway, governed by the principles and unit operations in the Theory of Sampling (TOS). This is perhaps the most enabling aspect of TOS: Since all materials are heterogeneous (it is only a matter to which degree large, intermediate, small, but never zero), TOS' generic sampling pathway is universally applicable to *all* types of material, appearing with *any* lot size, under *all* specified conditions. As one example, TOS applies with equal force for *any* primary size lot, but *also* in all the world's analytical laboratories for *all* menial sub-sampling operations needed here; it is only the *scale* differs (Principle of Sampling Scale-Invariance).

Focus #6: Theory of Sampling (TOS) - everything at a glance

The ultimate purpose of the use - and the scientific, regulatory, technological, or economic value of analytical results are all dependent on the imperative demand for all analytical aliquots to be *representative* of the *original heterogeneous lot/material* in question.

The sampling-to-analysis pathway is always a **multi-step process**, starting with primary sampling of the lot, ending with analysis of the aliquot (or test portion). This process always involves significant mass reductions with typical sampling rates (m/m) $1{:}10^3$ to $1{:}10^9$: lot (~tons) → primary samples (~kilograms) → secondary samples (~grams) → analytical aliquot (grams to micro grams) → analytical measurement. This is all required to be conducted in such a way that the final analytical result *represents* the salient properties of the original lot in an objectively documentable, fully reliable way [7].

The **Theory of Sampling (TOS)** is the only complete science-backed framework defining its role to be the guarantor of both **sampling accuracy** w.r.t., the original lot, and of **sampling precision** w.r.t., *reproducibility of the analytical aliquot*. Until physical delivery of the aliquot for analysis, this responsibility exclusively resides in the sampling domain. While the specific nature of 'the analyte' may imply various constraints (the analyte may for example be a physical characteristic of the sampled material, e.g., compression strength of concrete), this has no principal impact on how to conduct the *preceding* multi-stage sampling and sub-sampling process(es), which all takes place before analysis.

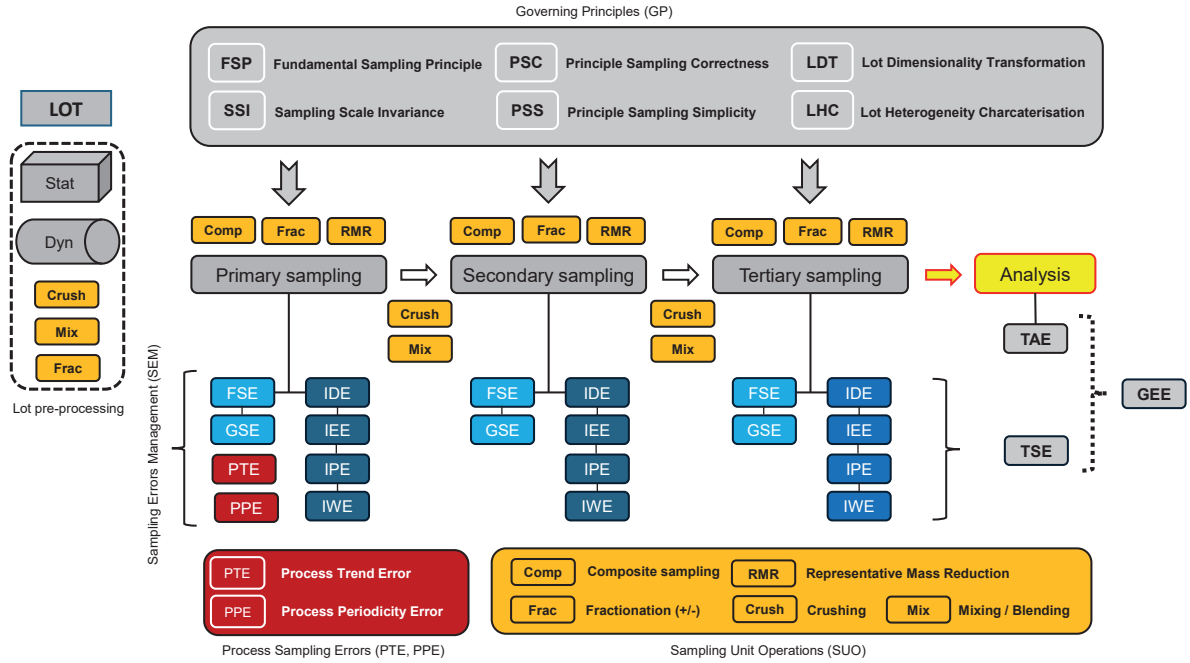


Figure 4: Theory of Sampling (TOS) Principal system's framework DS 3077: 2024 (3rd ed) [7].

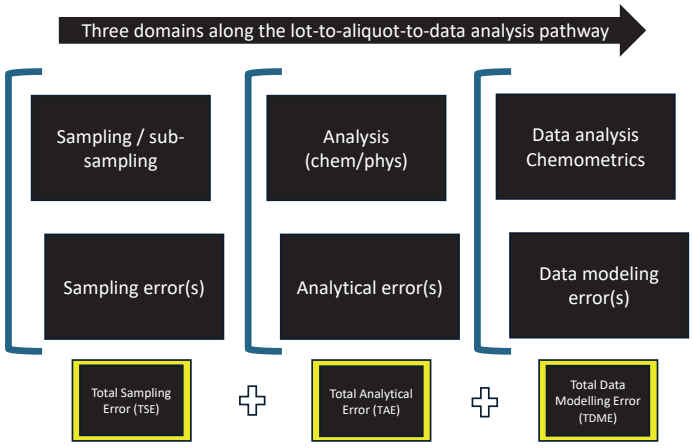
Focus #7: Three principal domains – necessary for optimal data quality

Complete quality assurance of analytical data necessitates acknowledgement of three fundamental *domains* along the full 'lot-to-aliquot-to analysis-to-decision making' pathway, which are:

1. The 'before analysis' domain (the sampling and sub-sampling domain)
2. The domain of analysis *s.s.*
3. The DSR domain (Data analysis; Statistics; Risk management; decision making).

The DSR domain is where the use of *analytical results* takes place, ranging from simple data analytical/statistical data treatment, via complex analytical signal calibration (chemometric multivariate calibration), modeling, prediction and validation), to higher level decision-making considerations, for example, as input to risk assessment [17].

It is counterproductive to view any single, or just two of these three domains in isolation. The feature 'data quality' has all too long been viewed as only related to analytical uncertainty, with seriously detrimental effects since leaving out all sampling uncertainty – and sometimes also ignoring errors and uncertainty effects associated with DSR operations on analytical data. Reliable use of quantitative 'data' must be based on acknowledgement of all three interconnected domains making up the full 'lot-to-aliquot-to-analysis-to-DSR' pathway.



Credit: KHE Consulting; used with permission.

Figure 5: Three successive domains are involved to cover the full 'lot-to-analysis-to-DSR' pathway.

Each domain is characterized by potential errors (TSE), (TAE), (TDSRE), which give rise to *uncertainty* effects (Focus #10).
It is the responsibility of specific domain expertise to minimize, or eliminate (where possible), all domain errors and effects (uncertainties). If no counteracting measures are taken, the 'before analysis' sampling domain will very nearly always dominate the total uncertainty budget: $MU_{total} = MU_{sampling} + MU_{analysis} + MU_{DSR}$.

In this context, from a logical, scientific and economic point of view all efforts and costs spent on analysis of what in reality are **specimens** is futile. The actions taking place in subsequent domains, i.e., data analysis/data modelling/statistical or critical decision-making domains, or regarding **Risk Management (RM)** will be seriously affected, with data quality unavoidably compromised. There is a ticking time bomb embedded in data for which the demands for *data quality* have not been adequately defined *before* sampling and analysis. Non-representative sampling is (like) the original sin: sampling error effects are passed on to the subsequent domains in the lot-to-analysis-to-DSR pathway in the form of hidden, uncontrollable additions to the total Measurement Uncertainty (MU_{total}) which will always be inflated to a degree larger than necessary. But there is no way to estimate the magnitude of such excess uncertainty incurred - and data quality issues cannot be rectified in any way in the post-analysis domain (sampling domain *corrections* are not possible). Data quality originates in, and must be optimized, starting with the sampling domain.

Perhaps the most prominent example of the need for complete domain comprehension concerns **Process Analytical Technologies (PAT)**, an approach for process sampling using appropriate sensor technologies to acquire spectral information from a delineated target volume in front of a sensor followed by powerful *multivariate calibration* (chemometrics) a.o.[3] But the PAT approach is overwhelmingly only concerned with the challenging analytical aspects together with the subsequent domain. PAT is an essential element of process monitoring and control, which are part of the DSR domain, all the while leaving the sampling domain overlooked. This has serious, often fatal consequences, as the delineated analytical volume is very nearly always just a minute fraction of the cross-section of the flowing stream resulting in a serious **sampling bias**. Very many current PAT solutions are flawed in this respect; full details can be found in [9,22].

3 This discipline is treated in full in SST#4, which is devoted to the theme Process Analytical Technology (PAT) vs. Process sampling.

Focus #8: A new, augmented scope for the Theory of Sampling (TOS)

From this three-domain context emerges a new paradigm regarding accuracy and precision in relation to sampling and analysis of heterogeneous lots/materials. It is critical that the full lot-to-aliquot is front and center:

Analytical accuracy and *analytical* precision, $MU_{analysis}$, *only* characterises the specific analytical method employed, thereby *missing* the dominating MU contributions stemming from sampling, $MU_{sampling}$. Users of analytical results cannot make valid and reliable decisions without information about *both* uncertainty contributions and should therefore always be supplied with information regarding MU_{total}, the *effective* total uncertainty. To this must be added the MU_{DSR} as appropriate to the situation.

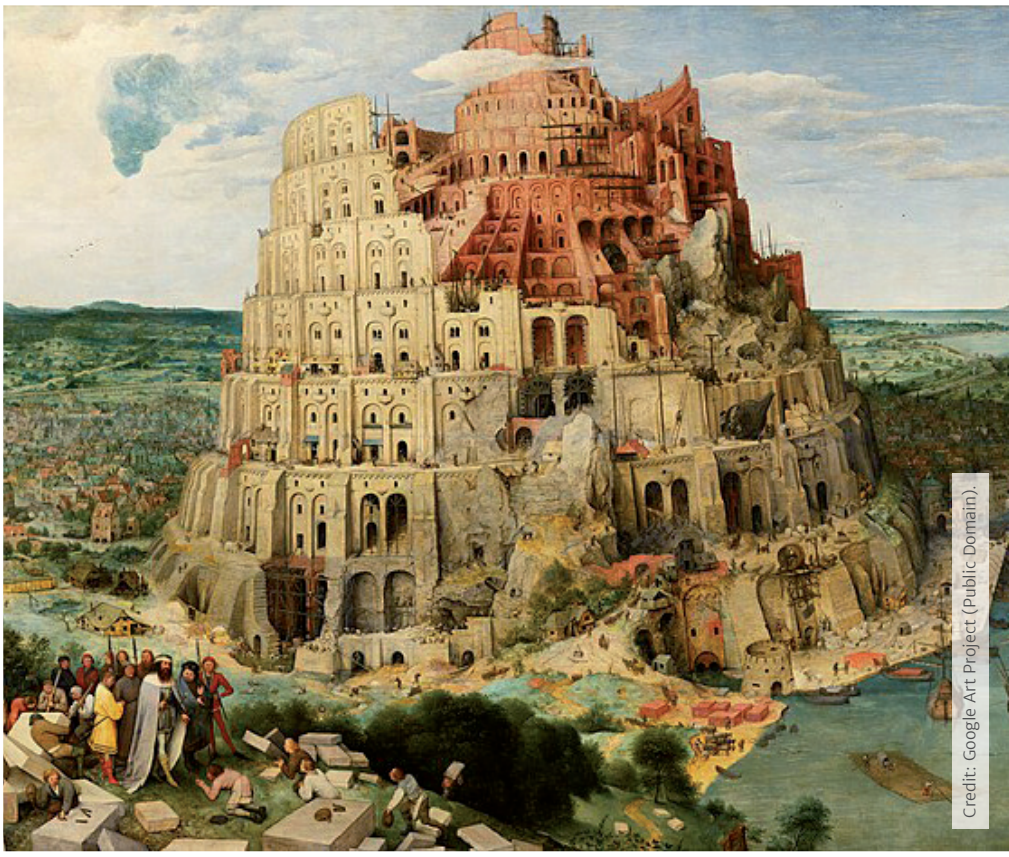


Figure 6: "Tower of Babel" by Pieter Bruegel the Elder (1563)

Focus #9: The analytical aliquot – only valid *creator* of relevant information

In the three-domain context, the analytical aliquot is the *physical manifestation* of translating from the domain of sampling to the domain of analysis. The *quality* of an aliquot is not related to the aliquot itself (which may come as a surprise to many) but is *exclusively* a function of the *sampling process* by which is was produced. In this context, the aliquot (strictly speaking, analysis of the aliquot) is the only valid creator of information about the original lot (stationary or moving). Also, it is not possible to ascertain whether a specific sample or aliquot is representative, or not, from any considerations only relating to the sample/sub-sample/ aliquot itself. Therefore, focus shall *exclusively* be on the *sampling process*, which must be fully TOS-compliant for the aliquot to be *documentable* as representative [1-10]. Also, there is no declination of the attribute 'representative'. A sampling process, or a resulting sample either is or is not representative.

Focus #10: Error vs uncertainty

There unfortunately exist scores of fundamentally *different* definitions of the concepts and terms *error* vs. *uncertainty* in various scientific disciplines, educational traditions and related literatures, often severely at odds with one another. The relationship between these is well likened to the Tower of Babel as has been extensively presented and debated in [8,9].

There is only one scientific way out of this quagmire: Clear, unambiguous definitions are mandatory as part of all outreach and educational endeavours. It is essential that analytical results are always reported *together* with a realistic total estimate of the associated Measurement Uncertainty, $MU_{total} = MU_{sampling} + MU_{analysis} + MU_{DSR}$. Currently, the uncertainty contribution from sampling is overlooked all too often with highly detrimental consequences because $MU_{sampling}$ can be 10-25-50 times *larger* than $MU_{analysis}$ depending on the level of sampling errors effects incurred by ignoring, or not being aware of, the critical adverse heterogeneity influence on the sampling process. An in-depth discussion of this status quo, with a critical focus for the DSRM domain, is presented in [9].

Focus #11: Global sampling standard DS 3077:2024 (3rd ed)

2024 saw publication of the 3rd revision of the standard DS 3077:2024 "Representative Sampling – Horisontal Standard" [7]. A succinct summary is as follows:

"The theory of sampling is a generic, matrix-independent framework for representative sampling of all types of aggregate and mixture materials (solid, slurries) in all grain-size brackets (from broken ores to powders). The universal sampling principles can be applied uniformly to all types of materials, and lots composed by aggregate particular matter and slurries (gasses and liquids are not covered by this document). This document describes a generic sampling process in sufficient detail and covers all elements necessary and sufficient for the stated objective enabling documentation of sampling representativity under the specified conditions for the sampling process employed. This document is based on the theory of sampling (TOS), constituting a complete competence basis for representative sampling, and ensuring appropriate levels of accuracy and precision for both primary sampling as well as for all subsampling procedures and mass-reduction systems at the subsequent laboratory stages before analysis..." [7, p.8]

This universal standard for representative sampling is aimed at all individuals with vested interest and/or responsibility for sampling (technical and supervisory personal, managers, stakeholders, companies, corporations, organisations and other relevant *legal persons*). The present compact initiation can be viewed as the 'simplest possible, self-contained introduction' for all of these agents (including legal and accounting departments).

All educational introductions are complemented by the following *call-to-action* [4,9], Fig. 7.

Credit: KHE Consulting; used with permission.

Figure 7: Proposal for a universal creed for responsible representative sampling, the Theory of Sampling (TOS) [7]

Henceforth, it is proposed to include these two statements wherever relevant in every commercial and trade contract and in any other guidance documentation that is based on, or includes sampling. This will reduce a substantial proportion of legal disputes stemming from isolated comparison of analytical results without the necessary three-domain recognition. In-depth treatment of this "assay exchange" issue can be found in [10], where it is shown that most, if not all such disputes simply reflect a lack of proper TOS understanding and competence. For the present initiation to TOS, a Glossary can be found in the Appendix. A broader selection of definitions and terms can be found in [7].

Focus #12: A historical timeline

Figure 8 presents a brief historical timeline of ca. 150 years of development of market needs/demands in societal sectors where early attempts at 'sampling' gradually emerged as first technological solutions. This development was fragmented and scattered until 1950, the year of a publication containing the first recorded vestiges of what came to be the Theory of Sampling (TOS) later on (in 1975). It took 25 more years until organised activities saw the light of day (2003) in the form of the 1st World Conference on Sampling and Blending (WCSB1), the founding of the International Pierre Gy Sampling Association (IPGSA) in 2017, and the start and continuation of dedicated publication activities: TOS-Forum, Sampling Science and Technology (2017-present). Two accounts of this development can be found in [22,23]. The International Pierre Gy Sampling Society (IPGSA) launched a fully updated homepage in 2025.

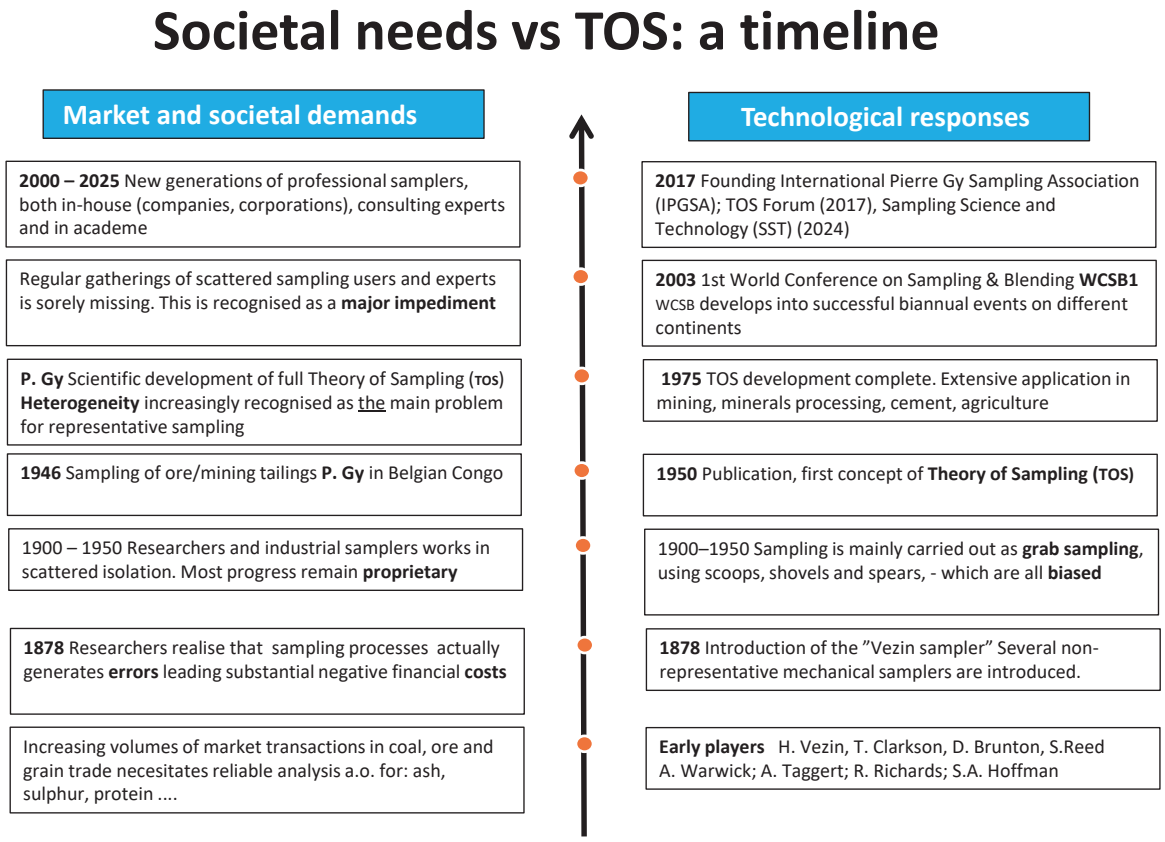


Figure 8: Timeline of development of sampling (practice and theory) as solutions to developing market and societal needs/demands in the last ~150 years.

Focus #13: Full professional competence

References [1-9] present a curated collection of general introductory literature on the Theory and Practice of Sampling (TOS) at entry level, [10-22] adding to this curriculum with accounts of applied TOS from selected scientific, technological and industrial application fields.

For guided competence building, TOS recommends the following reading order:

- Tier 1 (introductory): [1,2,7,8,6,12,11,21]
- Tier 2 (more advanced learning): [9, 4,3,5,17] [10,13-16,18-20, 22]

Selected contributions on the history of the theory and practice of sampling can be found in [23-24].

References

1. Esbensen, K.H. & Wagner, C. (2014). Why we need the Theory of Sampling, *The Analytical Scientist*, 21 (Oct.), pp. 30–38.

2. Esbensen, K.H. (2024). *Introduction to the Theory and Practice of Sampling*. p. 328, SIX-P Publishing, Wedel, DE.

3. Pitard, F.F. (2019). *Theory of Sampling and Sampling Practice*, 3rd ed., CRC Press, Boca Raton, Florida. p. 665. ISBN 978-1-1138-47648-6

4. Gy, P. (1998). *Sampling for Analytical Purposes*, Elsevier, Amsterdam, NE., p. 148. ISBN 0-471-97956-2

5. Lyman, G. J. (2020). *Theory and Practice of Particulate Sampling Theory – an Engineering Approach*. Materials Sampling & Consulting, Brisbane, Queensland, AUS. Geoff Lyman - Owner - Materials Sampling & Consulting Pty Ltd | LinkedIn

6. Petersen, L., Dahl, C., & Esbensen, K.H. (2004). Representative mass reduction in sampling – a critical survey of techniques and hardware. in Special Issue: 50 years of Pierre Gy's Theory of Sampling, In K.H. Esbensen & P. Minkkinen (Eds.), Proceedings of the First World Conference on Sampling and Blending (WCSB1), Esbjerg, DK., *Chemometrics and Intelligent Laboratory Systems*, 74(1), pp. 95–114.

7. DS 3077 (2024). *Representative Sampling— Horizontal Standard* (3rd ed.). Danish Standards, Fonden Dansk Standard, Nordhavn, Copenhagen, DK. https://webshop.ds.dk/en/standard/M374267/ds-3077-2024

8. Esbensen, K.H. & Pitard, F.F. (2022). Errors vs uncertainty vs measurement uncertainty. *Spectroscopy Europe/World*, 34(6/7).

9. Esbensen, K.H. (forthcoming 2025). Data Quality: Importance of the 'before analysis' domain (Theory of Sampling, TOS). *Journal of Chemometrics*. Special Issue "Tribute to Paul Geladi".

10. Esbensen K.H. & Vogel A.D. (2023). Unsubstantiated complacency re. the "assay exchange" paradigm: sampling uncertainties with hidden economic consequences. *ALCHEMIST 110* (July), pp. 14–21.

11a. Esbensen, K.H., Paoletti, C. & Theix, N. (2015). (Eds.) Journal AOAC International, Special Guest Editor Section (SGE): *Sampling for Food and Feed Materials*. pp. 249–320.

11b. Esbensen, K. H. (2015). Journal of AOAC International (JAOAC) Special Issue on Representative Sampling for Food and Feed Materials: A Critical Need for Food/Feed Safety. *TOS forum*, 4, p. 12. https://doi.org/10.1255/tosf.39

12. Petersen, L., Minkkinen, P. & Esbensen, K.H. (2005). Representative Sampling for reliable data analysis: Theory of Sampling, *Chemometrics and intelligent laboratory systems*, 77(1–2), pp. 261–277.

13. Esbensen, K. H., Romañach, R. J. & Román-Ospino, A. D. (2018). Theory of Sampling (TOS): A Necessary and Sufficient Guarantee for Reliable Multivariate Data Analysis, Chap. 4 in Ferreira, A. P., Menezes, J. C.; Tobyn, M. (Eds.), *Pharmaceutical Manufacturing in Multivariate Analysis in the Pharmaceutical Industry*, pp. 53–91. Academic Press, ISBN 978-0-12-811065-2

14. Esbensen, K.H., Paoletti, C. & Minkkinen, P. (2012). Representative sampling of large kernel lots – I. Theory of Sampling and variographic analysis, *Trends in Analytical Chemistry (TrAC)*, 32, pp.154–165, doi:10.1016/j. trac.2011.09.008

15. Minkkinen, P., Esbensen, K.H., & Paoletti, C. (2012). Representative sampling of large kernel lots – II. Application to soybean sampling for GMO control. *Trends in Analytical Chemistry (TrAC)*, 32, pp. 166–178, doi:10.1016/j.trac.2011.12.001 170.

16. Esbensen, K.H., Paoletti, C. & Minkkinen, P. (2012). Representative sampling of large kernel lots – III. General Considerations on sampling heterogeneous foods, *Trends in Analytical Chemistry (TrAC)*, 32, pp. 179–184, doi:10.1016/j.trac.2011.12.002

17. Paoletti, C. & Esbensen, K.H. (2022). Framing TOS in risk assessment: an outreach perspective for the future, in *Proceedings of WCSB10. TOS Forum Issue 11*, pp. 419–423

18. Esbensen, K.H. & Abu-Khalaf N. (2022). Before reliable near infrared spectroscopic analysis - the critical sampling proviso. Part 1: Generalised theory of sampling, *Journal of Near Infrared Spectroscopy* 30(6), pp. 291–310, doi:10.1177/09670335221124612

19. Esbensen, K.H. & Abu-Khalaf N. (2022). Before reliable near infrared spectroscopic analysis- the critical sampling proviso. Part 2: Particular requirements for near infrared spectroscopy, *Journal of Near Infrared Spectroscopy.* 30(6), pp. 311–321, doi:10.1177/09670335221124611

20. Esbensen, K. H. & Romañach, R. J., (2020). Sampling for NIR-analysis. Chap. 19, in *Handbook of NearInfrared Analysis*, 4th Ed. Ciurczak, E.W., Igne,B., Workman,J., Burns, D.A. (Eds.) CRC Press, Taylor & Francis Group. ISBN: 978-1-138-57648- 3

21. Abel Arkenbout, Stéphane Brochot, Trevor Bruce, Pedro Carrasco, Pablo Carrasco, Philippe Davin, Quentin Dehaine, Oscar Dominguez, Simon C. Dominy, Kim H. Esbensen, Dominique François-Bongarçon, Melissa Gouws, Ralph J. Holmes, Li Huachang, Eduardo Jara, Martin Lischka, Geoff Lyman, Zhu Mingwei, Pentti Minkkinen, Richard C.A. Minnitt, Claudia Paoletti, Christopher Robben, Rodolfo J. Romañach, Elke Thisted & D. Aldwin Vogel (2021). *Economic arguments for representative sampling. Spectroscopy Europe/World, 33*: 7, https://wcsb10.com/media/ogzbw012/wcsb10-economic-arguments-dec-2022-compressed.pdf

22. Esbensen. K.H. & Sivalingam, V. (2020). The sampling interface—A critical Theory of Sampling success factor in process sampling and PAT. *Proceedings of WCSB10: TOS Forum Issue 11*, pp. 115–129

23. Esbensen, K.H. (2004). "50 years of Pierre Gys 'Theory of Sampling'– a tribute", in Special Issue Proceedings: *First World Conference on Sampling and Blending (WCSB1)*. K.H. Esbensen & P. Minkkinen (Eds). *Chemometrics and Intelligent Laboratory Systems*, vol. 74, Issue 1, p. 3-6

24. Minnitt, R.C.A. & Esbensen, K.H. (2017). Pierre Gy's development of the Theory of Sampling: a retrospective summary with a didactic tutorial on quantitative sampling of one-dimensional lots. *TOS Forum* 7, pp. 7-19. doi: 10.1255/tosf.96

Appendix: Glossary

Lot: specified target material to be subjected to a specified sampling procedure. The term lot refers both to the material as well as to size (volume/mass), physical characteristics and geometric form. Lots are distinguished as *stationary* or *dynamic* (moving lots). A dynamic lot is a material stream where sampling is carried out at a fixed location. For both stationary and dynamic lots, sampling procedures must address the entire lot volume guided by the Fundamental Sampling Principle (FSP).

Mass-reduction: divisionary process leading to one or more *sub-samples* (portions) [m/m] of a larger lot/sample/sub-sample, with the objective of being representative of the original lot.

Sampling: (sampling procedure; sampling process): grab sampling or composite sampling.

Increment: fundamental unit of practical sampling, defined by a specific mass or volume extracted by a specified sampling tool.

Grab sampling: process of extracting a singular increment. Grab sampling cannot ensure representativity for heterogeneous materials.

Composite sampling: process leading to a compound sample (composite sample) made by aggregating a set of Q increments subject to the Fundamental Sampling Principle (FSP). Q can be set to make sampling fit-for-purpose according to a specific criterion.

Sample: extracted portion of a lot that can be documented to be a result of a representative sampling procedure (non-representatively extracted portions of a lot are termed specimens).

Sampling accuracy: Closeness of the analytical result of an aliquot w.r.t., to the true concentration of a lot. NB. Sampling accuracy always includes the analytical imprecision, since analysis is always based on an analytical aliquot, which is the end result of a complete 'lot-to-aliquot' sampling pathway. Therefore: "sampling accuracy" = "sampling + analytical accuracy".

Sampling precision: Variance of the series of analytical determinations in a Replication Experiment (RE). NB. Sampling precision always includes the analytical precision, since all analysis is always based on an analytical aliquot, which is the end result of a complete 'lot-to-aliquot' sampling pathway. Therefore, "sampling precision" = "sampling + analysis precision".

Analytical precision: Variance of repeated analytical determinations made on one-and-the-same aliquot. Analytical precision is *only* a characteristic of the analytical method.

Analytical accuracy: Deviation between the average of a series of repeated analytical determinations on one-and-the-same aliquot and the true average concentration of a lot. Analytical accuracy is *only* a characteristic of the analytical method.

Specimen: portion of a larger mass/volume extracted by a non-representative sampling process.

Fundamental Sampling Principle (FSP): principle governing a sampling process ensuring all increments an identical, non-zero extraction probability while covering the entire material lot (volume/mass). Sampling of a lot in which certain areas, volumes, or parts are not physically accessible cannot ensure representativity.

Fit-for-purpose representativeness: characteristic of a sampling process in which the Total Sampling Error (TSE) has been reduced to below a predefined threshold level.

Sampling bias: Difference between true lot concentration and grab sample or composite sample concentration determination (or average of replicate sample concentration determinations), whether sampled representatively or not.

Compositional heterogeneity (CH): compositional differences between *individual* fundamental units of a material (grains, particles, fragments). CH is an intrinsic characteristic of the target material to be sampled.

Distributional heterogeneity (DH): compositional differences between *groups* of fundamental units of a target material. Groups of units manifest themselves as practical increments used in sampling. DH is an expression of the spatial heterogeneity of a material to be sampled.

Grain-size heterogeneity (GH): compositional difference due to assemblages of units with different grain-size.

Lot heterogeneity: CH + DH + GH

Homogeneity: an assemblage of material units with *identical* unit size, composition, surface characteristics a.o. N.B. there are practically no homogenous materials in the realm of technology, industry, commerce a.o. of interest for sampling. All materials from these realms are in practice heterogeneous.

Representativity: prime objective of all sampling processes. Representativity refers to intrinsic material features, e.g., composition, grain size distribution, physical properties (e.g., intrinsic moisture). The representativity status of an individual sample cannot be defined, nor ascertained in isolation. i.e., if removed from the context of its full sampling-and-analysis pathway. The characteristic representative can only be accorded a sampling process in compliance with all relevant demands specified by TOS [1-10]. NB: For full mathematical-statistical definition see [3,4,5].

Theory of Sampling (TOS) (Theory and Practice of Sampling): necessary-and-sufficient framework of governing principles (GP), sampling unit operations (SUO), sampling error management rules (SEM) together with derived practices and skills needed to overcome adverse effects of material heterogeneity and non-representative sampling procedures.

Aliquot (analytical aliquot): ultimate sub-sample extracted in a 'lot-to-aliquot' pathway intended for analysis – or a *virtual* sample, e.g., a delineated volume of a stream of matter interacting with a spectroscopic analytical instrument (in the realm of Process Analytical Technologies, PAT).

Measurement Uncertainty (MU): (metrological term): MU expresses the variability interval of values attributed to a quantity measured. MU is the effect of a particular error, e.g. a sampling error, an analytical error or a data modelling error a.o. – or of combined effects (see MU_{total}).

- $MU_{sampling}$ reflects the variability stemming from sampling uncertainty
- $MU_{analysis}$ reflects the variability stemming from analytical uncertainty
- MU_{total} is the effective uncertainty stemming from both sampling and analysis
- $MU_{total} = MU_{sampling} + MU_{analysis} + MU_{DSR}$

Precision: Statistical variance $(STD)^2$. In practical sampling and analysis contexts 'precision' is a measure of imprecision.

Replication Experiment (RE): Replication of a series of independent complete 'lot-to-aliquot' analytical determinations, made under identical conditions. The number of replications is termed Q.

Total Sampling Error (TSE): TSE is causing the combined uncertainty effects resulting from material extraction along the full sampling pathway from-lot-to-aliquot.

Total Analytical Error (TAE): TAE is causing the combined uncertainty effects specifically resulting from analysis of the aliquot only.

Total Data modelling, Statistics, Risk management Error (TDSRE): TDSRE is causing the combined uncertainty effects resulting from post-analysis data treatment (DSR).

Stakeholder: legal person (company, corporation, agency or individual) with a vested interest or concern. Process Analytical Technologies (PAT): In the current process industry arena, analytical endeavors are increasingly sought to be served by the Process Analytical Technology (PAT) framework, offering a plentitude of on-line, mostly spectroscopic analytics: UV-VIS, NIR, RAMAN, NMR, 'acoustic chemometrics' a.o. See [12] as an introduction.

Sampling manager: legal person to whom responsibility is given for all actions related to sampling in a specified scientific, technological, industrial, business or other context.

Legal person: a legal person is any person or other legal entity that can do the things a human person is usually able to do in law – such as enter into contracts, commit to specified obligations and responsibilities.

TOS forum

Starting 2024 *Sampling Science and Technology* (SST) is a direct continuation of *TOS forum*, which was published by IMPublications in the decade 2013-2023.

The complete archive can be found here:

sst-magazine.info/tos-forum