



Metrology for data in life sciences, healthcare and pharmaceutical manufacturing: Case studies from the National Physical Laboratory

Paul M. Duncan¹, Nadia A. S. Smith¹, Marina Romanchikova¹

¹ Data Science Department, National Physical Laboratory, United Kingdom

ABSTRACT

Data metrology, i.e., the evaluation of data quality and its fitness-for-purpose, is an inherent part of many disciplines including physics and engineering. In other domains such as life sciences, medicine, and pharmaceutical manufacturing these tools are often added as an afterthought, if considered at all. The use of data-driven decision making and the advent of machine learning in these industries has created an urgent demand for harmonised, high-quality, content rich, and instantly available datasets across domains. The Findable, Accessible, Interoperable, Reproducible principles are designed to improve overall quality of research data. However, these principles alone do not guarantee that data is fit-for-purpose. Issues such as missing data and metadata, insufficient knowledge of measurement conditions or data provenance are well known and can be aided by applying metrological concepts to data preparation to increase confidence. This work conducted by National Physical Laboratory Data Science team showcases life sciences and healthcare projects where data metrology has been used to improve data quality.

Section: RESEARCH PAPER

Keywords: NMI; metrology; digital pathology; medicines manufacturing; metadata standards; data quality; ontologies; FAIR principles

Citation: Paul M. Duncan, Nadia A. S. Smith, Marina Romanchikova, Metrology for data in life sciences, healthcare and pharmaceutical manufacturing: Case studies from the National Physical Laboratory, Acta IMEKO, vol. 12, no. 1, article 10, March 2023, identifier: IMEKO-ACTA-12 (2023)-01-10

Section Editor: Daniel Hutzschenreuter, PTB, Germany

Received November 18, 2022; **In final form** February 20, 2023; **Published** March 2023

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the UK Government Department for Business, Energy & Industrial Strategy through the UK's National Measurement System.

Corresponding author: Paul M. Duncan, e-mail: paul.duncan@npl.co.uk

1. INTRODUCTION

Decision making in research, industry and healthcare is underpinned by the quality of data including its provenance, timeliness, reliability, and other aspects. Ascertaining the data quality using metrological principles of traceability, calibration and uncertainty can be described as *data metrology*. While certain disciplines such as radiation dosimetry or coordinate measurement of industrial components have long incorporated metrological tools such as calibration and traceability into their workflows, others such as laboratory medicine or pharmaceutical manufacturing are relatively new adopters who benefit profoundly from the European metrology networks for Traceability in Laboratory Medicine and Advanced Manufacturing.

Technological advancements in medicine and pharmaceutical manufacturing have been traditionally focused on advances in

drug discovery, experimental procedures, and manufacture. Medicines and treatments are becoming more expensive to produce, as pricing models drive down profit margins compounded with patents expiry [1]. Therefore, a greater emphasis is being placed on maximising the efficiency of medicines development and manufacture.

For the quality and repeatability of processes, most pharmaceutical firms operate at high variation levels in terms of accurately manufacturing materials. These variations, at levels between 3σ and 4σ , are estimated to cost ~\$20 bn annually through waste and inefficiency [2]. Therefore, companies are increasingly moving to developing controlled and flexible processes to offer digital health solutions for their customers.

The National Physical Laboratory (NPL) has set out to aid digitalisation in healthcare by focusing on the development of data metrology for life sciences, medicines and pharmaceutical manufacturing. Data metrology refers to the uncertainty present

in the data generated in each of these areas, from the quality of measurements accompanying the manufacturing to the quality of the data used for decision making processes.

This paper describes the similarities and differences between data metrology challenges addressed by NPL in the context of several cross-disciplinary projects with the goal of helping users to identify their data metrology needs and delivering confidence in the effective use of data.

2. DATA METROLOGY PROJECTS

The NPL Data Science team has been involved in multiple data metrology projects including life sciences, healthcare, and medicines manufacturing applications, exploring the similarities and domain-specific requirements in data quality and management. These projects and related data metrology challenges are outlined below.

2.1. Pharmaceutical manufacturing

Recent developments in digital pharmaceutical manufacturing are generating a large amount of data across varying temporal resolutions and manufacturing routes. This data provides unprecedented opportunities for pharmaceutical manufacturing to derive new insights and efficiencies from experiments but also imposes great challenges in data processing, management, sharing, and integration. Not only data integrity and authenticity are to be ensured, but the processes that lead to the generation of data must be traceable to enable trust.

The pharmaceutical industry introduced “Good Manufacturing Practices” (GMP) to standardise processes around quality, security and effectiveness, but did not make allowances for metrological concepts such as traceability and measurement uncertainty. Data metrology therefore becomes a critical component in understanding and controlling pharmaceutical processes and reducing the variation seen in the final product. NPL has worked with major pharmaceutical manufacturers and researchers to explore their data metrology needs and develop a set of applied research programs.

1) *Ontologies for clinical trial release.* NPL has developed techniques [3] to develop a Domain Agnostic Measurement Ontology, with a view to applying these techniques across different industries. For the past 2 years, NPL has worked with the Medicines Manufacturing Innovation Centre (MMIC) to develop an ontology to aid in the automation and digitalisation of all data required for regulators to approve drugs for consumption. Much of the approvals process is manual data processing which can be replaced by processing of data through modern data driven techniques. The ontology developed by NPL “codifies” all the data relationships which are pertinent to the identification of an expiry date for the release of a drug, facilitating true automation, reducing human input and significantly decreasing the potential for errors in the process due to the development of an approved automated decision process.

2) *Controlled vocabularies for pharmaceutical data exchange.* The development of the ontology for clinical trials release exposed an issue in how data from different companies and manufacturers can differ semantically when describing similar terms. For example, separate companies may use the terms “pill” and “tablet” to describe the same concept. This inconsistency decreases the quality of the information used in digitalised or automated systems. NPL has developed the first iteration of a controlled vocabulary for the clinical trials process to ensure that any automated system can understand the terminology used by each party [4]. This controlled vocabulary provides a traceable

link to quality processes for each company which can aid in automating the verification of the processes used. Within the context of the MMIC use case, this has provided a valid solution for partners and collaborators to deal with terminology harmonisation issues. NPL has also been exploring the idea of working with industry to create an industry-wide standard to create a unified approach to solving this problem and reducing the uncertainty of the information.

3) *Mapping of measurement uncertainty propagation in manufacturing.* NPL has been working on an approach to understand the measurement uncertainty generated at each stage of a continuous manufacturing process. Currently, uncertainty generated at each node is not propagated, so to ensure greater traceability of variation present in the final product, NPL are currently expanding upon pre-existing industrial case studies [5] to develop methods to “map” out the uncertainty present in manufacturing systems and propagate this to each stage.

The goal of these use cases under development is to truly understand the uncertainty of the information produced during pharmaceutical manufacturing and to provide industry with frameworks to understand their data metrology.

2.2. Minimum metadata for biological imaging

Biological imaging (bioimaging) encompasses a vast array of techniques, such as optical microscopy, spectroscopy, multispectral imaging, among others. In the pharmaceutical industry, these techniques are used both in R&D and in clinical studies that evaluate drug resistance, efficacy, targeting mechanisms and pharmacodynamics. Complexity, diversity, and volume of data generated by high-resolution imaging techniques drive the need for advanced analysis and data management methods and require new standards to ensure reproducibility of results and reliability of research [6]. While the efforts to improve data interoperability and definition of minimum reporting standards and metadata are ongoing [7]-[9], stronger engagement of equipment vendors, researchers and funding authorities is needed to create future-proof re-usable and reproducible data repositories.

Seeking such engagement, NPL has been working with two major pharmaceutical industry partners to work on three bioimaging case studies characterised by high data volumes and need for re-use: 1) mass spectrometry imaging (MSI); 2) high content screening; and 3) light sheet microscopy (LSM). To identify the minimum metadata requirements for the three bioimaging domains, subject matter experts (SMEs) were named by each partnering organisation. The SMEs worked with NPL Data Science team to collate the current practices on data management and annotation [10]. The review of the collated metadata revealed the metadata categories illustrated in Figure 1. The categorised domain-specific metadata sets were enhanced with metadata item descriptions, formats, and, where applicable, units of measure and suggestions of standardised terminologies or controlled vocabularies. The results were revised by all SMEs and published in the open NPL report MS 24 [11].

The findings of the study were used to define minimum microscopy metadata recommendations for research data repositories [12]. At NPL, the minimum metadata specifications for LSM and MSI were used extensively to develop frameworks and tools for bioimaging data capture and annotation [13].

2.3. Digital Pathology

Clinical histopathology describes a study of stained tissue sections on glass slides under a microscope, whereby

pathologists manually change the brightness, focus depth, and the region of interest. In digital pathology (DP), tissue samples are digitised using a whole slide imaging (WSI) scanner. The resulting high-resolution images (1-4 GB) can be studied in silico by image analysis software or on-display by a trained pathologist. The DP workflow poses multiple metrological challenges: reproducibility and repeatability of tissue processing, calibration and traceability of WSI, as well as uncertainty analysis to support diagnosis.

The NPL Digital Pathology inter-disciplinary project, launched in 2020, comprised a landscape exercise during which DP experts and stakeholders identified priority areas for metrology support [14], [15]. The outcomes of the landscape exercise were used to shape demonstrator studies with real-world data. Within the PITHIA trial collaboration (<http://www.pithia.org.uk>, Grant Reference Number PB-PG-1215-20033), the project team are studying the uncertainties in the diagnosis based on kidney biopsy images, aiming to 1) locate the sources of uncertainty in decision making and find tools to reduce it, as well as 2) find image features that correlate with clinical outcomes to increase reproducibility and explainability in WSI evaluation.

The preliminary findings (Figure 2) show how the on-display assessment method influences the diagnostic results: when a blood vessel wall thickness is measured directly (red line, lower left diagram), the assessors show preference for more uniform score assignment than if the wall thickness is calculated as a difference (outer diameter - lumen diameter) / 2 (blue line, lower right diagram). Further case studies will include analysis of measurable image features and their association with diagnostic predictions, as well as impact assessment of intra- and inter-WSI device variability on image features and diagnosis.

Future work will include engaging with standards bodies to include metrology-enabling contextual data such as calibration results, device settings etc. into clinical DP standards such as Digital Communications for Medical Imaging (DICOM) and Fast Healthcare Interoperability Resources (FHIR). These standards have high maturity levels and provide mechanisms to include metrological metadata and requirements such as units of measure, clinical terminologies, ontologies, unique identifiers etc.

2.4. Medical sensors case study

While WSI data and associated measurement information can be captured using the existing DICOM standard, novel medical devices require modification of existing standards to capture new data types and provide integration into the healthcare infrastructure. NPL worked with a UK-based medical device developer to create clinically interoperable data structures to store and manage the data from a novel surgical sensor. This opportunity facilitated the capture of valuable metrological information including traceability and calibration *ab initio*, creating a metrologically sound data model at the early stage of device development.

An example of how custom measurement-related information can be included into DICOM metadata is presented in Table 1.

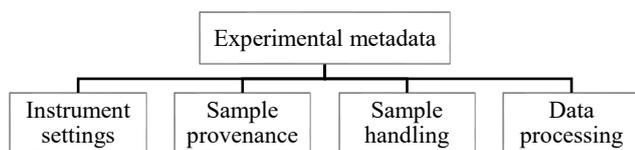


Figure 1. Metadata categories in bioimaging

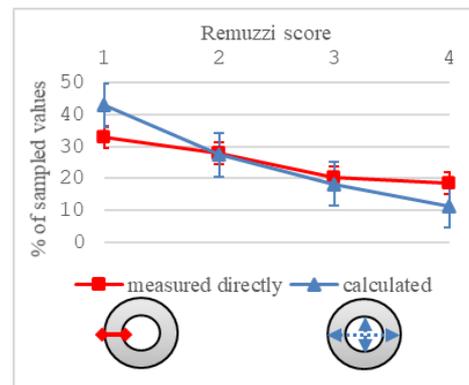


Figure 2. Impact of measurement method on clinical assessment (Remuzzi score). Red line: wall thickness is measured directly. Blue line: wall thickness is calculated from vessel outer diameter and lumen diameter. Image courtesy of Tobi Ayori.

A custom value (patient tilting angle in degrees) is enclosed in a Concept Name Code Sequence that refers to the coding document and provides the value inclusive of its format (value representation (VR)). Note that the value description includes the unit of measure and the reference terminology (Measurement Units Code Sequence).

2.5. Digital health

New measurement modalities within healthcare are creating vast amounts of high-dimensional data from disparate sources and of varying quality, including genomic and imaging data, biomarkers, electronic healthcare records and data from wearable devices. The current and future healthcare practices across the world are increasingly reliant on the integration of these diverse, complex, and large datasets as well as trusted and robust analysis methods [16].

The data curation process in healthcare includes extraction, de-identification, and annotation of datasets with metadata, as well as data fusion and linkage. Therefore, future-proof secure scalable curation methods that handle rapidly growing data volumes are needed.

NPL runs an ongoing inter-disciplinary Digital Health programme aimed to use data metrology tools to help solve some of the important and emerging challenges of utilising healthcare data [17]. The project includes several case studies, some of which are briefly described below, and further details can be found in the 2021 report [18].

Table 1. Including custom measurement value, units of measure and reference to ontology in DICOM metadata.

Tag description	Tag	VR	Value
Concept Name Code Sequence	(0040, A043)	SQ	-
Code Value	(0008, 0100)	SH	'1.2.2-1'
Coding Scheme Designator	(0008, 0102)	SH	'ASCODE'
Coding Scheme Version	(0008, 0103)	SH	'1.0'
Code Meaning	(0008, 0104)	LO	'Patient tilting angle'
Numeric Value	(0040, A30A)	DS	'-19.05'
Measurement Units Code Sequence	(0040, 08EA)	SQ	-
Code Value	(0008, 0100)	SH	'deg'
Coding Scheme Designator	(0008, 0102)	SH	'UCUM'
Coding Scheme Version	(0008, 0103)	SH	'1.4'
Code Meaning	(0008, 0104)	LO	'degrees'

One of the case studies investigates whether it is possible to improve the data quality and comparability by linking patient images with imaging device calibration data. The study set out to link megavoltage computed tomography (MVCT) images used for image-guided radiotherapy with MVCT device calibration data from the routine monthly quality assurance tests that check whether the scanner is fit-for-purpose. MVCT images are routinely used for patient positioning, radiation dosimetry, and in-treatment therapy effect assessment. Like other medical imaging modalities, MVCT images are subject to temporal and inter-device variations that are known to have negative influence on the accuracy of subsequent radiation dose calculation and image segmentation. We implemented a procedure that includes the device calibration information into the DICOM header information of the patient scan. We expect that the MVCT calibration data can be used to remove the device-related variability and make the patient images more inter-comparable, reduce the variations in the image quality, improving the accuracy of analysis, safety, and efficiency of data-driven clinical interventions [18].

Another case study focussed on the development of data-driven models to identify key prognostic markers in computerised medical records (CMR). CMR are a powerful source of information as they contain population level health indicators. These data can be used for estimations of disease incidence, provide insight into disease complexity and identify sub-groups of patients, among other things. National and regional level data aid decision-making in response to potential disease outbreaks, while identification of patient sub-groups can aid treatment planning, moving towards personalised medicine. Despite the enormous potential, identifying trends in large primary care data and inferring meaning from these data is extremely challenging due to their complexity, heterogeneity, dimensionality, incompleteness, and noisiness. CMR data are often mixed-type, making traditional data analysis tools unavailable. A generic data pre-processing and deep learning approach for visualisation and analysis of CMR data has been developed at NPL [19]. The tools enable the analysis of CMR data, as well as other related data types, such as demographics, metadata, medical histories, in a way that identifies non-linear patterns in an unlabelled manner. The features that form patient clusters can be linked back to the input data and interpreted by the clinician or stakeholder to aid in their decision making in complex healthcare scenarios. This framework can also be applied as a data exploration study to obtain data-driven hypotheses that can be tested with further data.

A further case study in the Digital Health programme evaluates how data linkage can be used to improve the quality of life and long-term treatment outcomes for prostate cancer patients by using the patient care data acquired outside of clinical trials. We developed an ontology-based data curation framework to identify and collate information about diagnosis, symptoms, and treatment side effects from routine primary care electronic health records. This work is a first step to increase the utility of primary care data for oncology by a) creating a knowledge base of data sources, b) mapping out the required integration efforts, and c) developing a practical ontology-based method for systematic and reproducible prostate cancer case identification and validating this method on real-world datasets. The developed ontology can be used to standardise the identification and retrieval of prostate cancer cases from primary care data [20].

NPL's most recent endeavours to increase availability and reliability of medical data include developing a curated data

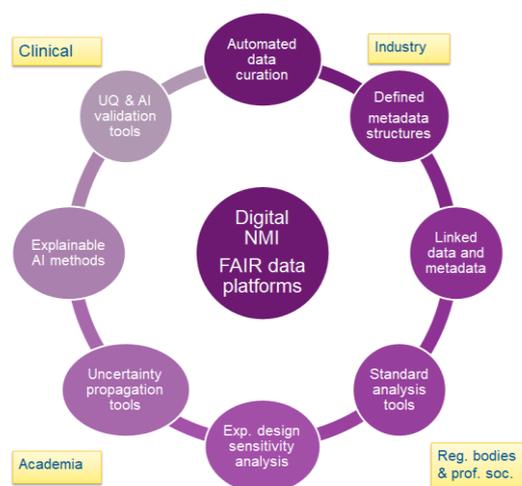


Figure 3. FAIR data platforms for clinically relevant research

platform. The platform will provide mechanisms for curation, storage, metadata annotation, linkage, and analysis of clinically relevant imaging, audit, and calibration data (Figure 3). Such a platform would provide a much-needed foundation to enable access to a richer and larger dataset than what is currently available, rendering the data FAIR-er, and thus increasing its value and utility.

3. CONCLUSIONS

This work presents a range of use cases and demonstrator studies in life sciences and healthcare developed by NPL through active collaborations with industry partners and researchers in digital pathology, bioimaging, pharmaceutical and bio-manufacturing. It is aimed to highlight the need for data metrology in life sciences and healthcare and to stress the role of National Measurement Institutes in these areas. Despite the relative heterogeneity of the presented case studies, the identified problems feature similarities including (a) missing metadata specifications, (b) lack of mechanisms to capture, exchange and propagate metrological information such as calibration data from data acquisition during measurement to its processing and (c) lack of methods to combine and propagate uncertainties in data processing chains.

The three problems listed above call for a systematic approach to data curation and metadata annotation based on the need for FAIR-ness and data reusability. Although the missing metadata specifications can be addressed using custom ontologies and controlled vocabularies, striving towards standards and minimum data quality requirements is recommended to increase data re-usability and impact across different sectors/companies. Furthermore, there is a variety of existing open standards and formats that can and should be used to manage data from new medical devices and imaging modalities. These standards can be adapted to incorporate information pertaining to metrological traceability and uncertainty. Lastly, while the use of, and need for, metrology methods is widely recognised in physics and engineering, in life sciences, medicine and pharmaceutical manufacturing these tools are often added as an afterthought, if considered at all. Therefore, work is required to demonstrate the need for and the impact of data metrology via case studies in the respective domains.

The NPL Data Science team believes that the identified challenge areas highlight both the need for heterogeneous

approaches to Data Metrology as well as common pain points across these fields. The findings presented in this paper call for a proactive and consistent approach to generating and using *quality* data. FAIR Data Platforms such as that shown in Figure 3 demonstrate an end-to-end approach to how data should be treated to ensure adherence to the FAIR principles and reduce any uncertainty generated due to the processing or labelling of the data.

ACKNOWLEDGEMENT

This work was funded by the UK Government Department for Business, Energy & Industrial Strategy through the UK's National Measurement System. We would also like to thank our partners at the MMIC; CPI, University of Strathclyde, UKRI, Scottish Enterprise, AstraZeneca and GSK as well as ArtioSense Ltd and the PITHIA trial investigators. Thanks to Michael Chrubasik, Louise Wright, and Peter Harris for providing feedback on the manuscript.

REFERENCES

- [1] D. Taylor, The Pharmaceutical Industry and the Future of Drug Development, Pharmaceuticals in the Environment, Edited by R. E. Hester; R. M. Harrison, 2015, pp. 1–33. DOI: [10.1039/9781782622345-00001](https://doi.org/10.1039/9781782622345-00001)
- [2] J. S. Srai, C. Badman, M. Krumme, M. Futran, C. Johnston, Future Supply Chains Enabled by Continuous Processing-Opportunities and Challenges, Continuous Manufacturing Symposium, 20–21 May 2014, J. Pharm. Sci., vol. 104, 3 (2015), pp. 840–849. DOI: [10.1002/jps.24343](https://doi.org/10.1002/jps.24343)
- [3] J.-L. Hippolyte, M. Chrubasik, F. Brochu, M. Bevilacqua, A domain-agnostic ontology for unified metrology data management, Meas. Sens., 18 (2021), p. 100263. DOI: [10.1016/j.measen.2021.100263](https://doi.org/10.1016/j.measen.2021.100263)
- [4] P. M. Duncan, D. S. Whittaker, Distribution identification and information loss in a measurement uncertainty network, Metrologia., 58 (2021), 034003. DOI: [10.1088/1681-7575/abef8](https://doi.org/10.1088/1681-7575/abef8)
- [5] M. Chrubasik, C. Lorch, P. M. Duncan, Ontology-Based Rest-APIs for Measurement Terminology: Glossaries as a service, IMEKO TC6 Int. Conference on Metrology and Digital Transformation, Berlin, Germany, 19-21 September, 2022. DOI: [10.21014/tc6-2022.023](https://doi.org/10.21014/tc6-2022.023)
- [6] B. J. Heil, M. M. Hoffman, F. Markowitz, Su-In Lee, C. S. Greene, S. C. Hicks, Reproducibility standards for machine learning in the life sciences, Nat Methods, 18 (2021), p. 1132–1135. DOI: [10.1038/s41592-021-01256-7](https://doi.org/10.1038/s41592-021-01256-7)
- [7] C. Allan, J.-M. Burel, J. Moore, C. Blackburn, M. Linkert, S. Loynton, D. MacDonald, W. J. Moore, C. Neves, A. Patterson, M. Porter, A. Tarkowska, B. Loranger, J. Avondo, I. Lagerstedt, L. Lianas, S. Leo, K. Hands, R. T. Hay, A. Patwardhan, C. Best, G. J. Kleywegt, G. Zanetti, J. R. Swedlow, OME Remote Objects (OMERO): a flexible, model-driven data management system for experimental biology, Nat. Methods, vol. 9, 3 (2012), pp. 245–253. DOI: [10.1038/nmeth.1896](https://doi.org/10.1038/nmeth.1896)
- [8] O. J. R. Gustafsson, L. J. Winderbaum, M. R. Condina, B. A. Boughton, B. R. Hamilton, E. A. B. Undheim, M. Becker, P. Hoffmann., Balancing sufficiency and impact in reporting standards for mass spectrometry imaging experiments, GigaScience, vol. 7, 10 (2018). DOI: [10.1093/gigascience/giy102](https://doi.org/10.1093/gigascience/giy102)
- [9] M. Huisman, M. Hammer, A. Rigano, U. Boehm, J. J. Chambers, N. Gaudreault, A. J. North, J. A. Pimentel, D. Sudar, P. Bajcsy, C. M. Brown, A. D. Corbett, O. Faklaris, J. Lacoste, A. Laude, G. Nelson, R. Nitschke, D. Grunwald, C. Strambio-De-Castillia; Minimum Information guidelines for fluorescence microscopy: increasing the value, quality, and fidelity of image data, ArXiv191011370 Cs Q-Bio, (2020). Online [Accessed 9 March 2020] <http://arxiv.org/abs/1910.11370>
- [10] E. Cooke, M. Hayes, M. Romanchikova, Acquisition and management of high content screening, light-sheet microscopy and mass spectrometry imaging data at AstraZeneca, GlaxoSmithKline and NPL: a survey report, NPL Report. MS 25, (2020). DOI: [10.47120/npl.MS25](https://doi.org/10.47120/npl.MS25)
- [11] F. Brochu, J. Bunch, E. Cooke, A. Dexter, M. Romanchikova, M. Shaw, T. R. Steven, S. A. Thomas, Federation of Imaging Data for Life sciences: current status of metadata collection for high content screening, mass spectrometry imaging and light sheet microscopy of AstraZeneca, GlaxoSmithKline and NPL, NPL Report. MS 24, (2020). DOI: [10.47120/npl.MS24](https://doi.org/10.47120/npl.MS24)
- [12] U. Sarkans, W. Chiu, L. Collinson, M. C. Darrow, J. Ellenberg, D. Grunwald, J.-K. Hériché, A. Iudin, G. G. Martins, T. Meehan, K. Narayan, A. Patwardhan, M. R. G. Russell, H. R. Saibil, C. Strambio-De-Castillia, J. R. Swedlow, C. Tischer, V. Uhlmann, P. Verkade, M. Barlow, O. Bayraktar, E. Birney, C. Catavittello, C. Cawthorne, S. Wagner-Conrad, E. Duke, P. Paul-Gilloteaux, E. Gustin, M. Harkiolaki, P. Kankaanpää, T. Lemberger, J. McEntyre, J. Moore, A. W. Nicholls, S. Onami, H. Parkinson, M. Parsons, Marina Romanchikova, N. Sofroniew, J. Swoger, N. Utz, L. M. Voortman, F. Wong, P. Zhang, G. J. Kleywegt, A. Brazma, REMBI: Recommended Metadata for Biological Images - enabling reuse of microscopy data in biology, Nat. Methods, 18 (2021), pp. 1–5. DOI: [10.1038/s41592-021-01166-8](https://doi.org/10.1038/s41592-021-01166-8)
- [13] S. Thomas, F. Brochu, A framework for traceable storage and curation of measurement data, Meas. Sens., 18 (2021), pp. 100201. DOI: [10.1016/j.measen.2021.100201](https://doi.org/10.1016/j.measen.2021.100201)
- [14] M. Adeogun, J. Bunch, A. Dexter, C. Dondi, T. Murta, C. Nikula, M. Shaw, A. Taylor, I. Partarrieu, M. Romanchikova, N. A. S. Smith, S. A. Thomas, J. Venton, Metrology for Digital Pathology. Digital pathology cross-theme project report, NPL Report. AS 102, (2021). DOI: [10.47120/npl.AS102](https://doi.org/10.47120/npl.AS102)
- [15] M. Romanchikova, S. A. Thomas, A. Dexter, M. Shaw, I. Partarrieu, N. A. S. Smith, J. Venton, M. Adeogun, D. Brettle, R. J. Turpin, The need for measurement science in digital pathology, Journal of Pathology Informatics, (2022), 100157, preprint. DOI: [10.1016/j.jpi.2022.100157](https://doi.org/10.1016/j.jpi.2022.100157)
- [16] The Topol Review - NHS Health Education England. Online [Accessed 31 March 2022] <https://topol.hee.nhs.uk/>
- [17] N. A. S. Smith, D. Sinden, S. A. Thomas, M. Romanchikova, J. E. Talbott, M. Adeogun, Building confidence in digital health through metrology, Br. J. Radiol., vol. 93, 1109 (2020), pp. 20190574. DOI: [10.1259/bjr.20190574](https://doi.org/10.1259/bjr.20190574)
- [18] N. A. S. Smith, M. Romanchikova, I. Partarrieu, E. Cooke, A. Lemanska, S. Thomas, NMS 2018-2021 Life-sciences and healthcare project “Digital health: curation of healthcare data” - final report, National Physical Laboratory, NPL Report. MS 31, (2021). DOI: [10.47120/npl.MS31](https://doi.org/10.47120/npl.MS31)
- [19] S. A. Thomas, N. A. S. Smith, V. Livina, I. Yonova, R. Webb, S. de Lusignan, Analysis of Primary Care Computerized Medical Records (CMR) Data with Deep Autoencoders (DAE), Front. Appl. Math. Stat., 5 (2019), 12 pp. DOI: [10.3389/fams.2019.00042](https://doi.org/10.3389/fams.2019.00042)
- [20] A. Lemanska, S. Faithfull, H. Liyanage, S. Otter, M. Romanchikova, J. Sherlock, N. A. S. Smith, S. A. Thomas, S. de Lusignan, Primary Care Prostate Cancer Case Ascertainment, Stud. Health Tech. Inf., 270 (2020), pp. 1369-1370. DOI: [10.3233/SHTT200446](https://doi.org/10.3233/SHTT200446)