Applied Ontology 17 (2022) 401–421
DOI 10.3233/AO-220271
IOS Press

CHAMEO: An ontology for the harmonisation of materials characterisation methodologies

Pierluigi Del Nostro [a], Gerhard Goldbeck [a] and Daniele Toti [a,b,*]
[a] *Goldbeck Consulting Limited, Cambridge, United Kingdom*
E-mails: pierluigi@goldbeck-consulting.com, gerhard@goldbeck-consulting.com
[b] *Department of Mathematics and Physics, Catholic University of the Sacred Heart, Brescia, Italy*
E-mail: daniele.toti@unicatt.it

Abstract. The field of materials characterisation encompasses a wide range of methods and related research communities. This has led to a proliferation of terminologies and data management approaches, hindering collaboration and interoperability. In this work, a domain ontology designed to model the common aspects across the different characterisation methodologies is presented. This ontology, called the CHAMEO ontology, is based on a recent CEN Workshop Agreement (CWA 17815) which introduced a standardised terminology and the Characterisation Data (CHADA) documentation scheme. The goal of CHAMEO is to provide a framework for harmonising the underlying method-specific ontologies, which can be developed by reusing and specialising the generic constructs of the CHAMEO ontology. This work is part of a broader initiative under the umbrella of the European Materials Modelling Council (EMMC), for the development of interconnected materials modelling ontologies based on a common root that is the Elementary Multiperspective Material Ontology (EMMO). The CHAMEO ontology was developed within the NanoMECommons European project that has the goal of harmonising characterisation protocols. The CHAMEO ontology has also been aligned with a number of recently developed, EMMO-based domain ontologies for the classification of materials, models, manufacturing processes and software products related to Materials Modelling.

Availability. The axiomatization of the ontology is stored in a GitHub repository available at: https://github.com/emmo-repo/domain-characterisation-methodology, and is published at the following URL: http://emmo.info/emmo/domain/chameo/chameo.

Keywords: Ontology, materials modelling, knowledge sharing, EMMO, materials characterisation, interoperability, materials

Accepted by: Emilio Sanfilippo

1. Introduction

The huge variety and complexity of materials led to the formation of multiple communities around the materials characterisation field, establishing different terminologies typically focusing on specific application domains. This work represents a contribution to the creation of a common knowledge framework for the documentation of characterisation methods, with the goal of facilitating reusability and transferability of knowledge across different communities and sectors.

Specifically, this work describes the CHAMEO ontology, conceived to be a reference framework for the description of materials characterisation procedures. The CHAMEO ontology models the generic

*Corresponding author. E-mail: daniele.toti@unicatt.it.

1570-5838/$35.00 © 2022 – IOS Press. All rights reserved.

aspects that are in common across the different characterisation techniques, providing definitions at the methodological level. Specific ontologies for modelling the different characterisation techniques could then be developed by specialising the CHAMEO definitions. This follows a modular design approach that increases the level of reusability and interoperability and reduces the effort for maintaining the ontologies.

The development of the CHAMEO ontology has been carried out under the European project NanoMECommons (NanoMECommons, 2021–2025) that aims to develop harmonised and widely accepted characterisation protocols. These protocols are meant to be integrated into real industrial environments in order to boost material, process, and product reliability, by reducing measurement discrepancies and improving data interoperability and traceability. To pursue this goal, ontologies play a central role, providing a unique and interoperable metadata structure and enhancing data quality and information management. Their purpose is to support the establishment of data-driven structure-property relationships and thus assist the quality assurance and material design procedures in industry. The ultimate objectives are to bring about semantics-based standards for materials characterisation and to contribute directly to Industry Commons, *i.e.* facilitating reusability and transferability of characterisation data across multiple manufacturing sectors.

The CHAMEO ontology is a key part of bringing about the overarching goal of NanoMECommons to provide a harmonised and standardised representation of materials characterisation, including multi-technique protocols.

NanoMECommons is one of a number of European projects under the umbrella of the European Materials Characterisation Council (EMCC, 2016–2022) and the European Materials Modelling Council (EMMC, 2014–2022) which have the broader goal of fostering the interaction and collaboration among different stakeholders in the area of Materials Science, and facilitating integration and interoperability of materials modelling and characterisation. For this purpose, in a multidisciplinary effort, the Elementary Multiperspective Material Ontology (EMMO) (The EMMC Consortium, 2014–2021) has been developed as a representational ontological framework based on the actual picture of the physical world, by following the principles of applied sciences, and in particular physics and materials sciences. CHAMEO is conceived as a domain ontology for materials characterisation methodologies, based on EMMO's top and middle level portions of the ontology.

This paper is organised as follows. In Section 2 an overview of relevant related work is provided. Section 3 illustrates what evidence is used for the definition of the ontology's scope, which is then described in Section 4 together with the corresponding design criteria. In Section 5 the structure of the CHAMEO ontology is explained in terms of classes and properties and how they relate to the top-level EMMO ontology. Section 6 describes the alignments with other ontologies, while Section 7 highlights CHAMEO's compliance with the FAIR principles. Finally, in Section 8 conclusions are drawn.

2. Related work

In this section pertinent background has been reviewed, including existing initiatives in the field of experiment and process standardisation used in analytical laboratories, as well as terminology, metadata and documentation of materials characterisation (CHADA) and related ontologies, in particular the Elementary Multiperspective Material Ontology (EMMO) which is used as the top- and middle-level framework for the ontology development in NanoMECommons.

2.1. Existing initiatives for standardising scientific data

The goal of introducing data standards is to effectively manage large quantities of scientific data produced by organisations, being able to share and combine them, in order to derive the greatest value from them. Over the years, efforts both at the research and the industrial level, including European and national research projects (for instance Nostro et al., 2013; Arosio et al., 2013; Capuano and Toti, 2019), have been pursuing standardisation of data representation in organisations for scientific, business and social purposes alike, including the usage, alignment and integration of ontologies.

There are many initiatives promoting and supporting the use of scientific data standards ranging from those advocating more general principles to more domain specific efforts. Some relevant examples are mentioned as follows.

Overarching initiatives include:

- The Committee on Data of the International Science Council (CODATA, 2020), promoting global collaboration to advance open science and to improve the availability and usability of data for all areas of research.
- The Research Data Alliance (RDA, Berman and Crosas, 2020), a community-driven initiative launched by institutions in EU, USA and Australia with the goal of building the social and technical infrastructure to enable open sharing and re-use of data, covering all lifecycle stages (production, usage, exchange, processing, storage).

More specifically in Chemicals and Materials Science, relevant associations and initiatives include the following, among which are long-standing efforts in crystallography and recent initiatives in Materials Science.

- The International Union of Crystallography (IUCr, 1947–2014), a scientific union whose objectives are to promote international cooperation in crystallography and to contribute to all aspects of crystallography, facilitating the standardization of methods, units, nomenclatures and symbols, among others. The Crystallography Information File and Framework (both known by the CIF acronym) are by-products of IUCr and are successful examples of standardizations of data exchange formats and protocols within the scientific area of Crystallography, often cited in the Materials Science community as references to what an effective standardization should be.
- The European Materials Modelling Council (EMMC, 2014–2022), promoting the digitalisation of materials knowledge and is the governing body of the EMMO ontology.
- The Materials Research Data Alliance (MaRDA, 2020), a community-led network focused on connecting and integrating U.S. materials research data infrastructure to realize the promise of open, accessible, and interoperable materials data.
- The FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy (FAIR-DI, 2014–2022), supporting the building of a reliable infrastructure for data from basics sciences and engineering.

In the context of the present work, the following initiatives from the life sciences are also very pertinent due to their maturity and as they deal with experimental procedures and analytical laboratories (*i.e.* a particular type of chemicals and materials characterisation).

- The Pistoia Alliance (PA, 2007), a non-profit group of more than 100 companies that aim at lowering innovation barriers in R&D through pre-competitive collaboration, developing best practices and working with regulators to adopt new standards.

- The Allotrope Foundation (AF, 2012), an international consortium of research-intensive industries developing a data architecture and a standards-based framework to acquire, exchange, and manage laboratory data in a standardised way. Both promote the use of the FAIR principles in the management of scientific data.

2.2. *The CHADA document template*

As described in Romanos et al. (2019), the purpose of CHADA is to provide a standard structure for documenting materials characterisation techniques. CHADA has been developed in the OYSTER project, following the 'template' of a similar structured documentation for materials modelling, the MODA. Recently, CHADA has been the subject of a CEN Workshop Agreement. Four types of concepts are used for the classifications of the different steps of an entire characterisation workflow (which can be simply called "characterisation"):

(1) User case, which includes the sample and the information on the environment of testing, the volume of probed material and the information on the surrounding environment, which interacts with the probe and generates a detectable (measurable) signal (information);
(2) Experiment, which represents the process by which the metrological chain is defined; within a single experiment, the following fundamental elements are identified: probe, signal, detector, noise (according to the previous definitions);
(3) Raw data, which is the set of data that is produced directly as output by the metrological chain, usually expressed as a function of time, position, or photon energy.
(4) Data processing, which represents any process (or sequence of processes) by which the data are analysed to arrive at the final shape.

The above concepts are used to structure the CHADA document into sections. Each of the sections groups a set of fields where it is possible to provide textual information about the characterisation methodology. The CHADA document has an introductory section that provides an overview of the overall methodology. The characterisation method can be single or multi-technique. Both cases are documented in the same document template. In case of a multi-technique approach a common overview of the methodology is provided and then, usually, all of the methods are described in each section of the CHADA (User, Experiment, Raw data, Data processing). Such kind of textual document is easily interpretable by humans, but lacks structured data that can help retrieve information on the characterisation methodologies on the base of different dimensions (*e.g.* material, probe, detector, properties).

In the NanoMECommons project, the CHADA textual document was used as a basis to build a more structured and shared knowledge represented by the CHAMEO ontology and taxonomies. The ontology, together with taxonomies, is meant to drive the user input and queries through which it will be implemented in an Open Innovation Environment (OIE), *i.e.* a web platform for the establishment of a wider manufacturing/modelling/characterisation network, whose development started under the OYSTER project (OYSTER, 2017–2022). The data collected through the OIE platform will be used to build a shared, harmonised knowledge base of the characterisation methods.

2.3. *Elementary multiperspective material ontology (EMMO)*

Ontology developments in NanoMECommons build on the groundbreaking work on EMMO, an ontology framework for the applied sciences that was initiated by the EMMC-CSA project (EMMC-CSA, 2016–2019) and is being further developed in a number of projects under the EMMC's governance.

In more detail, NanoMECommons also builds on the mechanical testing ontology work performed in the project (OYSTER, 2017–2022) that provides a first example of a characterisation ontology in the EMMO framework. Since then, there have been further developments in EMMO, for example in relation to workflows that are meant to be utilised in NanoMECommons. EMMO has been developed and is applied in a series of European Projects as shown in the Acknowledgements section of the EMMO repository[1]. EMMO has been conceived and demonstrated to work as a standard representational framework for applied sciences. Of particular importance to achieving this goal are the following points.

(1) EMMO has a minimal top level, which is based on the very few things one can say about the world based on fundamental science, in particular that everything is 4D, that there is a Universe object and a fundamental quantum 4D object and that everything in between cannot be specified a priori, other than to say it is either physical or a void. Everything else is a "Perspective".
(2) There are no abstract objects in EMMO; everything must be described in terms of physical objects.
(3) EMMO middle level ontologies are representations of different Perspectives, *i.e.* different ways of representing objects. Hence pluralism is fundamentally built into EMMO. This is very important in bringing together multi-disciplinary views on the world.
(4) All relationships between objects are of three fundamental types: topological (connectedness), mereological (parthood) and semiotic (signs that communicate meaning). In contrast to other ontologies that often have no or only a very loose way of organising relationships, this provides a strong logic framework to support reasoning.
(5) Semiosis, comprising any form of process that involves the use of symbols or signs for the production of meaning, is fundamental to representing what an individual means to communicate and what is known about objects. This is done in accordance with the so-called "semiotic triangle" (Ogden and Richards, 1923), where a symbol/sign is used to symbolise a thought that refers to a certain object: such a symbol or sign is then said that it stands for that object.
(6) The main Perspectives are:
    (a) Holistic: considers the importance and role of the whole as well as its parts without specific granularity hierarchy, with subclasses Whole (based on some criterion) and Part (as it appears in relation to the Whole, also regarding its role; hence Role is an alternative label for Part, as is also used, for instance, in theatre).
    (b) Persistence: considers the persistence in time (process) or space (object). This perspective also enables to map the widely used 3D ontologies BFO (BFO Discussion Group, 2002) and DOLCE (ISTC-CNR Laboratory for Applied Ontology, 2002) which are based on a similar high-level categorisation.
    (c) Physicalistic: uses applied science concepts to provide meaning to objects (*e.g.* a material as a scientific object, interrogated by scientific means).
    (d) Reductionistic: focuses on a strict hierarchy of objects in terms of granularity levels (in space and/or time). Also useful for a "System of systems" view on engineering.
    (e) Perceptual: includes recognisable patterns in space and/or time such as sounds, languages, alphabets symbols, mathematics, graphics.
(7) EMMO makes extensive use of annotations and alternative labels to cover the varied meaning attached to words by different communities and different standards (*e.g.* the definition of Product

[1] https://github.com/emmo-repo/EMMO#acknowledgement

varies between ISO 9000 and ISO 14040, hence the name Product as such is not sufficient to describe the meaning, even within ISO standards.)

3. Evidence for scope

Evidence for the ontology scope development was gathered from the NanoMECommons partners from various perspectives in order to capture a wide range of aspects. These include:

- An overview of industrial cases, distinguishing 9 specific cases in the industrial domains tackled by the project (Organic Electronics, Avionics and Automotive, Additive Manufacturing, Energy and Chemicals).
- Competency Questions for the NanoMECommons ontology, using an established method for the development of and agreement on the ontology scope.
- Collection and review of CHADA from projects partners following the NeOn methodology (Suarez-Figueroa and Gomez-Perez, 2009) and using CHADA as a "tabular technique" which provides the basis for structuring information similar to that described in METHONTOLOGY (Fernandez-Lopez et al., 1997).

Details are provided in the following subsections.

3.1. Competency questions

Forty seven competency questions were collected from NanoMECommons' industrial partners. For each question there is a unique ID, the partner, the characterisation method (if the question is related to a specific method), notes describing how the question is addressed by the ontology, and if the question is in the scope or not. From the competency questions a first set of terms has been extracted. In Table 1 three examples of competency questions are reported, where the extracted concepts are highlighted in bold. The full list of competency questions can be found in the Appendix.

What arose is that, except for the questions referring to the detailed data, all of the other questions are in the ontology scope and can in principle be answered by querying the NanoMECommons knowledge base (concepts, relationships, individuals). Detailed operational data, like the sequences of measurements performed by the device, can be stored in external systems and referenced as external resources.

3.2. CHADA from the industrial cases

A number of industrial cases from NanoMECommons have been selected for the design of the CHAMEO ontology. For each of them a CHADA document with the description of the characterisa-

Table 1

Three examples of competency questions

Competency questions				
ID	Question	Method	Notes	In scope
CQ1	Which kind of **material** has been tested (**metal**, **ceramic**, **polymer**, ...)?	Nanoindentation	Material class and instances	Yes
CQ19	What is the speed per indent? (*i.e.* 1 indent/second, 1 indent/minute, ...)	Nanoindentation	Actual measurement data is stored outside the ontology	No
CQ38	What was the **characterisation environment** (**ambient**, **inert atmosphere**)?	Spectroscopic Ellipsometry	Characterisation environment class and instances	Yes

Table 2
An example of an industrial case

Industrial cases					
ID	Industrial case/material	Objectives	Surface preparation	Workflow methods	Properties
IQ1	Batteries (*e.g.* Li-based cathode composites)	A protocol for in-situ assessment of nanomechanical properties (fracture toughness, residual stress, and elastic modulus) in battery materials	none	Scanning Electron Microscopy (SEM) Nanoindentation Electrical testing	SEM image Modulus and Hardness Impedence State of charge State of health

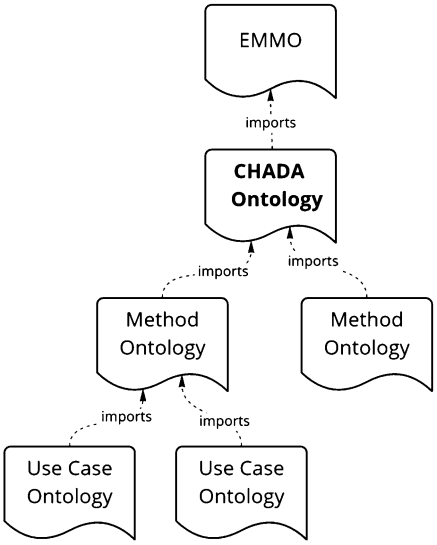


Fig. 1. Modular ontology design.

tion technique has been analysed. In Table 2 there is an example of an industrial case with its title, objectives, the surface preparation method, the workflow methods used for selecting the area of testing or topology and for obtaining the material properties, and the properties of interest that are output of the characterisation.

By analysing how the different CHADA documents were filled, especially for the cases of complex workflows where multiple techniques are used to obtain the final material property, it arose that the way the information is provided by different users is subject to interpretation, and this makes it hard to get homogeneous information about the different characterisation techniques. The CHAMEO ontology is meant to address this issue by providing a framework for defining a clear, machine-readable documentation, based on shared concepts and definitions.

4. Ontology scope and design

A modular approach is adopted for the design of the ontologies for the characterisation of materials, as depicted in Fig. 1.

The EMMO is the reference framework for the applied sciences, as described in Section 2.3, and includes upper and middle level concepts useful for the development of domain specific ontologies. Based on the EMMO, the CHAMEO ontology provides the constructs that are generic and transversal with respect to the different methods of characterisation. Specific method ontologies are then developed using the CHADA constructs for describing each characterisation method (*e.g.* Mechanical Testing). For each characterisation method there can be use case ontologies (*e.g.* Nanoindentation, Fatigue, Cable Bundle). The modular approach allows to maximise reusability and reduces the effort for maintaining the ontologies.

As a semantic model built on the textual CHADA document, the CHAMEO ontology retains the flexibility to have both structured information, also exploiting taxonomies where possible, and human readable textual descriptions.

The definitions listed below, extracted from the CEN/CENELEC Workshop Agreement (CWA 17815, CEN, 2021) were used for the ontology design:

- **Measurement**: A specific activity that results in the determination of values (*e.g.* coordinate values, length values, false/true values).
- **Characterisation method**: Characterisation method is the combination of the experiment and the associated experimental plan, pre-processing and post-processing activities, to arrive at a measured material structure/property.
- **Characterisation workflow**: Combination of more than one characterisation method to form the entire workflow that is needed to describe the desired material's behaviour and/or performance. Example: mechanical testing can be combined "in-situ" with microscopy and spectroscopy tools to develop a multi-technique characterisation workflow.

The ISA-88 standard (American National Standards Institute and International Society of Automation and Instrument Society of America, 2010) is considered as a base to model the process hierarchy in the ontology. In the ISA-88 the *Process Model* describes how a batch process can be decomposed into a hierarchy, namely:

- **Process**: A process that leads to the production of finite quantities of material by subjecting quantities of input materials to an ordered set of processing activities over a finite period of time using one or more pieces of equipment.
- **Stage**: A part of a process that usually operates independently from other process stages and that usually results in a planned sequence of chemical or physical changes in the material being processed.
- **Operation**: A major processing activity that usually results in a chemical or physical change in the material being processed.
- **Action**: Minor processing activities that are combined to make up a process operation. Process actions are the lowest level of processing activity within the process model.

The ISA88 specifically refers to the concept of *Batch Process*, but its main principles can be adapted for the CHAMEO ontology. Figure 2 depicts the process hierarchy referred to, where it has been made explicit that a process can have sub-processes. The following mapping between the CWA and ISA-88 definitions clarifies how the process is modelled in the ontology. The Characterisation Workflow is the overall process, composed by Characterisation methods (can be many in the multi-technique approaches); thus the latter are sub-processes of the Characterisation Workflow. Each Characterisation method is typically made up of multiple stages (*e.g.* calibration, measurement, post-processing). Stages

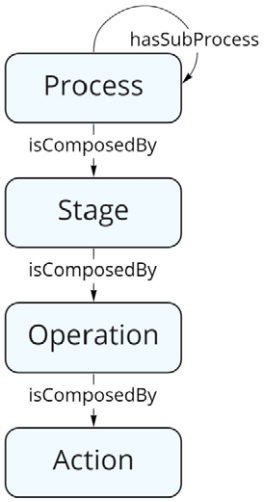


Fig. 2. ISA88 process hierarchy.

are currently the level of the hierarchy where the CHADA ontology stops. Further decompositions into operations and activities can be done in sub-ontologies of the CHAMEO ontology describing the details of specific methods (*e.g.* Mechanichal Testing, Focus Ion Beam, RAMAN).

5. The CHAMEO ontology

This section provides a description of the CHAMEO ontology classes and properties and how they are mapped to the EMMO ontology. The alignment with other external ontologies will be discussed in Section 6.

In the following description, class names are in bold, whereas object properties are in bold and italic, and datatype properties are in italic. In the ontology diagrams, classes are depicted as rectangles (yellow for the CHADA ontology classes, light green for the EMMO classes, white for other external ontology classes); individuals have been depicted as blue rectangles in order to visually distinguish them from the classes; and relationships as arrows with dotted lines when referring to object/datatype properties (in blue and gray, respectively) and arrows with solid lines when referring to subclass properties/*rdf:type* properties (black and purple, respectively). Datatype property values are also displayed within white rectangles where appropriate. Firstly, an overview on the characterisation workflow and method is provided (see Fig. 3); subsequently, details are described on how the stages, materials and devices involved in the characterisation method are modelled.

- **chameo:CharacterisationWorkflow**: This class models the overall characterisation workflow that can be a composition of different methods used to obtain the final property. The CharacterisationWorkflow can have a graphical representation, modelled as **emmo:Icon**, linked through the ***emmo:hasIcon*** object property. This class is defined as a subclass of **emmo:Process** and **emmo:Whole**. The latter class in EMMO is part of the holistic perspective and is used to model an entity which is defined according to a unity criteria that relates holistically its parts to form a whole.
- **chameo:CharacterisationMethod**: Represents a specific characterisation method that is defined as a subclass of **emmo:Observation** and **emmo:SubProcess** (being it a sub-process

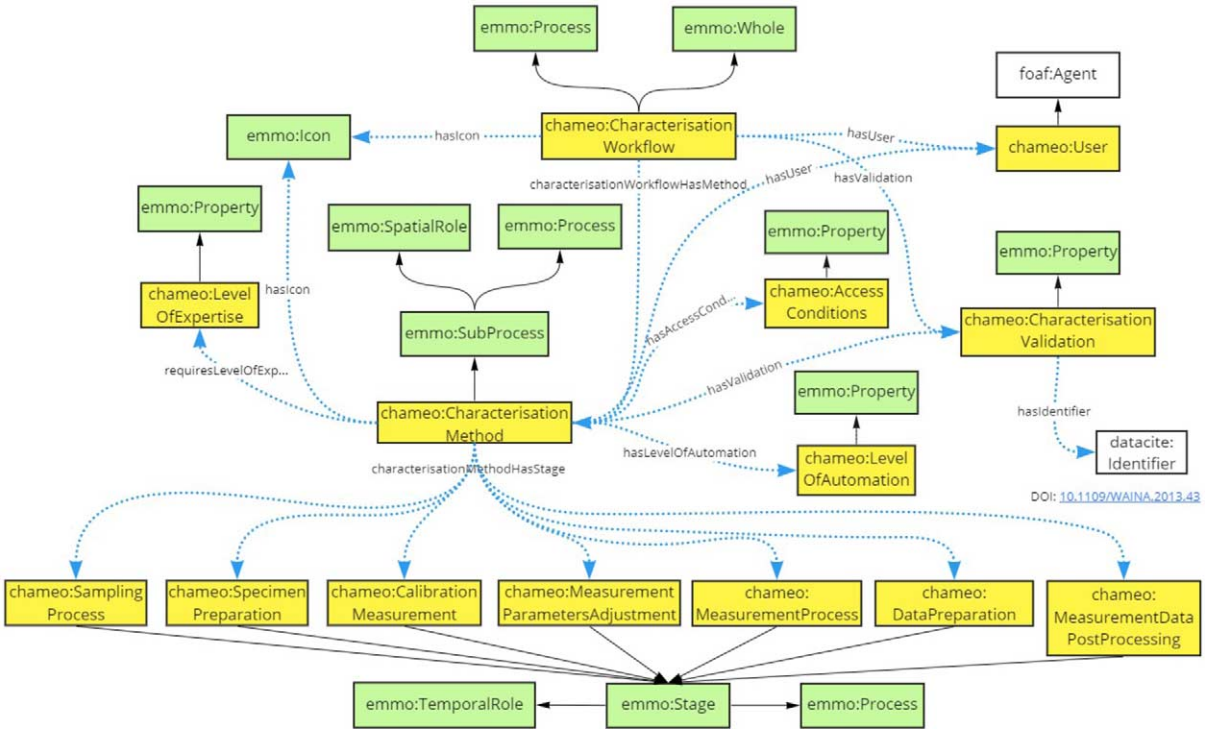


Fig. 3. Characterisation workflow and its components.

of the overall workflow and then a spatial part of it as a whole). For each characterisation method it is also possible to specify a required level of expertise (**chameo:LevelOfExpertise** through ***chameo:requiresLevelOfExpertise***) and whether there is a certain level of automation (**chameo:LevelOfAutomation** through ***chameo:hasLevelOfAutomation***). The characterisation method is typically made up of several stages (manufacturing process, sampling process, specimen preparation, calibration process, measurement, data preparation, data post processing), which are detailed below.

- **chameo:User**: This is the class of the users participating in a characterisation method, defined as a subclass of **foaf:Agent**. Each method can be run by a different user, since a different expertise can be required.
- **chameo:CharacterisationValidation**: A scientific validation of the characterisation technique can be specified both for the single method or for the entire workflow, using the ***hasValidation*** object property. The datatype property *hasPeerRevArt* allows to specify, for each instance of **chameo:CharacterisationValidation**, the reference to a scientific paper that validates the approach (*e.g.* using a Document Object Identifier).

In Fig. 4 the class hierarchy of specimen, sample and material is depicted, together with the processes that are related to manufacturing, sampling and specimen preparation. The following definitions are adopted.

- **chameo:Sample**: Portion of material selected from a larger quantity of material (**emmo:Material**). Through the EMMO object properties ***emmo:hasPart*** and ***emmo:hasProperty*** it is possible to spec-

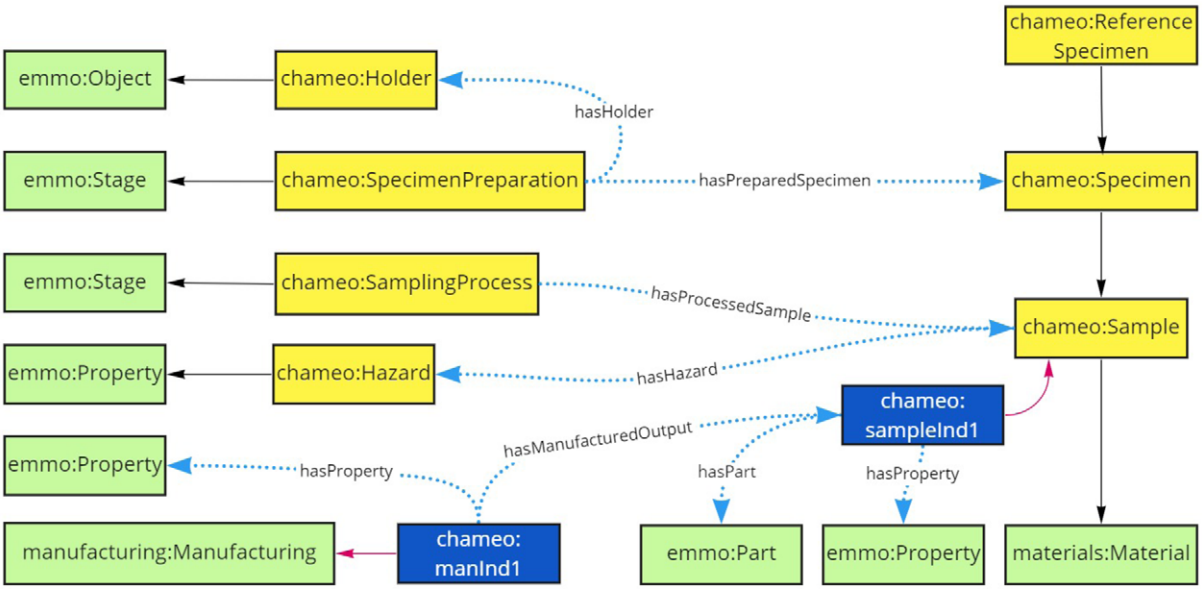


Fig. 4. Specimen, sample, material.

ify, at the individual level, if a sample is composed by different parts/layers and if there are properties of interest.

- **chameo:Specimen**: Portion of material taken under conditions such that the sampling variability cannot be assessed (usually because the population is changing and is assumed, for convenience, to be zero). The specimen is the object that is actually used for the observation. Specimen can also be the result of a series of sample preparation processes before actual observation and measurement.
- **chameo:SamplingProcess**: Act of extracting a portion (amount) of material from a larger quantity of material. This stage results in obtaining a sample that is representative of the batch with respect to the property or properties being investigated.
- **chameo:SpecimenPreparation**: Class for the specimen preparation processes (*e.g.* , machining, polishing, cutting to size, *etc.*). The preparation of a specimen can also imply the use of a holder (**chameo:Holder**).
- **manufacturing:Manufacturing**: This class represents the processes for transforming a precursor object into a product by using manual labour, machinery or chemical/biological processes. For the scope of this ontology, the interest is not on the details of the manufacturing process of the sample but on some of its properties, in order to be able to correlate them with the properties that are output of the characterisation. These properties can be specified at the individual level by attaching them to an individual of **manufacturing:Manufacturing**, using the ***emmo:hasProperty*** object property.

The portion of ontology in Fig. 5 illustrates how the characterisation system is modeled, as described below.

- **chameo:CharacterisationSystem**: This class represents the whole system used for the characterisation. It is a subclass of the **emmo:HolisticSystem**, defined as an object that is made up of a set of sub-objects working together as parts of a mechanism or an interconnecting network (natural or artificial); a complex whole. It includes the probe and the characterisation machine.

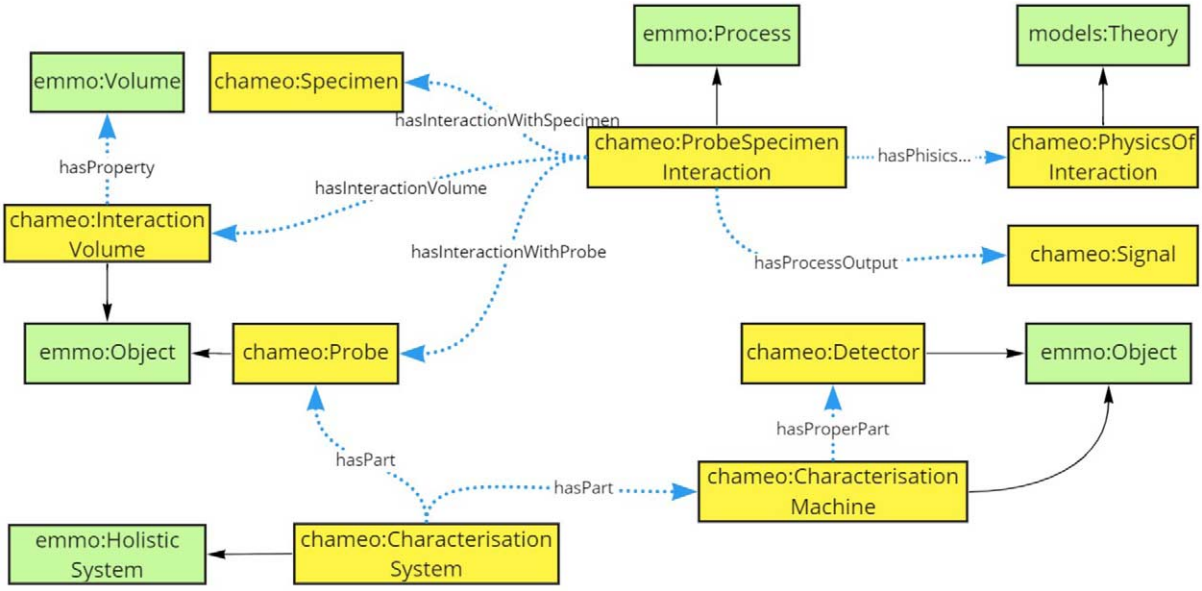


Fig. 5. Characterisation system and interaction with the specimen.

- **chameo:Probe**: is the physical tool (*i.e.* a disturbance, primary solicitation, or a gadget), controlled over time, which generates measurable fields that interact with the specimen to acquire information on specimen's behaviour and properties.
- **chameo:CharacterisationMachine**: is the equipment used for the characterisation, which includes the detector.
- **chameo:Detector**: Physical device used to measure, quantify and store the signal after the interaction with the specimen.
- **chameo:InteractionVolume**: The volume of material, and the surrounding environment, which interacts with the probe and generates a detectable (measurable) signal (information).
- **chameo:ProbeSpecimenInteraction**: The process related to the interaction between the probe and the specimen. Its participants are the probe, the specimen and the interaction volume.
- **chameo:PhiysicsOfInteraction**: Set of physics principles (and associated governing equations) that describe the interaction between the sample and the probe.

The calibration process is illustrated in Fig. 6, whose classes and properties are described as follows.

- **chameo:CalibrationMeasurement**: This stage represents the act of measurement using a reference sample, with the goal to produce reference data that will be used to calculate the final characterisation property. The participants of this process are the specimen, the probe and the detector. This class is subclass of the generic **emmo:Measurement** class. Through the *chameo:hasExecutionDate* datatype property the information about the execution timestamp of each calibration is stored.
- **chameo:RawData**: The raw data is a set of (unprocessed) data that is given directly as output from the detector, usually expressed as a function of time or position, or photon energy. This class is subclass of the emmo:Data. For each raw data the acquisition rate (**chameo:AcquisitionRate**) and the measurement unit (**emmo:ReferenceUnit**) are specified.

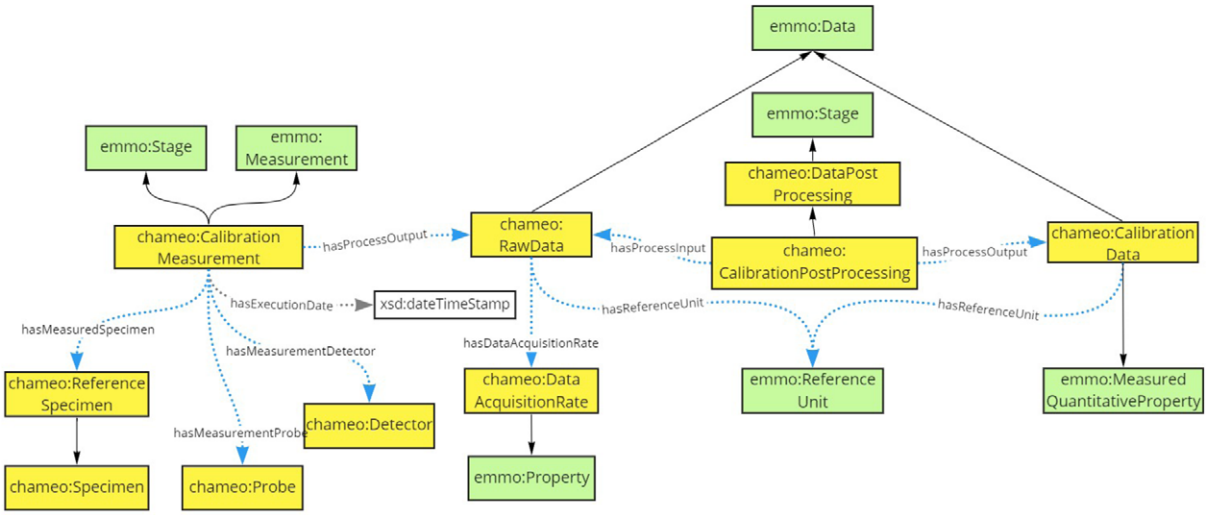


Fig. 6. Calibration process.

- **chameo:CalibrationDataPostProcessing**: In the calibration process the post-processing provides reference data to be used for calculation of the characterisation property. It is subclass of the generic **chameo:DataPostProcessing** class.
- **chameo:CalibrationData**: This is the output of the **chameo:DataPostProcessing** in the calibration process. This class is subclass of the **emmo:MeasuredQuantitativeProperty**, being it the result of interest for the measurement.

The last portion of the ontology is the one related to the actual measurement of the material's characterisation property, as depicted in Fig. 7 and explained as follows.

- **chameo:MeasurementProcess**: This is the process that results in the determination of values using an equipment. It is a subclass of **emmo:Stage**, being it a stage of the characterisation method, and the generic **emmo:Measurement**. The participants of this process are the specimen, the detector and the probe, and its output is raw data.
- **chameo:MeasurementTime**: It quantifies the time needed for the acquisition of the raw data. It is a property of the measurement process.
- **chameo:MeasurementParameter**: This class represents the input parameters that are necessary for acquiring the signal.
- **chameo:CharacterisationEnvironment**: It models the environment of the experiment (*e.g.* temperature, pressure, working environment – in air, controlled pressure, or vacuum – humidity, noise, vibrations). Specific classes deriving from this one are defined for each of the physical quantities used for describing the environment of the experiment.
- **chameo:DataPreparation**: This stage represents all of the activities that are performed for preparing the raw data for the post-processing. Input and output for this type of process is the raw data. Specific subclasses are added for data normalisation **chameo:DataNormalization** and **chameo:DataFiltering**.
- **chameo:MeasurementDataPostProcessing**: Analysis that allows the calculation of the material's property of interest, taking as input the raw data generated by the measurement and the calibration data. It is a subclass of the generic **chameo:DataPostProcessing** class.

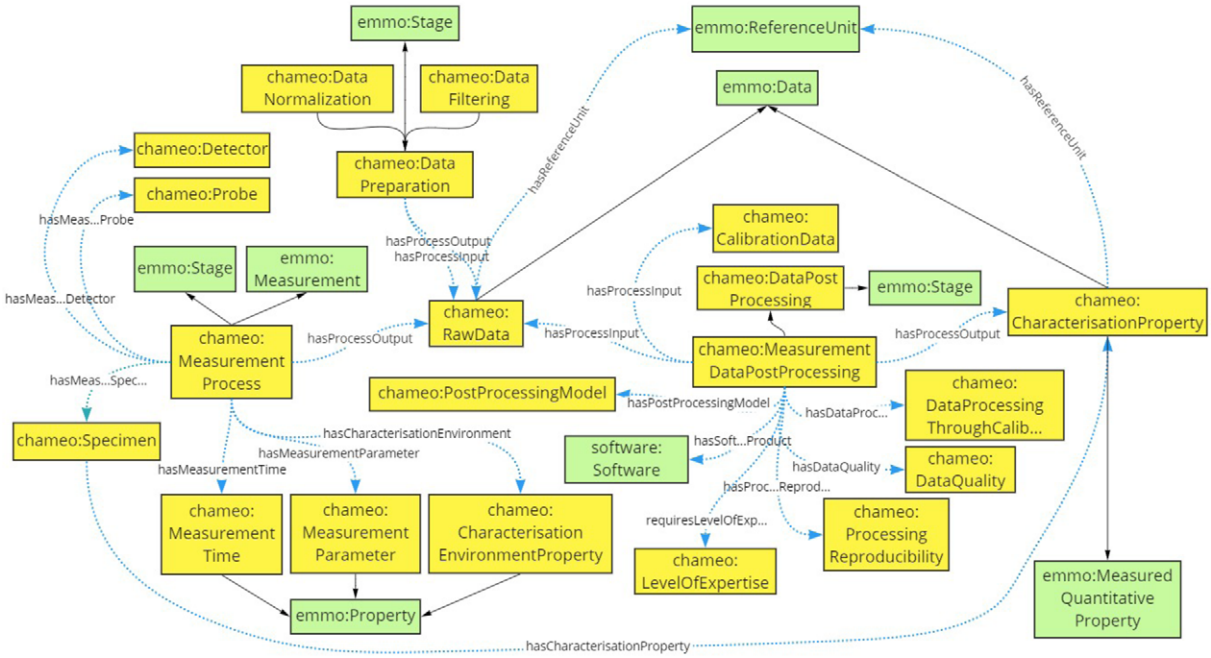


Fig. 7. Measurement and data processing to get the characterisation property.

- **chameo:PostProcessingModel**: This class represents the model used for the post processing (*e.g.* Oliver Pharr for nanoindentation data).
- **chameo:SoftwareProduct**: Via this class it is possible to specify the software used for post-processing.
- **chameo:LevelOfExpertise**: This class represents the level of expertise required for the post-processing.
- **chameo:ProcessingReproducibility**: Via this class it is possible to provide a description of the statistical analysis performed in order to ensure data reproducibility.
- **chameo:DataQuality**: It is used to provide example evaluation of S/N ratio, or other data quality indicators (limits of detection/quantification, statistical analysis of data, data robustness analysis).
- **chameo:DataProcessingThroughCalibrations**: It is used to describe how the data are corrected and/or modified through calibrations.
- **chameo:CharacterisationProperty**: This class models the material property, which is the final output of the characterisation method. The characterisation property is defined as a **emmo:MeasuredQuantitativeProperty** and is assigned to the specimen through ***chameo:hasCharacterisationProperty***.

6. Alignments with other ontologies

Following the best practices in ontology development, knowledge from existing ontologies has been reused when possible for developing the CHAMEO ontology. Besides, the CHAMEO ontology models generic aspects of the characterisation methodology and therefore requires connections with other ontologies/taxonomies for specific use-cases. Alignments and connections between CHAMEO and other

ontologies are detailed in the following subsections. Further ontologies and taxonomies are expected to be linked to CHAMEO in the future.

6.1. OIE ontologies

A set of EMMO-compliant domain ontologies have been developed to be integrated into the Open Innovation Environment (OIE), a web-based platform that enables sharing and exchanging of semantic, ontology-based, information across members of the European Materials Science community. These ontologies are called *OIE ontologies* (GitHub repository available here: https://github.com/emmo-repo/OIE-Ontologies/). They are catalogues and use-cases related to materials, models, manufacturing process, characterisation methods, software systems, and include concepts that have been reused in the CHAMEO ontology, namely:

- Manufacturing (http://emmo.info/emmo/domain/oie/manufacturing): As depicted in Fig. 4, in CHAMEO the manufacturing process of the sample is modeled by the `manufacturing:Manufacturing` class, related to the sample through the `manufacturing:hasManufacturedOutput` property.
- Material (http://emmo.info/emmo/domain/oie/materials): the sample is a subclass of the `materials:Material` class; again, see Fig. 4.
- Model (http://emmo.info/emmo/domain/oie/models): The physics of the interaction between the probe and the specimen is modelled by using the `models:Theory` class, as shown in Fig. 5.
- Software (http://emmo.info/emmo/domain/oie/software): The software used for the data post-processing is represented by the `software:Software` class (see Fig. 7).

6.2. Friend-of-a-friend (FOAF)

The connection with the FOAF ontology (Brickley and Miller, 2014) is made by defining the CHAMEO **User** class as a subclass of `foaf:Agent`. Being the FOAF ontology not compliant with OWL-DL, *i.e.* the decidable OWL profile adopted for CHAMEO, the whole FOAF ontology has not been imported, but the `foaf:Agent`, the only FOAF construct used by CHAMEO, has been simply referenced by using the FOAF namespace. This allows CHAMEO to be kept decidable (compliant with OWL-DL), while at the same time using the necessary concept from the FOAF ontology.

6.3. Dublin core

The Dublin Core ontology (Dublin Core Metadata Initiative (DCMI), 2020) is used to annotate the ontology with the information about its license, in compliance with the FAIR principles (see Section 7). Just like the FOAF ontology, Dublin Core is not compliant with OWL-DL, therefore it is not imported in CHAMEO; instead, a subset of its constructs is referenced.

6.4. DataCite ontology

The DataCite Ontology (Shotton and Peroni, 2020) is used to add scientific references for the validation of the specific characterisation methods or the overall characterisation workflow. More precisely, the class **CharacterisationValidation**, for which a human-readable description can be provided, has a unique identifier, represented by the `datacite:Identifier` class, through the `datacite:hasIdentifier` property. This identifier belongs to a particular scheme, `datacite:ResourceIdentifierScheme` (*e.g.* DOI, ISBN, ISNN).

7. Compliance with the FAIR principles

The FAIR data principles (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al., 2016) are guidelines to support the reusability of digital assets with an emphasis on the machine-readability of data. Ontologies, besides their key role for supporting interoperability, are digital artefacts and as such should follow the FAIR principles, in order to ensure findability, accessibility, reusability and interoperability. The CHAMEO was developed as a FAIR ontology, as detailed below:

- *Findable*: The URI of the ontology is resolvable (http://emmo.info/emmo/domain/chameo/chameo), and is also available via a persistent URL (http://purl.org/chameo/chameo); version IRI, namespace and prefix are correctly specified, and the latter is also available in the prefix.cc (Galway, 2022) public repository. Standard metadata have been used to annotate the ontology and the GitHub repository where the ontology is stored, in order to facilitate the search for this ontology (and those aligned with it). The conventions used for all of the EMMO-compliant ontologies have been followed, including title, description, license information, authors, contributors, aligned ontologies, *etc.*
- *Accessible*: CHAMEO's metadata are retrievable via a standardized, open communication protocol (HTTP) by using the ontology's URI or permanent URL, and as stated in the Abstract is also stored and publicly accessible in a GitHub repository available at: https://github.com/emmo-repo/domain-characterisation-methodology.
- *Interoperable*: CHAMEO has been implemented using standard RDF (W3C, 2014a), RDFS (W3C, 2014b) and OWL (W3C, 2012) formalisms, and uses other vocabularies which in turn follow the FAIR principles (like Dublin Core Metadata Initiative Terms (Dublin Core Metadata Initiative (DCMI), 2020), VANN (Davis, 2005), *etc.*). Furthermore, a number of alignments with other ontologies was also carried out, as detailed in Section 6.
- *Reusable*: CHAMEO's documentation is available in HTML format in the GitHub repository, includes all of the recommended metadata (namespace prefix, version info, contributor, creation date, *etc.*), a number of optional metadata and basic provenance metadata as well. All of the ontological classes have labels (defined via the `skos:prefLabel` construct), and CHAMEO's license is clearly specified, available and resolvable. The CHAMEO ontology is licensed under the Creative Commons (CC) Attribution license international version 4[2] which gives maximum freedom to reuse the work of the licensor. The annotation is done through the *dcterms:license* property and the CC license is available in RDF, in order to let all the information related to the license be machine-understandable. Besides, CHAMEO is a domain-level ontology that is part of EMMO's ecosystem, and reuses and connects with existing domain-level ontologies stored, annotated and available in a similar manner; CHAMEO, in complying with all of the principles stated above, is of course also available for reuse for a variety of purposes.

CHAMEO's compliance with the FAIR principles has also been tested via the experimental FOOPS! tool (Garijo et al., 2021), obtaining a score of 85%. However, it is important to underline that, at the time of this writing, the tool still displays some bugs and inconsistencies when analyzing the ontologies (including the inconsistent detection of the ontologies' persistent URIs and/or URL redirections). As a result, CHAMEO's actual score should be even higher than the one currently returned by the tool.

[2] https://creativecommons.org/licenses/by/4.0/

8. Conclusion

The CHAMEO ontology presented in this work has been developed as a part of a broader initiative towards the harmonisation of different characterisation techniques, aligned with the objectives of the European Characterisation and Modelling Councils, EMCC and EMMC, respectively. The goal is to model generic aspects of the characterisation methodologies, in order to provide a common framework for the development of ontologies related to specific techniques. Taxonomies and catalogues are planned to be further developed to provide specific concepts to be used in order to specialise the CHAMEO classes and properties. CHAMEO is based on EMMO, a top-level ontology conceived to be a reference for modelling knowledge in the area of Materials Science, and is aligned with a number of other domain ontologies. The CHAMEO ontology has been developed within the OYSTER and NanoMECommons projects, and is meant to be exploited in an Open Innovation Environment, a platform for collecting data about characterisation experiments and for sharing knowledge across communities.

Acknowledgements

This work received funding from the European Commission via the Horizon 2020 projects "NanoMECommons" (Grant Agreement n. 952869) and "OYSTER" (Grant Agreement n. 760827).

Appendix. Full list of competency questions

Table 3
Full list of the forty-seven competency questions

ID	Question	Method	Notes	In scope
CQ1	Which kind of material has been tested (metal, ceramic, polymer, ...)?	Nanoindentation	Material class and instances	Yes
CQ2	Which reference sample has been used for system calibration (fused quartz, sapphire, *etc.*)?	Nanoindentation	Types of Sample	Yes
CQ3	Which indenter tip has been used (Vickers, Berkovich, *etc.*)?	Nanoindentation	Types of Probe	Yes
CQ4	How was the sample mounted on the sample holder?	Nanoindentation	Requires a taxonomy of sublcasses under SampleHolder, Sample preparation	Yes
CQ5	How was the sample prepared for testing (mechanical polishing, electropolishing, ...)?	Nanoindentation	Sample preparation	Yes
CQ6	Which is the surface roughness of the sample?	Nanoindentation	Roughness is a Sample property, the actual data is stored outside the ontology	No
CQ7	Which were testing humidity and temperature?	Nanoindentation	Humidity and Temperature are part of the Characterisation environment.	Yes
CQ8	How many measurements were completed?	Nanoindentation	Actual measurement data is stored outside the ontology	No
CQ9	At which locations? (for mapping)	Nanoindentation	Actual measurement data is stored outside the ontology	No

Table 3
(Continued)

ID	Question	Method	Notes	In scope
CQ10	Which loading history has the sample undergone (quasi-static, continuous stiffness measurement, high-speed, ...)?	Nanoindentation	List of measurement types for a certain sample	Yes
CQ11	Which method has been used for data analysis (Oliver-Pharr, *etc.*)?	Nanoindentation	Subclass of Post processing	Yes
CQ12	Have other properties been investigated, apart from Hardness and Modulus?	Nanoindentation	Subclass of properties	Yes
CQ13	Were the tip area function and machine compliance calibrated immediately after testing? If not, when was the last calibration performed?	Nanoindentation	Calibration datatype property executionDate	Yes
CQ14	Have images (micrographs) of the indentation marks been acquired?	Nanoindentation	Images linked to the Measurement?	Yes
CQ15	Who is the person (user) who did the experiment?	Nanoindentation	For GDPR information about the laboratory will be provided, not the specific user.	No
CQ16	What type is the material? (thin film, multilayered, bulk – single phase, bulk – complex, MEMS, nanopatterned)	Nanoindentation	Subclass of Material	Yes
CQ17	What is the method used for the validation of nanoindentation? (correlate individual indents with microstructure: X-ray mapping, AFM, EDS, EBDS)	Nanoindentation	Subclass of CharacterisationMethod	Yes
CQ18	What is the speed of measurement? (is it high-speed? – *i.e.* fast protocol? 1h, high-resolution nanoindentation protocol 1 day).	Nanoindentation	Subclass of the Measurement + Measurement Properties	Yes
CQ19	What is the speed per indent? (*i.e.* 1 indent/second, 1 indent/minute, ...)	Nanoindentation	Actual measurement data is stored outside the ontology	No
CQ20	What is the instrument type?	Nanoindentation	Subclasses of CharacterizatonMachine	Yes
CQ21	What is the (total) thickness of the thin-film (multi-)layer?	Nanoindentation	Sample parts, like the Layer, and their properties can be described in the ontology	Yes
CQ22	How many layers compose the multilayered film? (1, 2, 3, ...)	Nanoindentation	Sample parts, like the Layer, and their properties can be described in the ontology	Yes
CQ23	How many layers is the sample made up of?	FIB-DIC (Focused Ion Beam-Digital Image Correlation)	Sample parts, like the Layer, and their properties can be described in the ontology	Yes
CQ24	What is the thickness of each layer?	FIB-DIC	Sample parts, like the Layer, and their properties can be described in the ontology	Yes
CQ25	Which materials are the layers composed of?	FIB-DIC	Layer is a Material. Different materials are subclasses of Material.	Yes
CQ26	Has the surface been etched to reveal microstructure?	Nanoidentation of multiphases steel	Defined in the specific cases	Yes

Table 3
(Continued)

ID	Question	Method	Notes	In scope
CQ27	Which etchant has been used?	Nanoidentation of multiphases steel	Subclass of Probe	Yes
CQ28	Which reference sample has been used for system calibration (fused quartz, sapphire, Polymeric Material, *etc.*)?	Nanoindentation	Subclass of Material, Sample	Yes
CQ29	Was the sample embedded in a matrix for sample preparation (*e.g.* resin)?	Nanoindentation	Sample preparation	Yes
CQ30	Type of resin?	Nanoindentation	Sample holder type/material	Yes
CQ31	What is the thickness of the sample?	Nanoindentation	Sample dimension	Yes
CQ32	Was the sample previously submitted to a conditioning (temperature, humidity, Duration)?	Nanoindentation	Sample preparation properties	Yes
CQ33	What are the testing parameters: Maximum load, targeted depth, strain rate, drift rate, maximal holding time?	Nanoindentation	Measurement parameters	Yes
CQ34	Which kind of material has been tested (Reflective solid samples, bulk and thin films)?	Spectroscopic Ellipsometry	Subclass of Material	Yes
CQ35	Which reference sample has been used for Monochromator Calibration and optical alignment (Al sample)?	Spectroscopic Ellipsometry	Subclass of Sample	Yes
CQ36	Which reference sample has been used for evaluation of calibration state (bulk c-Si reference sample)?	Spectroscopic Ellipsometry	Subclass of Sample	Yes
CQ37	How was the sample mounted on the ellipsometer stage (Positioning and alignment of the sample adjusting height, tilt, *etc.*)?	Spectroscopic Ellipsometry	Not such level of detail in the ontology as structured information. May be a textual description of the Sample preparation.	No
CQ38	What was the characterisation environment (ambient, inert atmosphere)?	Spectroscopic Ellipsometry	Characterisation environment class and instances	Yes
CQ39	Which was the selected measurement method (Spectroscopic Mono/MWL, Kinetic Mono/MWL)?	Spectroscopic Ellipsometry	Subclass of Measurement	Yes
CQ40	Which were the input parameters for the measurement (AOI, M-A Angle, Spectral range, Spectrum acquisition step, Light integration duration, Accumulation...)?	Spectroscopic Ellipsometry	Measurement parameters	Yes
CQ41	How many measurements were completed (Static measurements, real time)?	Spectroscopic Ellipsometry	Count of the different Measurements in the Experiment	Yes
CQ42	At which locations? (mapping...)	Spectroscopic Ellipsometry	Actual measurement data is stored outside the ontology	No
CQ43	Which were the data analysis procedures used for the post-processing of the raw data (Spectroscopic or kinetic theoretical optical model)?	Spectroscopic Ellipsometry	Subclass of Post processing	Yes
CQ44	Which were the investigated properties? (Thickness, optical properties of materials, information on the optical functions, surface roughness, interface layers...)	Spectroscopic Ellipsometry	Properties that are output of the Method	Yes

Table 3
(Continued)

ID	Question	Method	Notes	In scope
CQ45	Which software used for post-processing (DeltaPsi2, Excel, Origin, *etc.*)?	Spectroscopic Ellipsometry	Software product used for the Post processing	Yes
CQ46	Was the system calibrated immediately after the measurement(s) end? If not, when was the last calibration performed?	Spectroscopic Ellipsometry	Calibration datatype property executionDate	Yes
CQ47	Who is the person (user) who did the experiment?	Spectroscopic Ellipsometry	For GDPR, information about the laboratory will be provided, not the specific user.	No

References

AF (2012). Allotrope Foundation. https://www.allotrope.org/.

American National Standards Institute and International Society of Automation and Instrument Society of America (2010). ANSI-ISA-88.00.01-2010: Batch control part 1: Models and terminology. American national standard. ISA, https://books.google.it/books?id=jbK0nQAACAAJ.

Arosio, G., Bagnara, G., Capuano, N., Fersini, E. & Toti, D. (2013). Ontology-driven data acquisition: Intelligent support to legal ODR systems. In *Frontiers in Artificial Intelligence and Applications* (Vol. *259*, pp. 25–28). doi:10.3233/978-1-61499-359-9-25.

Berman, F. & Crosas, M. (2020). The research data alliance: Benefits and challenges of building a community organization. *Harvard Data Science Review*, *2*. doi:10.1162/99608f92.5e126552.

BFO Discussion Group (2002). Basic formal ontology. https://basic-formal-ontology.org/.

Brickley, D. & Miller, L. (2014). Friend-of-a-friend. http://xmlns.com/foaf/spec/.

Capuano, N. & Toti, D. (2019). Experimentation of a smart learning system for law based on knowledge discovery and cognitive computing. *Computers in Human Behavior*, *92*(C), 459–467. doi:10.1016/j.chb.2018.03.034.

CEN (2021). Materials characterisation – terminology, metadata and classification. https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/ICT/cwa17815.pdf.

CODATA (2020). The Committee on Data of the International Science Council. https://codata.org.

Davis, I. (2005). VANN: A vocabulary for annotating vocabulary descriptions. http://purl.org/vocab/vann/.

Dublin Core Metadata Initiative (DCMI) (2020). DCMI metadata terms. http://purl.org/dc/terms/.

EMCC (2016–2022). The European Materials Characterisation Council. http://characterisation.eu/.

EMMC (2014–2022). The European Materials Modelling Council. https://emmc.eu/.

EMMC-CSA (2016–2019). The European Materials Modelling Council, Grant agreement n. 723867. https://cordis.europa.eu/project/id/723867.

FAIR-DI (2014–2022). The FAIR data infrastructure for physics, chemistry, materials science. https://www.fair-di.eu.

Fernandez-Lopez, M., Gomez-Perez, A. & Juristo, N. (1997). METHONTOLOGY: From ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium*, Stanford, USA (pp. 33–40).

Galway, D. (2022). Namespace lookup for RDF developers. http://prefix.cc.

Garijo, D., Corcho, O. & Poveda-Villalón, M. (2021). FOOPS!: An ontology pitfall scanner for the FAIR principles, 2980.

ISTC-CNR Laboratory for Applied Ontology (2002). Descriptive ontology for linguistic and cognitive engineering. http://www.loa.istc.cnr.it/dolce/overview.html.

IUCr (1947–2014). International Union of Crystallography. https://www.iucr.org/iucr.

MaRDA (2020). The Materials Research Data Alliance. https://www.marda-alliance.org.

NanoMECommons (2021–2025). Harmonisation of EU-wide nanomechanics protocols and relevant data exchange procedures, across representative cases; standardisation, interoperability, data workflow, Grant agreement n. 952869. https://cordis.europa.eu/project/id/952869.

Nostro, P.D., Orciuoli, F., Paolozzi, S., Ritrovato, P. & Toti, D. (2013). ARISTOTELE: A semantic-driven platform for enterprise management (pp. 44–49). doi:10.1109/WAINA.2013.43.

Ogden, C. & Richards, I. (1923). *The Meaning of Meaning – a Study of the Influence of Language upon Thought and of the Science of Symbolism*. Routledge.

OYSTER (2017–2022). Open characterisation and modelling environment to drive innovation in advanced nano-architected and bio-inspired hard/soft interfaces, Grant agreement n. 760827. https://cordis.europa.eu/project/id/760827.

PA (2007). Pistoia Alliance. https://www.pistoiaalliance.org/.

Romanos, N., Kalogerini, M., Koumoulos, E.P., Morozinis, A.K., Sebastiani, M. & Charitidis, C.A. (2019). Innovative data management in advanced characterization: Implications for materials design. *Materials Today Communications*, *20*. doi:10.1016/j.mtcomm.2019.100541.

Shotton, D. & Peroni, S. (2020). The DataCite ontology. http://purl.org/spar/datacite.

Suarez-Figueroa, M.C. & Gomez-Perez, A. (2009). NeOn methodology for building ontology networks: A scenario-based methodology. In S..S.T.S..I.C. on SOFTWARE (Ed.), *Proceedings of the International Conference on Software, Services & Semantic Technologies (S3T 2009)*.

The EMMC Consortium (2014–2021). The European materials modeling ontology. https://emmc.info/emmo-info/.

W3C (2012). Web ontology language, https://www.w3.org/OWL/.

W3C (2014a). RDF resource description framework, http://www.w3.org/RDF/.

W3C (2014b). RDF schema, http://www.w3.org/TR/rdf-schema/.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific data*, *3*. doi:10.1038/sdata.2016.18.