

Review

Materials informatics for mechanical deformation: A review of applications and challenges

K. Frydrych¹, K. Karimi¹, M. Pecelerowicz¹, R. Alvarez-Donado¹, F. J. Domínguez-Gutiérrez^{1,2}, F. Rovaris¹, S. Papanikolaou^{*1} 

¹ NOMATEN Centre of Excellence, National Centre for Nuclear Research, ul. A. Sołtana 7, 05-400 Swierk/Otwock, Poland

² Institute for Advanced Computational Science, Stony Brook University, Stony Brook, NY 11749, USA

* Correspondence: stefanos.papanikolaou@ncbj.gov.pl; +48-22-273-2507 (PL)

Abstract: In the design and development of novel materials that have excellent mechanical properties, classification and regression methods have been diversely used across mechanical deformation simulations or experiments. The use of materials informatics methods on large data that originate in experiments or/and multiscale modeling simulations may accelerate materials discovery or develop new understanding of materials' behavior. In this fast-growing field, we focus on reviewing advances at the intersection of data science with mechanical deformation simulations and experiments, with a particular focus on studies of metals and alloys. We discuss examples of applications, as well as identify challenges and prospects.

Keywords: Metal alloys; Machine learning; Informatics; Defects; Dislocations; Mechanical deformation; Data science; Ontology

1. Introduction

Materials informatics (MI) is an interdisciplinary field of study at the interface of materials engineering and data science which aims to boost discovery, characterization, component design, synthesis, and screening of new or alternative materials based upon modern state-of-the-art machine learning (ML) techniques [1–3]. The Materials Research Society bulletin provides an exhaustive list across almost all aspects of materials science and engineering that are greatly influenced by this relatively new field including (but not limited to) thermoelectric materials, superconductors, novel bio-materials, batteries, and fuel cells, dielectric elastomers and so on [4–6]. An emerging category pertains to the ongoing research of deformation and failure in metal alloys owing to an ever-increasing demand for optimal combinations of material lightweightness, strength, ductility, corrosion-resistance, and toughness as essential infrastructural and industrial components [7]. In this framework, ML has been proven to be an efficient and powerful computational tool in micro-structural tailoring and characterization [8–10], with robust predictions of fundamental thermomechanical and physical properties [11–13], and systematic development of improved atomic potentials based on high-throughput materials computations [14–17].

Materials data scientists make routine use of advanced statistical learning elements (i.e. regression, classification, regularization, dimensionality reduction, and cross-validation [18]) for insights into inherent patterns, underlying physics, and structure-property correlations across a broad range of lengthscales and timescales. It is indeed a fact that only very few of these features can be systematically explored solely based on theory, experiments, or computations, and thus, prediction of complex materials behavior has been limited. The main challenge concerns the high-dimensional and scattered nature of the parameter space associated with multi-component materials making the design problem intractable based on conventional materials science approaches [19].

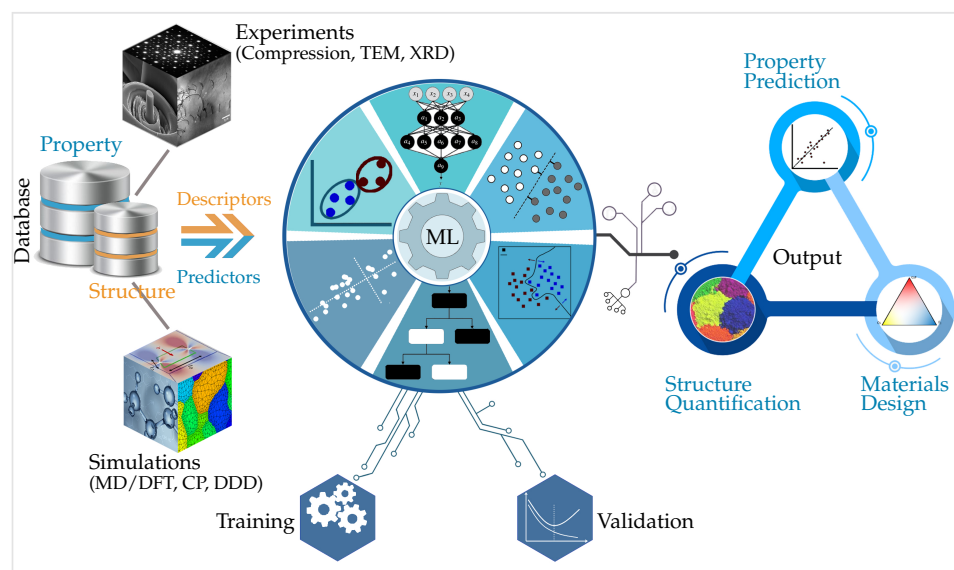


Figure 1. A typical materials informatics workflow for applications related to mechanical deformation. Databases (left) generated through microscopy and (nano)mechanics experiments are combined with multiscale modeling simulations to provide relevant descriptors and predictors. The latter are used to train and validate ML algorithms (center) leading to the property prediction, structural characterization, and materials design/discovery (right).

High entropy alloys, for instance, are formed by mixing multiple principal components with a tendency to form unary, binary, or more complex phases that may also undergo phase transformations depending on additional processing [20]. A massive combinatorial phase space exists, even though it has been only partially investigated through experiments and/or simulations. ML provides possible frameworks to integrate search strategies for novel materials with tailored properties in a timely yet robust and efficient manner which is otherwise impossible using traditional frameworks.

As schematically illustrated in Fig. 1, the implementation of an ML-based approach involves two elementary steps: pattern detection and predictions, both essential in understanding fundamental material properties [21]. The former demands continuous access to sufficient amounts of data which are essential for any quantitative assessment of materials behavior. The sources of data are typically diverse including those (re)produced from high throughput (*ab initio*, atomistic, and continuum-based) simulations, and extensive experimentation and/or those collected from existing materials databases [22–24]. In the informatic study of structure-property relationships, a relevant material dataset contains a *feature* space normally spanned by a set of atomic and/or local (micro-)structural *descriptors* that presumably capture salient structural details for an existing set of materials [25]. This structural information may cover a broad range of attributes including element-wise atomic properties [11,26–28], atomic fraction in multicomponent materials [12,29,30], and topological measures such as coordination number, packing efficiency, and local ordering [25,31], lattice distortion [11,32], and local crystallography [33] in metallic systems. Alternatively, the input data might be composed of a set of microstructural images fed into a “deep” learning model that can learn about embedded features directly from the image-based information (see Sec. 2 and references therein). In an unsupervised learning context, ML users typically seek for analyzing underlying trends, clustering features, and multivariate correlations associated with the input features which are presumably tied to distinct physical processes governing certain (bulk) material properties. Supervised learning methods, on the other hand, additionally incorporate a set of *predictors*, commonly attributed to fundamental material constants or properties, which are potentially influenced by descriptors but without any *a priori* knowledge about the degree of inter-correlations. It should be noted

that, depending on the context, a meaningful combination of local and global attributes may serve as the features dataset in order to make predictions and optimizations on a different set of bulk physical properties as the target data. In studying metallic alloys, for instance, Vickers hardness relates (intuitively!) to intrinsic inter-atomic interactions but also strongly correlates with the metals elasticity, yielding, and toughness [34,35]. The predictive aspect of ML assists in establishing a quantitative structure-property association with no prior assumptions and purely based on the existing materials data, playing a key role in the study of novel engineering materials.

The present overview revisits the growing role of informatics in several subcategories pertaining to materials deformation and failure. There are quite a few existing reviews in the scientific literature with a somewhat similar overarching theme—that is to say, applications of data analytics in the solid state materials science, novel materials discovery, design, and associated property predictions [19,21], data analytics and image processing for microstructural identification and physics inference [2] as well as data-centric science relevant to thermodynamics/mechanics in different families of alloyed metals (including metallic glasses and high entropy alloys) [36] and other structural (continuum-based) materials [37,38] or bulk glassy systems [39], and learning methods on searching for property-related descriptors [19,40]. Our motivation is somewhat different; we aim to address a niche question in this field as to how one could efficiently train the machine to learn from certain (often linear) properties through materials “precursors” or “footprints”, and use the acquired knowledge to forecast unknown functionalities beyond linear elastic limits [41]. Upon successful training, this leads to a functionality-based materials discovery as an essence of the inverse design framework which aims to connect properties to underlying mechanisms related to deformation and failure in alloyed metals [42]. A special focus will be made on the extraction of relevant physics from simulation-based and experimentally-derived (image) data integrated with data analytic approaches. Owing to the multi-scale notion of deformation and flow in metals and alloys, the numerical frameworks discussed here range from first principle *ab initio* calculations and atomistic (Molecular Dynamics and Monte Carlo) simulations to coarse-grained mesoscale frameworks (Discrete Dislocation Dynamics) and phenomenological continuum models, all well-established techniques within the computational materials science community. The experiment-based contributions will be mainly addressed from the feature-identification point of view where high-resolution electron microscopy, X-ray scattering, and digital image correlation outputs highly demand efficient use of deep learning approaches to unravel embedded topological irregularities (e.g. dislocations in crystals, grain boundaries in polycrystals, quenched disorder in metallic glasses, or micro-cracking in brittle solids). Further discussions will follow on how the outcomes of such an analysis may improve existing lattice-based and continuum models by incorporating key micro-mechanical/structural ingredients (e.g. defect density, dislocation interactions).

The organization of the paper is as follows. Section 2 begins with a description of common image-based characterization techniques combined with advanced machine learning tools to explore relevant microstructural features in metals and alloys. We highlight, in particular, applications of deep learning in materials classification, phase identification, topological characterization, and physics extraction in both supervised and unsupervised contexts. In Section 3, we turn to mechanical deformation experiments and highlight the machine learning capabilities in identifying prominent materials features that are intrinsically linked to desired physical and mechanical properties, within structure-property frameworks. Focusing on crystalline alloys, in Section 4, we move on with a discussion of dislocation plasticity on the basis of a combined discrete dislocation dynamics and machine learning framework. We further present several ML-based illustrations of property derivation from a set of high-resolution strain images that contain important information about dynamical mechanisms of dislocation ensembles. Validation of dislocation-related mechanisms may only be performed at the

nanoscale, so Section 5 considers the application of materials data science in the in-situ characterization and classification of nanomechanics experiments. Section 6 explores the current state of the art in using machine learning in studies of phase stability and thermodynamics properties related to materials discovery and high-throughput metal-alloy design, and a brief account of the key developments for ML interatomic potentials in atomistic simulations. Finally, Section 7 provides a brief discussion on different aspects of the materials information ontology relevant to the data-driven research of materials deformation as well as a broad overview of associated impacts, existing challenges, and prospects.

2. Materials informatics in microstructural image classification

A common approach towards materials informatics has been to use large imaging datasets to obtain valuable information about a given material. Modern computational techniques such as ML enable the analysis of data coming from various imaging techniques. Such techniques have a potential for providing faster (by orders of magnitude!), more reliable, and repeatable image analysis than traditionally applied human-eye inspections. The images can be prepared using various experimental techniques, and in the context of mechanical deformation, these include optical microscopy, digital image correlation (DIC) for the analysis of strain information, scanning electron microscopy (SEM) and electron backscatter diffractometry (EBSD). Characteristic applications, in this context, include the automatic identification of key defected regions that may compromise mechanical properties, and an example of a characteristic challenging prospect is the classification of irradiation-induced defects.

2.1. Microscopy and analysis of images

Early applications of informatics methods on microstructure imaging emerged in the context of optical microscopy (OM). Using random forest (RF) algorithms [43,44] and unsupervised learning approaches [45], the steel microconstituent segmentation based on OM images was achieved. Beyond OM, and in the context of mechanical properties, DIC has emerged as a promising technique. DIC uses specimen surface images taken in the initial and deformed configuration to produce displacement and strain fields. Although DIC is a computationally intensive technique, some recent progress suggests that its speed can be increased by the application of convolutional neural networks (CNNs) [46], and therefore its potential in producing big data sets will fully develop. Recently, it has been noticed [47,48] that DIC results contain very rich information related to plastic deformation gradients, but only a part of it is utilized. Through the use of unsupervised ML approaches [47,49] it is possible to gain insights into the history of deformation of the crystalline sample, as well as consistently predict mechanical properties such as the yield stress, without the use of stress information. Furthermore, ML methods have been applied to determine the fracture toughness of composite materials based on DIC results, typically using artificial neural networks (ANNs) [41,50,51]. Crack detection, measurement, and characterization based on DIC can be performed using image processing methods [52] and fatigue crack detection in DIC images may be automatically performed using CNNs [53].

At the nanoscale, transmission electron microscopy (TEM), SEM, and EBSD can be used to prepare image libraries for training ML frameworks. An important application of ML trained on SEM images is the steel microstructure classification. In Refs. [8,9], CNNs were applied to classify microconstituents of ultrahigh carbon steels (UHCS). This is a characteristic application that can be extended in a large collection of material classes [41]. Fig. 2 presents the t-distributed stochastic neighborhood embedding (t-SNE) map for the UHCS data set and the constituents were classified based on SEM images. In Refs. [54,55] the fully convolutional neural network (FCNN) accompanied by a max-voting scheme was applied to classify microconstituents of steel. In order to achieve the goal, fine-tuning and data augmentation were applied. The ML framework was trained

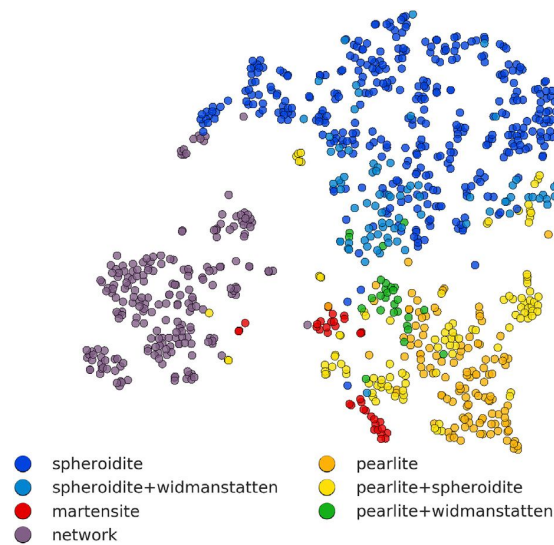


Figure 2. The map of t-distributed stochastic neighborhood embedding (t-SNE) for the entire ultra-high carbon steel dataset. Spheroidite, pearlite, etc. are various steel microstructure constituents that are important from the point of view of steel properties and were traditionally classified by human experts. The figure is reproduced after [8] with permissions from Elsevier.

on both SEM and light optical microscopy images. Beyond neural networks, in [56], the classification of steel microstructures can be performed by the application of support vector machines (SVM), or RF and gradient boosting machine.

It is possible to combine DIC with an SEM image collection to perform high-resolution digital image correlation (HR-DIC). In this way, it is not only possible to classify material microstructures, but also to investigate novel deformation mechanisms. For example, the identification of deformation twins in HR-DIC of Mg alloys was undertaken in [57,58].

The use of SEM technology has also led to the emergence of the EBSD technique as an important data-rich experimental method commonly used to investigate the microstructure of polycrystalline materials. An application of ML methods in this context is the segmentation and classification of patterns such as grains and grain boundaries. Segmentation methods may directly incorporate the detection of anomalies and ML methods can be applied to increase the speed of indexing and classification of EBSD patterns as in [59–62]. Given that EBSD connects to lattice misorientations in crystals, another application of ML methods includes physics extraction based on the acquired EBSD data. For example, ML methods can be used [63] to find correlations between microstructural parameters and twinning, by using decision trees, ultimately identifying the influence of various crystallographic and morphological attributes on twin nucleation and propagation in AZ31 Mg alloys. By connecting to dislocation physics [64], it is also possible to apply deep learning to estimate the total dislocation density based on a geometrically necessary dislocation (GND) density. The method was applied to EBSD images of α -Fe.

At the atomic scale, TEM and high resolution transmission electron microscopy (HRTEM) are efficient experimental techniques revealing detailed microstructural information. However, as the analysis of TEM and HRTEM images is a highly tedious and labor-consuming task, it is an open challenge to perform materials informatics to automate TEM image classification. Nevertheless, in [65], deep convolutional neural networks trained on simulated data were only able to interpret HRTEM images and the analysis of nanoparticles observed in TEM [66] based on an evolutionary algorithm (EA) may be applied, while the performance of various CNNs for the task of TEM image segmentation has been evaluated in [67]. However, an important problem is the point defect detection in TEM images. To resolve this challenge, a possible approach is to use

supervised ML that requires simulation-generated data with appropriate labels [65,68]. It is also possible to consider, as in [69], the problem of defect detection as an anomaly detection one: in such a case, self-supervised ML could be only trained on defect-free experimental images. When there is no imaging noise, applying principal component analysis (PCA) can be shown to be enough to locate defects, but CNNs may improve the model performance in the presence of imaging noise that may also originate in quenched disorder [41,42,47,48,70]. Another way to circumvent the noise issues in TEM is by using scanning transmission electron microscopy (STEM) [68] where the size and quality of STEM datasets have been increasing exponentially. For example, artificial neural networks were applied [71] towards the detection of twins in STEM images [72,73].

2.2. A challenging application: Imaging classification of irradiation-induced defects

Irradiation in contemporary and future nuclear fission and fusion power plants inherently leads to changes of the material microstructure[74], in ways that are highly complex in terms of predictive abilities and combination of processing routes, posing an immense challenge in the inference of irradiated structure-property relationships. In particular, there is a highly complicated nature and variability of defects across scales, irradiation dose, and temperature, as well as the nature of the irradiation beam of incoming particles (alpha, beta, ions).

Despite the apparent complexity, MI methods can facilitate defect classification. A characteristic example was shown through the use of STEM images of irradiated steels that were analyzed using ML approaches [75], in particular CNNs, leading to the identification of irradiation-induced defects such as dislocation loops. Furthermore, it has been possible to apply ML to detect He bubbles in TEM images of Ni-based alloys irradiated with He ions at 650°C. The comparison of manually and automatically detected He bubbles is shown in Fig. 3. Finally, an automated system for analyzing in-situ TEM videos was presented in Ref. [76], where a deep-learning tool called YOLO (you only look once) can be used to analyze dislocation loops present in FeCrAl alloys irradiated to various damage doses.

3. Informatics in deformation experiments & simulations

While detecting and classifying defects can lead to insightful conclusions, their behavior *in-situ* as the sample is mechanically loaded, provides a unique view in the structure-property relationships that become accessible through materials informatics. Indeed, for material design, a combination of optimal ductility and strength requires

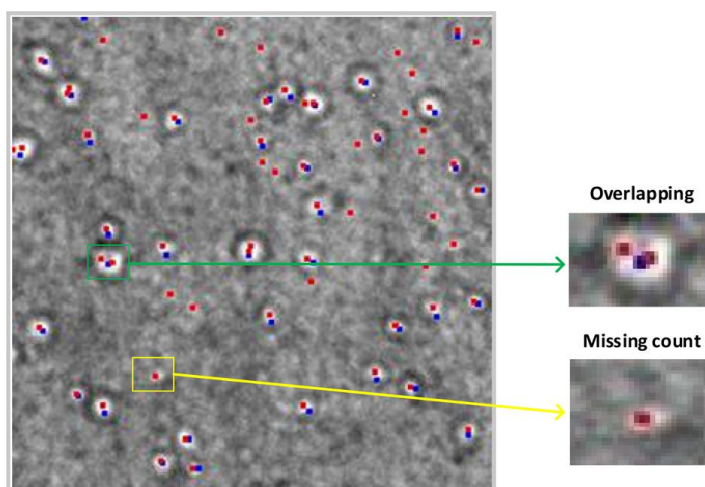


Figure 3. The location of He bubbles in an TEM image of an irradiated Ni-based alloy marked manually by TEM analysis experts (blue) and using the ML approach (red). Reproduced after [77] based on Creative Commons CC BY license.

prior knowledge about the nucleation and growth of relevant microstructural phases as well as their consequences in terms of macroscopic properties of interest, a training task which seems quite amenable to ML.

3.1. Searching for microstructural features and machine learning

Bulk elasticity of metals, as a main functional property, is largely controlled by electronic structures, chemical bonding, atomic/molecular arrangements, preparation history, and thermal treatments, forming an intractable design space from a materials engineering point of view. The materials science literature provides numerous instances where an appropriate combination of surrogate data-driven models and deformation-based experiments led to robust and meaningful predictions of key functional properties out of a tremendous (seemingly intractable!) features space. Earlier attempts in this context originated in the use of data analytics (neural networks in particular) in flow stress prediction of metallic materials, aluminum alloys [78,79], steel [80,81], Ti alloy [82] under varying temperature, deformation rates, and applied strains. Neural network training was also practiced in terms of the alloy composition search that, along with heat treatment parameters, lead to accurate evaluations of yield strength, hardness, ductility, elasticity, fatigue properties, and fracture toughness [83–89]. Furthermore, in [11], an extensive neutron-diffraction characterization was carried out to obtain elastic constants associated with an $\text{Al}_{0.3}\text{CoCrFeNi}$ alloy under uniaxial tension. The empirical measurements along with an appropriate set of relevant elemental parameters (i.e. group number, cohesive energy, density, atomic radii) were used to feed a gradient-boosted trees algorithm. More generally, intensive research effort was undertaken in the field of amorphous metals. For example in Ref.[28], a set of dynamically-measured elastic moduli based on ultrasonic-wave-propagation tests led to an SVM regressor that was used to relate bulk and shear moduli to a set of training features with the atomic electronegativity, atomic volume, and atomic size differences selected as best-performance descriptors.

However, beyond elasticity, the ability for property prediction is limited by the complexity of plastic deformation mechanisms. For example, in the heavily studied amorphous alloys, it has been known for many years now that local topological structures must have strong bearings on the physics and mechanics. There has been a growing body of research attempting to establish robust microstructural origins associated with common glassy features including slow relaxation dynamics, aging phenomenon, and dynamic heterogeneity (see [90] and references therein). The main focus has been on the prevalence of certain kinds of local ordering, purely based on (static) structural information, that will strongly correlate with dynamical properties. Commonly used structural quantities include (but are not limited to) the coordination number, local potential energy, free volume [91], radial distribution function¹ [31], close packing and local polyhedral order [91,92], or order parameters describing bond-orientational order in condensed phases [93]. It has been further argued that the predictive capacity, i.e. degree of correlations with dynamical quantities, associated with each individual structural descriptor (or in combination together) could significantly vary across a wide range of supercooled liquids and glasses [90,94]. Similar connections have been also established in the context of mechanically driven glasses where the local structure is suggested to provide strong evidence for bulk mechanical properties [95]. In metallic glasses, for instance, local atomic configurations are believed to strongly correlate with elastic heterogeneities [96,97]. Atomistic simulations of [98,99] revealed the emergence of local structural motifs that imprint soft environments susceptible to irreversible shear-induced rearrangements and, therefore, improve the plasticity of metallic glasses. This has been contrasted with “short-range ordering” in an amorphous matrix with insignificant contributions to the deformation mechanism but instead playing an important role in slowing down the

¹ The radial distribution function describes the distribution of pair-wise distances of particles within a specified volume.

dynamics of supercooled liquids [95]. In [100], an abundance of the former structural patterns was linked to the enhanced deformability in rejuvenated glass structures which is at odds with aging samples, generally rich in terms of the latter local features.

Characteristically, in the context of amorphous metals, pattern detection in the research of glassy matter is generally regarded as a supervised learning process where appropriate machine learning classifiers (and/or regressors) are trained based on a large set of atom-wise dynamical measures (atomic trajectories, mobility, or rearrangements) and corresponding local structural information (e.g. average pair correlation function, coordination number, bond orientation, and so on) [101]. These are referred to as *responses* and *features* in the modern language of machine learning. The concept was largely put forward in a series of important papers applying machine learning to describe the interplay between structure and dynamics in several glass formers as well as polycrystals [33]. The former includes supercooled liquids and glassy solids [91,94,102,103] and sheared amorphous solids [91,104–106]. In this framework, the machine learning model successfully learned from training datasets to distinguish a population of rearranging atoms from that of local frozen regions in a typical classification context. The developed methodology was further validated by its predictive capabilities in terms of the subsequent atomic mobility solely based on the existing local topology. The observed binary separation led to defining a local metric called “softness” by measuring a (signed) distance of each input feature from the predicted decision boundary within the multi-dimensional feature space. Physically speaking, the softness field should entail the atoms’ propensity to undergo localized rearrangements (as a dynamical quantity) based upon a fully structural notion (see Fig. 4(a,b)). The authors further hypothesized and validated (numerically) that the softness probability has an Arrhenius-type dependence on the temperature and were able to measure local energy barrier height accordingly. In the context of metallic glasses [31] a local dynamical metric called “flexibility volume” was defined based on the product of the vibrational mean squared displacement and average interatomic distance which is similar in essence to the notion of softness. The authors used the former as a target set along with an input structural metric associated with the radial distribution function of individual atoms to train a support vector regression. The regression analysis lead to a robust prediction of the structural flexibility based on local atomic environments and was shown to strongly correlate with a broad range of elasto-plastic properties. A related study [107] addressed the structure–dynamics relationship in metallic glasses linking the “interstice” distribution and “quench-in” softness with the latter metric being the gradient-boosted decision tree-based probability of plastic rearrangements.

In addition, several unsupervised learning frameworks have been developed where there is no outcome measure (dynamical properties in our context), and the goal is to describe the associations and patterns among a set of local structural measures [91,93] which further investigated strong correlations between the outcomes of the unsupervised clustering scheme and the dynamics in classical model glass formers, as displayed in Fig. 4(c,d).

3.2. Materials informatics and in-situ loading: DIC and surrogate models based on plasticity

In the context of crystalline deformation and beyond elasticity in polycrystalline metals and alloys, the primary focus has been the use of DIC methods and then, the associated pursuit of plasticity models that reproduce strain behaviors. On the microstructural level, several research papers have been devoted to coupling HR-DIC with plasticity: in [108], various materials (BCC tantalum, FCC Nickel, Ti-6Al-4V) have been investigated using a combination of crystal plasticity and finite element methods (CPFEM) and three experimental techniques, namely EBSD, HR-DIC and surface profilometry. The strain fields predicted by CPFEM were directly compared with HR-DIC measurements and very good agreements were obtained [109–111]. Another possibility of the same kind would be the identification of local elastic strains and lattice rotations by

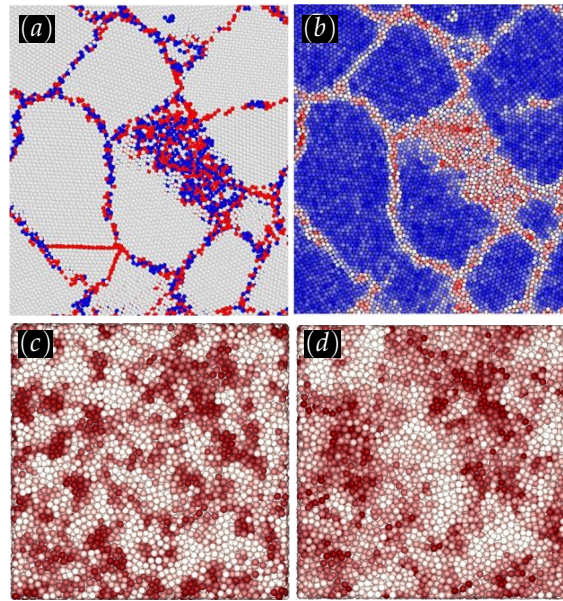


Figure 4. Linking microstructure with dynamics **a)** Distinct microstructures (bulk grains versus grain boundaries) illustrated based on the Voronoi topology analysis of a simulated polycrystalline aluminum. The white, red, and blue colors denote the FCC phase, hexagonal close-packed (HCP) structure, and atoms with neither the former nor the latter structure, respectively. **b)** The same snapshot displaying the associated softness (reproduced from [33]). Atoms with a high softness are red and those with a low softness are blue. **c)** Binary microstructural classification of particles in a model glass former colored according to the probability of being a member of the corresponding cluster. Dark red and white denote high and low probabilities. **d)** The same particles but color-coded in accordance with the associated mobility. Dark red and white indicate high and low mobility (reproduced from [93] licensed under CC BY 4.0).

applying the HR-EBSD variant. CPFEM and HR-EBSD were applied to investigate the heterogeneous elastic strains near the impurity-originated inclusion in nickel superalloy in [112,113].

In this context, an interesting materials informatics application was presented in [114,115]. The approach is based on coupling experimental techniques, crystal plasticity (CP) simulations and Bayesian networks to investigate the short crack growth in beta-metastable BCC titanium alloy (VST-55531) subjected to high cycle fatigue. The applied experimental technique was a combination of diffraction contrast tomography, multiple phase contrast tomography scans, and modern reconstruction and segmentation techniques. Tree augmented Naive Bayes was applied to extract correlations between postulated short crack driving force metrics and experimental results. Another interesting application is the investigation of the cyclic stress-strain curves of various dual-phase steels differing in pearlite-to-ferrite proportion as predicted by Miyazawa et al. [116]. First, the microstructural analyses were conducted using EBSD in order to extract the grain morphology and determine the location of both phases. Then, the stress-strain curves obtained from low-cycle fatigue experiments were utilized to obtain the parameters of the macroscopic J2 plasticity model of pearlite phase and crystal plasticity model of ferrite. Then, the 2D FEM simulations were conducted in order to provide the dataset for ML. Two ML frameworks, namely the linear regression and neural network, were applied to dimensionally reduced datasets thus enabling to provide cyclic stress-strain predictions. Another example was presented in [117] where the artificial neural network was trained using data coming from the rate-dependent CPFEM modelling in order to predict mechanical response and texture evolution. The approach was applied to AA6063-T6 alloy for tension and simple shear of both single and polycrystals. The 2D CPFEM model was built based on EBSD results. The feed-forward backpropagation

ANN was used. The strain and crystallographic orientation have been applied as an input vector. The output consisted of stress and updated orientation. Interestingly, it was demonstrated that the trained ML lead to correct predictions outside the bounding box as well. First, it predicted correct stress-strain curves and the texture evolution at larger strain under simple shear. Second, it correctly predicted the stress-strain curves after changing the strain path. Note that, as the CPFEM results were demonstrated to be in good agreement with the experimental results, the ANN approach automatically provides predictions consistent with experimental data.

A key direction, in the domain of building a reduced order model is the application of a materials knowledge systems (MKS) approach. The aim of this formulation as developed by Kalidindi and coworkers [118,119] was to optimize the microstructure of the material to meet the required properties. In [120] the MKS framework was extended to polycrystalline microstructures. Namely, CPFEM simulations for various artificial microstructures of α -Ti were performed. The MKS was then calibrated so that the elastic stiffness and yield strength under uniaxial tension obtained with PCA as compared to CPFEM results would provide the desired accuracy. This way it is possible to obtain these two mechanical properties for a given texture much faster as compared to CPFEM approach. Note, however, that the new texture should lie between the bounds of the calibration set and that the calibration would have to be repeated if one is interested in different boundary conditions or different mechanical properties. A similar approach was then applied to evaluate the fatigue performance first in the high cycle fatigue [121,122] and then in the transition fatigue regime [123].

4. Learning from crystal defects: Dislocation ensembles

Dislocations and their dynamics have been a key tool towards understanding polycrystalline deformation in metals and alloys, so it is natural to pursue MI applications in this context. Discrete Dislocation Dynamics (DDD) has been one of the major tools used for the numerical simulation of mechanical properties of materials [124]. The wide adoption of DDD in this field is motivated by the intrinsic multi-scale nature of crystal plasticity: dislocations are defects at the atomic scale but their collective movement and interaction determine the mechanical properties of a material at the macroscopic scale. DDD approaches retain the discrete definition of the dislocation lines while considering them as inserted in an elastic continuum. DDD relies on a spatial discretization of the dislocation lines, generally achieved by considering small straight segments [125] whose motion is dictated by the local stress field acting at their position through the Peach and Kohler expression [126]. DDD stands in between atomistic and pure continuum approaches in terms of achievable time/length scales, offering the possibility to directly compare simulations with laboratory conditions. The application of DDD approaches

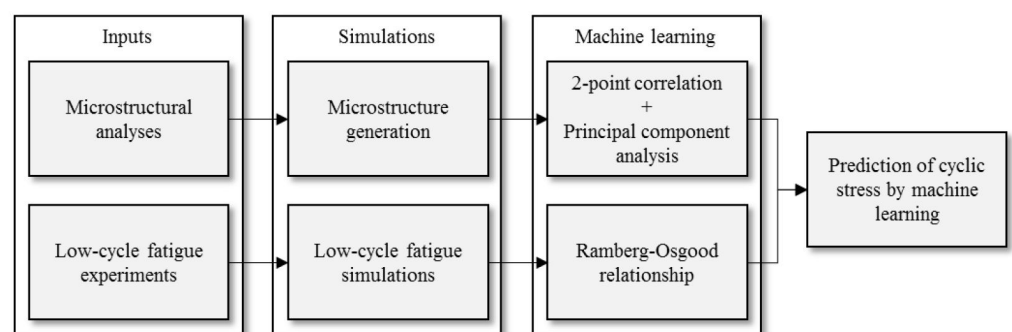


Figure 5. Framework applied by Miyazawa et al. [116] in order to predict cyclic stress-strain property of dual-phase steels. The approach combines EBSD, mechanical experiments, FEM simulations with macroscopic and crystal plasticity, 2-point correlations, principal component analysis and Ramberg-Osgood relationship with ML model. Reproduced after [116] based on Creative Commons Attribution License.

in the past three decades has provided major insights into the deformation behavior of materials [70,127,128], elucidating fundamental mechanisms underlying material deformation: size effects in plasticity [129–135], the role of dislocation junctions in material hardening [136–139] and the identification of dislocation avalanches [70,140–145].

4.1. Machine learning models based on features of dislocation ensembles

Despite the wide range of successful applications achieved by DDD in the field of mechanical properties of materials, the advent of materials informatics in materials science is setting a new paradigm on how traditional DDD frameworks can be utilized towards insightful predictions for structure-property relationships [146,147]. The ability of ML models such as ANN to learn complex non-linear behaviors provides an opportunity to guide the understanding of complex collective dislocation behaviors and their influence on mechanical deformation of materials. In this way, DDD can be regarded as a high-throughput approach to generate large data, to be analyzed by ML approaches in the search of new physical correlations, generally hidden by the ensemble and relevant mechanisms' complexity. In this context, Ref. [148] demonstrated that it is possible in simple models to predict stress/strain curves as simulated by two-dimensional DDD simulations starting from features extracted from initially random dislocation microstructures. The importance of the result is encapsulated in the complexity of the initially random configurations, that may also acquire fractal characters [70,149,150]. The descriptors exploited in this work were features extracted from the initial dislocation microstructures (Fig. 6(a)) such as the density of GNDs and the internal stresses as shown in Fig. 6(b) and (c), respectively. The authors addressed the issue of deformation predictability starting from systems with different sizes and also considering the effect of pre-straining the system before training (Fig. 6(d) and (e)). Results regarding the performances of their NN-based model (Fig. 6(f)) on the prediction of the stress-strain curves showed that better performances can be obtained for larger systems and higher deformation strain. Also, an initial pre-deformation improves the predictive ability of the model. Similar topological descriptors were employed in [?] to characterize precipitates-controlled jamming-to-pinning in terms of the topology the dislocation network. variations in the

In more physically consistent conditions, Ref. [151] discussed high-throughput DDD simulations [47] towards generating synthetic experiments for 2D thin aluminum films, where the pre-processing route, in the context of the level of tensile pre-straining, controlled the initial dislocation configuration. Then, prediction of stress-strain curves and mechanical properties was achieved by repeating synthetic DIC experiments, and by using both unsupervised and deep-learning ANNs. The comparison between deep-learning and unsupervised learning method showed a drastic advantage towards deep ANN approaches. In this way, progress in DIC techniques may provide a tool for a complete, non-destructive characterization of mechanical behavior of materials.

Finally, it is worth noticing that ML methods may be used towards a complete classification of dislocation ensembles. While physical tools such as local misorientations, or topological features [41] may allow for ensemble classification, Ref. [152] investigated the possibility of defining ML descriptors for the classification of dislocation microstructures. The authors carried out DDD simulations with free surfaces in order to test different density fields commonly used in continuum models of dislocations such as the dislocation density, Nye's tensor, and higher-order fields. In this context, it was possible to classify the sets of descriptors commonly used in Continuum Dislocation Dynamics (CDD). This result is important as it paves the way for a multi-scale framework where data are automatically passed from lower to higher-scale models. Other works [153] focused on calibrating CP models based on results from DDD by exploiting concepts from ML such as regularized regression and cross-validation.

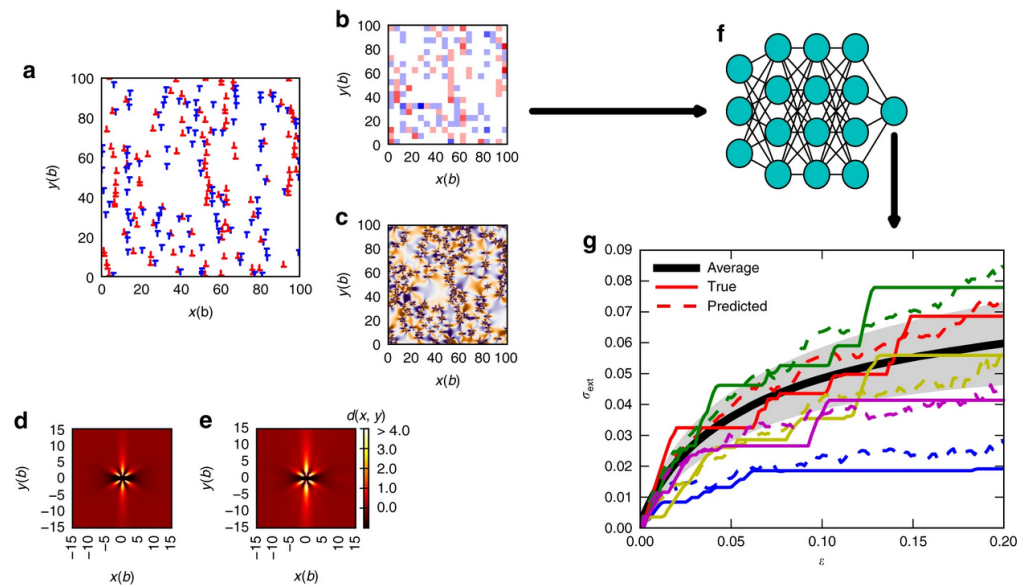


Figure 6. Workflow showing a ML approach applied to DDD simulations for the prediction of mechanical properties of materials, reproduced from Ref. [148]. Features are selected from the initial dislocation microstructure (a) like the geometrically necessary dislocations (GND) density of positive and negative dislocations (b) and the internal stress field (c). After defining as initial states relaxed random configurations (pair correlation function shown in (d)), or pre-deformed dislocation configurations (e), the authors trained a neural network (NN) (f) in the prediction of stress/strain curves based on these descriptors. Results showing NN predictions against true stress/strain curves are shown in (g). The figure was reproduced after [148] based on Creative Commons Attribution 4.0 license.

4.2. Coupling dislocations to microstructural imaging experiments

Due to the inherent length-scale, DDD simulations have been considered as an ideal candidate for the direct comparison with experiments, particularly regarding small-scale testing of thin films [154,155], micropillars [156–158] or nanoindentation experiments [159]. However, the advent of materials informatics has raised new opportunities for the closer coupling of experiments and modeling [160]. The improvements associated with the increase of the computational power and the progression in detector technology, with better acquisition rate and spatial resolution of the detectors used in characterizations, are opening the way to the application of MI techniques in experiments-modeling coupling. Large-scale simulations permit the modeling of more realistic systems and *in-situ* characterization techniques leading to significantly richer data sets that can be used for improving the physical models, the determination of model parameters and the validation of results. A possible application of this coupled approach will be the generation of virtual dislocation microstructures based on *in-situ* experiments to be used as initial conditions for DDD simulations. Although this direct coupling has not been widely investigated so far, several attempts have been made in the generation of synthetic experiments starting from DDD simulations to be used for the application of MI techniques.

5. Learning dislocation features from nanomechanics *in-situ* experiments

Experimental exploration of optimal metallic alloys with promising mechanical properties at extreme conditions need to be augmented and supported by numerical modeling towards describing physical and chemical mechanisms of plastic deformation. Thus, nanoscale investigations of mechanical deformation (e. g. nanoindentation) are imperative towards multiscale modeling strategies for closer coupling between numerical models and experiments. However, severe size effects and microstructural

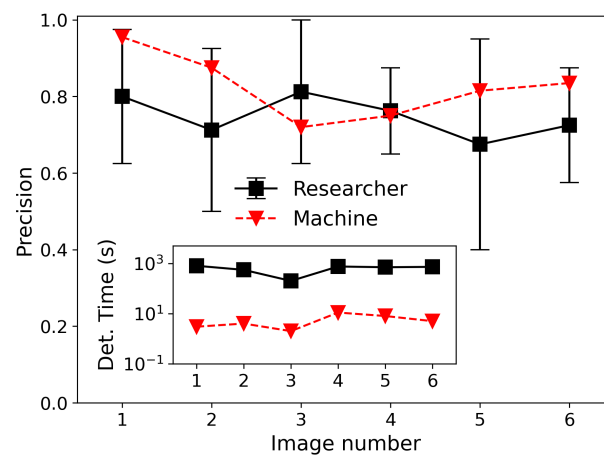


Figure 7. A scientific software based on CNN is applied to quantify defects in materials. A comparison of precision and time efficiency of both the human and machine labeling is presented. The image number correspond to selected one from the complete test set and error bars correspond to the lowest and maximum precision obtained by the 5 researchers. The selected images were also analyzed by five researchers [75].

noise [70] require statistical tools operating on large data, collected from experiments at (nano-)scales, where comparisons to simulations are viable. For example, validation of numerical simulations can be provided by *in-situ* nanomechanical experiments, thus providing more information about plastic deformation and a detailed characterization of the samples being tested, to enable direct comparison of the time evolution (with mechanical or thermal loading) response of modeling[41,160] This motivates the development of MI software workflows that capture critical resolved shear stress and dislocation behaviors on various slip systems, noting that a detailed analysis of atomic-scale processes of plastic deformation, such as dislocation propagation and twin growth provides an understanding of the defects nucleation, propagation, and multiplication, as well as interactions with other surrounding defects. This aims to create a benchmark of mechanical properties and dislocation dynamics for further experimental data analysis [161] that provides insights into the dislocation reactions at an atomistic level where slip-grain boundary (GB) interactions occur [1,162].

A characteristic case of large data in this context, is the scientific exploration of the detailed characterization of dislocation nucleation and propagation[163], that requires the visualization of dislocations during mechanical loading, by adjusting crystal orientations at every single loading step. The method requires a reconstruction of a large number of 3D frames where materials informatics is involved due to its advantages given by automated analysis of these emerging and inevitably increasing data generation capabilities to detect and classify materials defects. As an example of MI applications in this context[2,75], a CNN is applied to automate a defect analysis in electron microscopic images. In Fig. 7 the authors show the performance of the associated scientific software compared to work done by five expert researchers (human work) through identifying material defects. Due to observable limitations of human assessment, the authors concluded that the machine performance vs. human performance clearly demonstrated the need for high quality ML models vs. the ground truth. Into this research scope, modern methods in computer vision and ML have been applied, such as unsupervised statistical representation of microstructure images [164], where reported results show that pre-trained neural networks represented micrographs well with no previous knowledge of the nature of shapes. Also, clustering or classification in various materials images with deep convolutional neural networks is utilized to automatically analyze position averaged convergent beam electron diffraction patterns in scanning 4D scanning-TEM images. Therefore, the aforementioned methods and results indicate that computer

vision techniques are very promising for microscopic image analysis by replacing human labor and producing a reliable analysis output.

Modern advances in software and the development of fast computer processors potentially increase real-time image recognition, by coupling them with the electron microscopic system to make possible a direct, on-the-fly *in-situ* analysis. Here, efforts have been done to apply CNN architecture to analyze and identify common crystallographic defects in structural alloys like dislocation lines, precipitates and voids[165]. In Fig. 8 the authors reported the machine vs human performances by computing materials evaluation metrics for defect quantification performed, and several categories of the defect quantification such as dislocation density, number density, diameter, and diameter standard deviation of precipitates and voids. Overall the CNN based scientific software obtained more accurate result than the expert researchers. The authors concluded that an automated identification of common crystallographic defects in metals using deep learning semantic segmentation, based on high-quality microscopy data, providing more accurate results than conventional manual counting, and thus advance model performance. However, current scientific software needs to overcome several technical issues like multiple contrast mechanisms that may contribute to the observation of defect features by misidentifying them, particularly those with divergent contrast.

6. Beyond mechanical property empirical rules: Learning how to design metal alloys

The compositional complexity of alloys has had a vertiginous increase since the last century (see Fig 9). Therefore, a necessity for new materials capable to support harsh environments together with the improvement of experimental techniques have made possible the creation of a complex solid solution (CSS) with more than five elements in a single phase[166,167]. Since the spatial arrangement of a given phase as well as the particle packing determine the electrical, chemical, and mechanical properties[168–170] of every alloy, atomistic simulations are generally accompanied with post-processing algorithms that allow us to gain some insight into systems by means of assigning a structural type of each particle based on its atomic environment. Among the conventional classification methods we have the common neighbor parameter (CNP)[171], based on a combination of the common neighbor analysis (CNA)[172,173], and the centrosymmetry parameter (CSP)[174–176]. The idea is to define a parameter as in CSP,

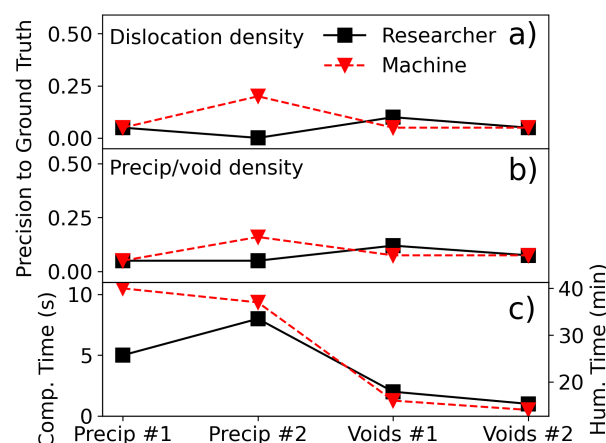


Figure 8. A modern electron microscopic system coupled with a CNN based scientific software is applied to directly analyze materials defects. This figure presents the comparison of materials evaluation metrics for defect quantification performed by computer and by expert researchers. Materials metrics include dislocation density in a), and precipitates and voids number density in b), the diameter and the time spend for computer and human experts to quantify these defects in c) [165].

but instead of related with the centrosymmetry of the lattice, the parameter depends on the common atomic neighborhood like the CNA method. Thus, the CNP combines the strength of both CNA and CSP. Therefore their success in correctly describing for example amorphous Ni[177], grain boundaries[178,179], metallic glasses[180], deformation of nanocrystalline metals[181,182], and shock deformation of metals[183]. Another frequently used method is the Voronoi analysis (VA)[184]. However, due to its high sensitivity to lattice distortions[185,186] the method is generally used to characterize atomic structures of liquid and amorphous systems[187–192]. Other less common techniques are also employed to characterize the atomic structures[193,194]. Beyond conventional approaches, MI provides a set of useful tools for the classification of microstructures that emerge in CSSs.

ML techniques have proven to be very useful for the discovery of new materials, as well as for the creation of interatomic potentials capable of describing alloys with even more than five elements[196,197]. Thus, if we can provide sufficiently accurate features to an ML algorithm, we can accelerate the exploration of the compositional space by orders of magnitude faster than those performed in first-principles simulations. For this purpose, MI has focused on advances in effectively and efficiently exploring the compositional space to identify novel materials in crystalline and amorphous states, as well as developing ML interatomic potentials to correctly describe the physical properties and enable studies of longer size and longer time than those accessible using *ab initio* methods.

The discovery of new materials was traditionally made by trial and error design strategies, taking different ways for crystals and glassy systems. For crystalline solid solutions, the Hume-Rothery rules appear as a first empirical attempt to predict the solid solution forming ability (SSFA) i.e., the formation of a single crystalline phase, in binary alloys[198,199]. Recently, as a consequence of the discovery of HEAs, the Hume-Rothery rules have been shown insufficient, and new empirical rules have risen to predict SSFA[200–202]. The glass-forming ability (GFA) for metallic alloys, on the other hand, is predicted based on the Inoue empirical rules[203] in an analogous manner as Hume-Rothery rules. However, these rules cannot correctly predict the GFA, and metallic glasses that can be easily created experimentally, such as CuZr[204,205], do not satisfy all Inoue's rules. Thus, despite the efforts to propose new criteria to predict SSFA and

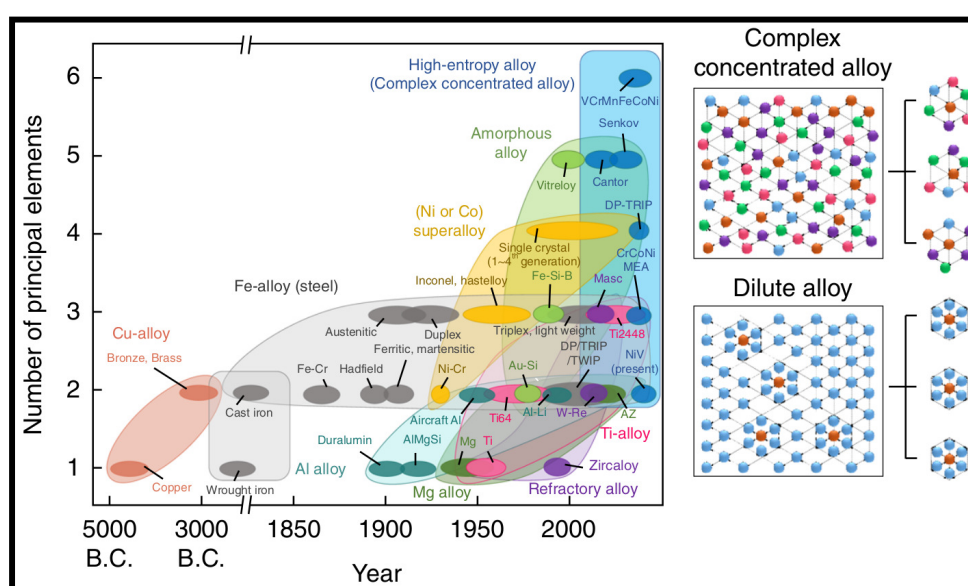


Figure 9. Historical evolution of the compositional complexity of alloys. From simple unary systems to the more complex multi-principal element alloys. Reproduced after [195] based on Creative Commons Attribution 4.0 International License.

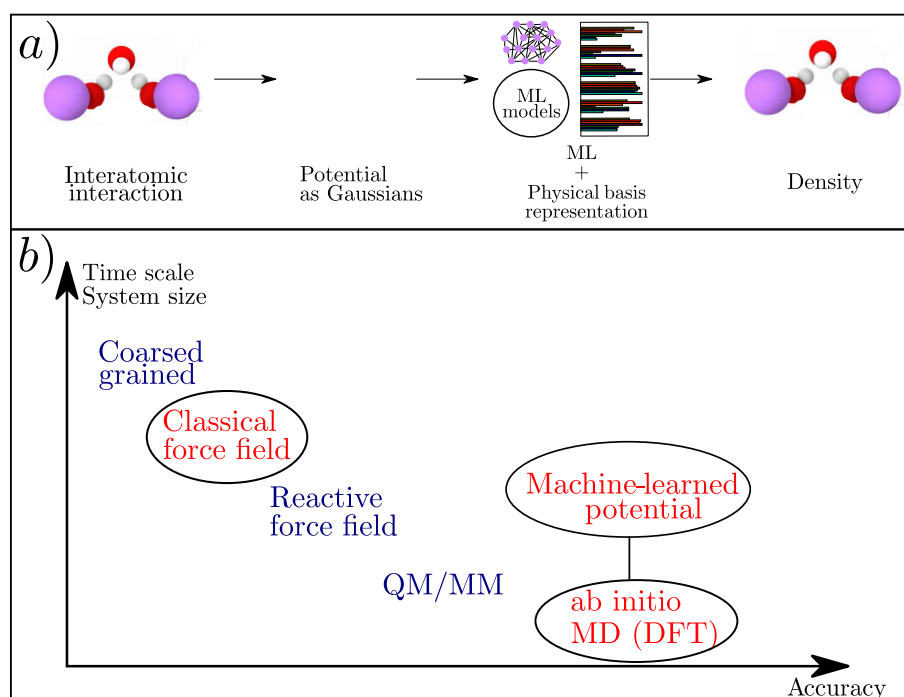


Figure 10. Schematic description of the two approaches employed to calculate the particle interaction in atomistic simulations using ML methods. a) A Hohenberg-Kohn map using ML models to bypass the calculation of Kohn-Sham equations. b) A schematic representation of the trade-off between cost and accuracy of force field and MLPs fitted from DFT calculations.

GFA, these empirical rules usually present compromised predictability. Replacing these traditional methods with ML approaches has resulted in very accurate predictability of novel high-throughput material design[206–208]. Concrete examples of the success of the composition exploring are the material genome project, idealized to uncover properties of inorganic materials[209], Gaussian process classification algorithms towards efficient predictions of promising solutes[210], the impact of local lattice relaxation to stabilize single-phase of BCC HEA[211], and last but not least, phase prediction of HEAs[212,213].

In the context of using ML models to predict interatomic interactions between atoms of alloys, there is a variety of applicable strategies that are based on quantum mechanical methods (such as DFT), that provide quantum mechanical elements (forces, energies), that allow ML methods to develop Machine Learning Potentials (MLP) to infer interatomic potentials or/and forces, with critically decreasing computational time for achieving *ab initio* accuracy. Fig. 10 shows a schematic representation of common approaches employed to calculate particle interactions in atomistic simulations using ML models. Fig. 10a is an example of accelerating the computational cost by avoiding solving the Kohn-Sham equations and obtaining the total energy directly from the charge density[214–219]. Fig. 10b is a schematic representation of the trade-off between conventional atomistic simulations and MLPs. The generation of these MLPs is a problem of regression involving map of the potential energy landscape as well as its derivatives (the force field) by fitting training data generated from thousands of DFT calculations. Readers interested in calculations based on these approaches can found additional information in references [14,15,196,220].

7. Materials deformation informatics: Challenges, prospects and ontology

As of today, digitalization is one of the main driving forces of technological and scientific progress. Jointly with a variety of quantitative modeling and simulation techniques, digitalization is a passage towards the so-called Industry 4.0 [221], and in particular, it can accelerate the discovery and design of new materials. To predict various

physical properties of materials, many data-based models are incorporated, which is supported by increasingly complex simulations run on increasingly performant infrastructure [222,223]. That leads to the creation of an enormous amount of heterogeneous data, originating from different methods and workflows, maintained by a significant number of groups within different communities.

Managing large data sets is at the core of Materials Informatics, where it allows for obtaining valuable insights on a given material. For example, image-recognition computational techniques rooted in Machine Learning are used in the data analysis of images originating from different experimental sources and various imaging techniques. This includes optical microscopy (OM), digital image correlation (DIC and HR-DIC), scanning electron microscopy (SEM), and others. Machine Learning is also relevant for other data-heavy techniques, where different sets of parameters are considered (data sets with atomic positions, velocities, defects, lengths, orientations, etc.). From the standpoint of data processing another important domain is 3D visualizations of dislocation nucleation, where materials informatics allows for automated analysis of the emerging defects which inevitably increases data generation capabilities. Data processing and management are very relevant also in the canonical techniques of simulating material properties: Molecular Dynamics (MD) and Monte Carlo (MC). While the MD explores the phase space in a deterministic way by solving the equation of motion for physical evolution, the MC samples the configuration space according to the Maxwell-Boltzmann statistics. Both methods however are very data-intensive and lead to the creation a very significant amount of data.

The abundance of incompatible data sources often causes reduced interoperability of the data, which in turn prompts the development of model data repositories, often equipped in particular analytical tools, allowing to integrate and share the scattered data [224]. Among the multitude of names, it is worth noting: Integrated Computational Materials Engineering (solids) [225], Computational Molecular Engineering (fluids) [224], Process Data Technology or Computer-Aided Process Engineering (process technology)[226].

The process of digitalization is two-step: first, the data is turned into digital form (digitization), which results in raw unannotated data (dark data) [227,228]. This is then followed by giving the data a suitable structure and description (metadata) to ensure that it remains findable, accessible, inter-operable, and reusable i.e. conform to the FAIR principles [229]. Even though all the FAIR principles are fundamental, mutually dependent, and can not be separated, in some cases (like complex data-driven workflows), it is the interoperability that becomes the most relevant [230]. Therefore to provide a shared, standardized representation of the domain knowledge, significant efforts were made towards creating metadata standards and sophisticated classification schemes (ontologies) which are utilized to make complex data standardized, compatible, and efficiently searchable [231].

7.1. Interoperability

Interoperability can be loosely defined as an agreement between multiple parties (ideally the whole community) on a common terminology standard, defined by an ontology. The very notion can be split into three main aspects, all originating from theoretical linguistics: syntax, semantics, and pragmatics [221]. Syntactic interoperability is associated with the grammar of a formal language (folders structure, file formats, data items arrangement, etc.), while semantic interoperability is focused on an agreement on the meaning of implications of the data content [232]. Semantic interoperability is only achieved with accepted metadata standards, which support understanding the data annotation by all parties in an agreed way. This subsequently allows to integrate all the communicated data within a single platform with multiple sources and users and ultimately enables cross-system usability [233]. The above two aspects are insufficient without defining the context in which the communication occurs and the general un-

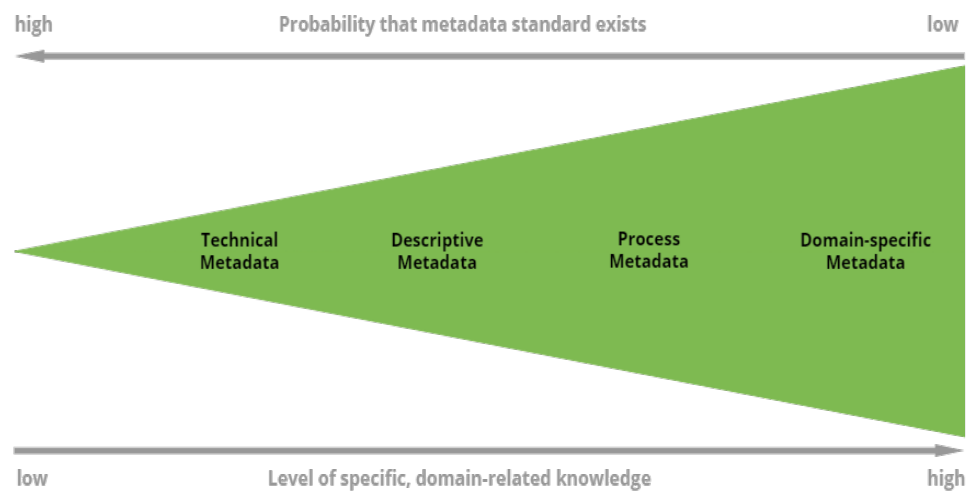


Figure 11. Metadata categories with specificity level. With an increasing level of specificity, the probability of existing metadata standard is decreasing (based on [221]).

derstanding of performative roles or, in other words, what different participants in an exchange can reasonably expect from each other [221]. The latter is addressed by pragmatic interoperability, which concerns such requirements as well as recommendations of the practice of communicating and dealing with data [234–236].

7.2. Metadata

As pointed out, according to the FAIR principles data should be findable, accessible, interoperable, and reusable. A commonly acknowledged way of conforming the data management to those principles is by incorporating *metadata*. Metadata can be defined as *data about data* or more formally – a structured form of knowledge representation [237] which serves the purpose of describing a research asset and its selected aspects. It therefore allows to relieve the direct communication of the parties exchanging data, which lowers “science friction” [221,238] and supports workflows specific of a domain [239]. It is a common practice to divide metadata into categories based on the metadata specificity – some of them are rather general (for example file sizes, authors, etc.) whereas others are very particular and only applicable to a single domain. Below, we briefly present the semantic description of four main classes, which are at the core of every sensible data description and hold true for engineering (including modeling materials) and multiple branches of science [221].

- technical metadata - technical aspect of the research asset, mostly the file attributes on a file system level and similar syntactic information (file sizes, checksum information, storage location, access dates, file formats),
- descriptive metadata - general information about the research asset (authors, keywords, title),
- process metadata – information on the generation process of the research asset (for example the computational environment and software used to generate or process the data). It may consist of several consecutive steps,
- domain-specific metadata – domain-specific description of the research objects. For example in computational engineering, this includes details about the simulated system, methods of simulation, resolution, etc.

The specificity of the categories is presented in ascending order in Fig. 11 [221]. While the technical and descriptive metadata keys are rather generic and relevant for various branches of science, the categories of process metadata and domain-specific metadata are bound to the research process and the research object respectively. To an extent, different classes may overlap and a metadata key can be part of several categories.

The likelihood that suitable standards exist for a given category is smaller for its larger specificity [221] (see figure 11). While multiple standards are existing for technical and descriptive metadata, this does not hold for the process and domain-specific metadata, where a significant development effort is often required. For technical and descriptive metadata, the semantic information is similar throughout all disciplines (DataCite is the standard for a general description and citable data objects [241]). On the other hand, process metadata are deeply connected with the research process, where metadata standards only exist for specific processes (CodeMeta [242], Citation File Format[243]). Similarly for domain-specific metadata, the appropriate standards only exist for few research objects.

Importantly, the distinction between the above four categories is critical also for automated extractability. While technical information is usually easy to extract (mostly file system attributes), and process- and domain-specific information is relatively easy to extract automatically for computational engineering applications. On the other hand, the descriptive information is hardly extractable, since it provides a higher-level description of the research. Therefore, human interpretation is often crucial[221].

7.3. Ontologies

A convenient way of organizing metadata is by incorporating an *ontology*, which serves as a framework for metadata design [240]. Ontologies and ontology-based techniques are aimed at providing a shared standardized representation of terminology over a domain. In particular, by imposing an ontology-based description of the data, the level of interoperability and reusability is increased, while introducing metadata derived from ontology allows promoting findability and accessibility. Ontologies are defined as sophisticated classification schemes aimed at making complex data searchable. They have a hierarchical structure and define the basic terms and relations over a domain, as well as the rules for combining them [240]. Ontologies consist of four components: (i) concepts that represent sets or classes of entities in a domain, (ii) instances that represent the actual entities, (iii) relations (*is-a* or *has-a*), and (iv) axioms representing restrictions imposed on the domain. The tree-like ontology structure allows organizing entities based on their granularity, from general to specific. The actual set of components might be the basis for

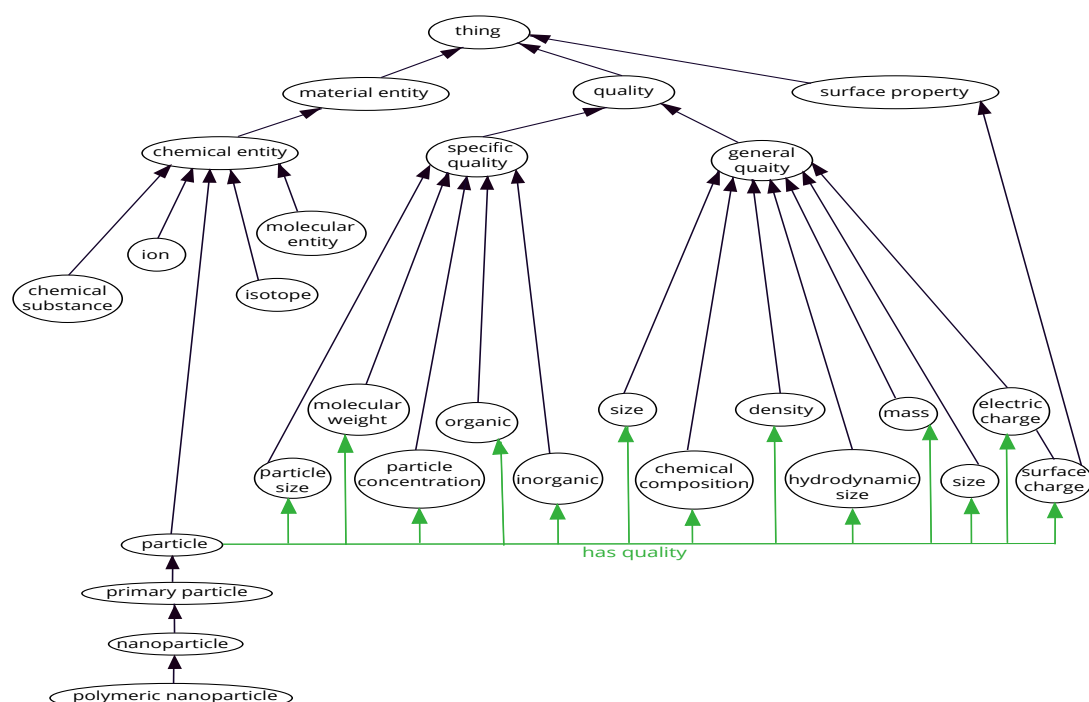


Figure 12. An ontology example: Sample of the NanoParticle ontology (based on [240]).

ontology classification; for example, in Fig. 12 a little sample of NanoParticle ontology is presented where the black arrows represent axioms of is-a relations, i.e. if A is B, then all entities belonging to A belong to B as well. Also, A is referred to as a sub-concept of B. In this example chemical substance, particle, ion, isotope, and molecular entity are all sub-concepts of a chemical entity. Therefore, all chemical substances, particles, ions, isotopes, and molecular entities are also chemical entities. Further, all primary particles are particles, all nanoparticles are primary particles, and so on. It is clear that the is-a relation is transitive i.e. describes inheritance of properties. Another important relationship is a has-a relation, which is associated with the axiom that entities have qualities (green arrows).

A metadata model can be seen as an intermediate state between a non-formal data description and a complete formalization of metadata keys. Its purpose is to describe a research object as well as the relation to other objects, which derives from the common understanding of the domain. This approach could also be called ontology-based metadata since the metadata model is build based on an object model.

The process of engineering a hierarchical metadata model includes several consecutive steps, some of which might be repeated until a detailed enough description of the research is reached. In the first part, a clear understanding of the research object is reached which is achieved by the analysis of the research process with all potentially involved stakeholders and using a natural language. In this step, the information on relevant entities is found and grouped which is subsequently followed by identifying the attributes describing entities (names, units, etc.) as well as terms, relations, and rules (in material modeling an example of relevant entity is a *component*, which represents a chemical species). Everyone potentially involved contributes to this step, since the metadata will serve as a semantic convention for further communication. Finally, the relations between the entities must be understood, including how different entities are connected to create an in-depth description. This part is subjective, arbitrary, and strongly specific to the research.

Once the object model is defined in a formal language, some attention has to be paid if its certain parts exist in other standards (the more general metadata, the more probably they are to be found in other metadata standards). In the last step, the metadata model representation is created in terms of choosing the formal language. In most situations, it is XSD2 or JSON, which offers the possibility of defining the entities, attributes, and relations.

8. Conclusions

In this review, we have addressed common applications of data science and machine learning methods relevant to a broad range of deformation and failure properties specific to metal alloys. In this framework, the materials informatics capacity has been highlighted in various contexts related to microstructural classification, phase stability, in-situ characterization, constitutive modeling, property prediction, and materials discovery and high-throughput design, in both supervised and unsupervised settings. The associated discussions have provided relevant data analytics tools to process, probe, and interpret multi-scale and multi-dimensional (simulation-based and/or experimentally-derived) datasets within the structure-property framework.

We conclude this overview by restating the fact that data-driven research of materials necessitates smooth access to well-maintained and reusable databases which are diverse and heterogeneous in nature, as noted earlier in Section 7. Sources of disparities in data are typically diverse and often stem from different measurement techniques and associated uncertainties, various experimental conditions, and underlying assumptions. Therefore, it is quite likely that a careless combination of information from different data repositories often leads to misinterpretation of spurious trends and significant bias in predicted properties. This, in fact, highlights a need for human supervision in building

meaningful synergies among diverse data sources to ensure soundness and robustness of machine-generated output.

Author Contributions: Conceptualization, S.P.; methodology, K.F., K.K., M.P., R.A-D., F.J.D-G., F.R.; investigation, S.P., K.F., K.K., M.P., R.A-D., F.J.D-G., F.R.; resources, S.P.; d writing—original draft preparation, K.F., K.K., M.P., R.A-D., F.J.D-G., F.R.; writing—review and editing, S.P.; supervision, S.P.; project administration, S.P.; funding acquisition, S.P. All authors have read and agreed to the published version of the manuscript, please turn to the [CRediT taxonomy](#) for the term explanation.

Funding: This research was funded by the European Union Horizon 2020 research and innovation program under grant agreement no. 857470 and from the European Regional Development Fund via Foundation for Polish Science International Research Agenda PLUS program grant no. MAB PLUS/2018/8

Acknowledgments: We would like to thank Mikko Alava for inspiring discussions and fruitful suggestions and Paweł Sobkowicz for showing us the inspiring motivations for this work.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Rajan, K. Materials informatics. *Materials Today* **2005**, *8*, 38–45. doi:10.1016/S1369-7021(05)71123-8.
2. Rickman, J.M.; Lookman, T.; Kalinin, S.V. Materials informatics: From the atomic-level to the continuum. *Acta Materialia* **2019**, *168*, 473–510.
3. Ramakrishna, S.; Zhang, T.Y.; Lu, W.C.; Qian, Q.; Low, J.S.C.; Yune, J.H.R.; Tan, D.Z.L.; Bressan, S.; Sanvito, S.; Kalidindi, S.R. Materials informatics. *Journal of Intelligent Manufacturing* **2019**, *30*, 2307–2326.
4. Rodgers, J.R.; Cebon, D. Materials informatics. *MRS bulletin* **2006**, *31*, 975–980.
5. Tanaka, I.; Rajan, K.; Wolverton, C. Data-centric science for materials innovation. *MRS Bulletin* **2018**, *43*, 659–663.
6. Reyes, K.G.; Maruyama, B. The machine learning revolution in materials? *MRS Bulletin* **2019**, *44*, 530–537.
7. Davis, J.R.; Allen, P.; Lampman, S.; Zorc, T.B.; Henry, S.D.; Daquila, J.L.; Ronke, A.W. *Metals handbook: properties and selection: nonferrous alloys and special-purpose materials*; ASM international, 1990.
8. DeCost, B.L.; Francis, T.; Holm, E.A. Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures. *Acta Materialia* **2017**, *133*, 30–40. doi:https://doi.org/10.1016/j.actamat.2017.05.014.
9. DeCost, B.L.; Lei, B.; Francis, T.; Holm, E.A. High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel. *Microscopy and Microanalysis* **2019**, *25*, 21–29.
10. Liu, G.; Jia, L.; Kong, B.; Feng, S.; Zhang, H.; Zhang, H. Artificial neural network application to microstructure design of Nb-Si alloy to improve ultimate tensile strength. *Materials Science and Engineering: A* **2017**, *707*, 452–458.
11. Kim, G.; Diao, H.; Lee, C.; Samaei, A.; Phan, T.; de Jong, M.; An, K.; Ma, D.; Liaw, P.K.; Chen, W. First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation. *Acta Materialia* **2019**, *181*, 124–138.
12. Wen, C.; Zhang, Y.; Wang, C.; Xue, D.; Bai, Y.; Antonov, S.; Dai, L.; Lookman, T.; Su, Y. Machine learning assisted design of high entropy alloys with desired property. *Acta Materialia* **2019**, *170*, 109–117.
13. Xue, D.; Balachandran, P.V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated search for materials with targeted properties by adaptive design. *Nature communications* **2016**, *7*, 1–9.
14. Botu, V.; Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *International Journal of Quantum Chemistry* **2015**, *115*, 1074–1083.
15. Bartók, A.P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J.R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Science advances* **2017**, *3*, e1701816.
16. Morgan, D.; Jacobs, R. Opportunities and challenges for machine learning in materials science. *Annual Review of Materials Research* **2020**, *50*, 71–103.
17. Friederich, P.; Häse, F.; Proppe, J.; Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nature Materials* **2021**, *20*, 750–761.
18. Friedman, J.; Hastie, T.; Tibshirani, R.; others. *The elements of statistical learning*; Vol. 1, Springer series in statistics New York, 2001.
19. Schmidt, J.; Marques, M.R.; Botti, S.; Marques, M.A. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials* **2019**, *5*, 1–36.
20. Shang, Y.; Brecht, J.; Psitidda, C.; Liaw, P.K. Mechanical behavior of high-entropy alloys: A review. *arXiv preprint arXiv:2102.09055* **2021**.
21. Liu, Y.; Zhao, T.; Ju, W.; Shi, S. Materials discovery and design using machine learning. *Journal of Materiomics* **2017**, *3*, 159–177.
22. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K.A. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002.

23. Villars, P.; Cenzual, K. *Pearson's crystal data®: crystal structure database for inorganic compounds*; ASM International Materials Park, OH, 2007.
24. Saal, J.E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *Jom* **2013**, *65*, 1501–1509.
25. Ward, L.; O'Keeffe, S.C.; Stevick, J.; Jelbert, G.R.; Aykol, M.; Wolverton, C. A machine learning approach for engineering bulk metallic glass alloys. *Acta Materialia* **2018**, *159*, 102–111.
26. Salvador, C.A.; Zornio, B.F.; Miranda, C.R. Discovery of Low-Modulus Ti-Nb-Zr Alloys Based on Machine Learning and First-Principles Calculations. *ACS Applied Materials & Interfaces* **2020**, *12*, 56850–56861.
27. Furmanchuk, A.; Agrawal, A.; Choudhary, A. Predictive analytics for crystalline materials: bulk modulus. *RSC advances* **2016**, *6*, 95246–95251.
28. Xiong, J.; Zhang, T.Y.; Shi, S.Q. Machine learning prediction of elastic properties and glass-forming ability of bulk metallic glasses. *MRS Communications* **2019**, *9*, 576–585.
29. Chang, Y.J.; Jui, C.Y.; Lee, W.J.; Yeh, A.C. Prediction of the composition and hardness of high-entropy alloys by machine learning. *Jom* **2019**, *71*, 3433–3442.
30. Mansouri Tehrani, A.; Oliynyk, A.O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T.D.; Brgoch, J. Machine learning directed search for ultraincompressible, superhard materials. *Journal of the American Chemical Society* **2018**, *140*, 9844–9853.
31. Fan, Z.; Ding, J.; Ma, E. Machine learning bridges local static structure with multiple properties in metallic glasses. *Materials Today* **2020**, *40*, 48–62.
32. Jafary-Zadeh, M.; Khoo, K.H.; Laskowski, R.; Branicio, P.S.; Shapeev, A.V. Applying a machine learning interatomic potential to unravel the effects of local lattice distortion on the elastic properties of multi-principal element alloys. *Journal of Alloys and Compounds* **2019**, *803*, 1054–1062.
33. Sharp, T.A.; Thomas, S.L.; Cubuk, E.D.; Schoenholz, S.S.; Srolovitz, D.J.; Liu, A.J. Machine learning determination of atomic dynamics at grain boundaries. *Proceedings of the National Academy of Sciences* **2018**, *115*, 10943–10947.
34. Chen, W.C.; Schmidt, J.N.; Yan, D.; Vohra, Y.K.; Chen, C.C. Machine learning and evolutionary prediction of superhard BCN compounds. *npj Computational Materials* **2021**, *7*, 1–8.
35. Avery, P.; Wang, X.; Oses, C.; Gossett, E.; Proserpio, D.M.; Toher, C.; Curtarolo, S.; Zurek, E. Predicting superhard materials via a machine learning informed evolutionary structure search. *npj Computational Materials* **2019**, *5*, 1–11.
36. Hart, G.L.; Mueller, T.; Toher, C.; Curtarolo, S. Machine learning for alloys. *Nature Reviews Materials* **2021**, pp. 1–26.
37. Guo, K.; Yang, Z.; Yu, C.H.; Buehler, M.J. Artificial intelligence and machine learning in design of mechanical materials. *Materials Horizons* **2021**, *8*, 1153–1172.
38. Bock, F.E.; Aydin, R.C.; Cyron, C.J.; Huber, N.; Kalidindi, S.R.; Klusemann, B. A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Frontiers in Materials* **2019**, *6*, 110.
39. Liu, H.; Fu, Z.; Yang, K.; Xu, X.; Bauchy, M. Machine learning for glass science and engineering: A review. *Journal of Non-Crystalline Solids* **2021**, *557*, 119419.
40. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials* **2017**, *3*, 1–13.
41. Papanikolaou, S. Microstructural inelastic fingerprints and data-rich predictions of plasticity and damage in solids. *Computational Mechanics* **2020**, *66*, 141–154.
42. Papanikolaou, S. Learning local, quenched disorder in plasticity and other crackling noise phenomena. *npj Computational Materials* **2018**, *4*, 1–7.
43. Bulgarevich, D.S.; Tsukamoto, S.; Kasuya, T.; Demura, M.; Watanabe, M. Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures. *Scientific reports* **2018**, *8*, 1–8. doi:https://doi.org/10.1038/s41598-018-20438-6.
44. Bulgarevich, D.S.; Tsukamoto, S.; Kasuya, T.; Demura, M.; Watanabe, M. Automatic steel labeling on certain microstructural constituents with image processing and machine learning tools. *Science and technology of advanced materials* **2019**, *20*, 532–542. doi:https://doi.org/10.1080/14686996.2019.1610668.
45. Kim, H.; Inoue, J.; Kasuya, T. Unsupervised microstructure segmentation by mimicking metallurgists' approach to pattern recognition. *Scientific Reports* **2020**, *10*, 1–11. doi:https://doi.org/10.1038/s41598-020-74935-8.
46. Boukhtache, S.; Abdelouahab, K.; Berry, F.; Blaysat, B.; Grediac, M.; Sur, F. When Deep Learning Meets Digital Image Correlation. *Optics and Lasers in Engineering* **2021**, *136*, 106308.
47. Papanikolaou, S.; Tzimas, M.; Reid, A.C.; Langer, S.A. Spatial strain correlations, machine learning, and deformation history in crystal plasticity. *Physical Review E* **2019**, *99*, 053003.
48. Papanikolaou, S.; Tzimas, M. Effects of rate, size, and prior deformation in microcrystal plasticity. *Mechanics and Physics of Solids at Micro-and Nano-Scales* **2019**, pp. 25–54.
49. Papanikolaou, S.; Alava, M.J. Direct detection of plasticity onset through total-strain profile evolution. *Physical Review Materials* **2021**, *5*, 083602.
50. Cidade, R.A.; Castro, D.S.; Castrodeza, E.M.; Kuhn, P.; Catalanotti, G.; Xavier, J.; Camanho, P.P. Determination of mode I dynamic fracture toughness of IM7-8552 composites by digital image correlation and machine learning. *Composite Structures* **2019**, *210*, 707–714.

51. Rezaie, A.; Achanta, R.; Godio, M.; Beyer, K. Comparison of crack segmentation using digital image correlation measurements and deep learning. *Construction and Building Materials* **2020**, *261*, 120474.
52. Gehri, N.; Mata-Falcón, J.; Kaufmann, W. Automated crack detection and measurement based on digital image correlation. *Construction and Building Materials* **2020**, *256*, 119383.
53. Strohmman, T.; Starostin-Penner, D.; Breitbarth, E.; Requena, G. Automatic detection of fatigue crack paths using digital image correlation and convolutional neural networks. *Fatigue & Fracture of Engineering Materials & Structures* **2021**, *44*, 1336–1348.
54. Azimi, S.M.; Britz, D.; Engstler, M.; Fritz, M.; Mücklich, F. Advanced steel microstructural classification by deep learning methods. *Scientific reports* **2018**, *8*, 1–14.
55. Mulewicz, B.; Korpala, G.; Kusiak, J.; Prah, U. Autonomous interpretation of the microstructure of steels and special alloys. Materials Science Forum. Trans Tech Publ, 2019, Vol. 949, pp. 24–31.
56. Gola, J.; Webel, J.; Britz, D.; Guitart, A.; Staudt, T.; Winter, M.; Mücklich, F. Objective microstructure classification by support vector machine (SVM) using a combination of morphological parameters and textural features for low carbon steels. *Computational Materials Science* **2019**, *160*, 186–196.
57. Chen, Z.; Daly, S. Deformation twin identification in magnesium through clustering and computer vision. *Mater. Sci. Eng. A* **2018**, *736*, 61–75.
58. Chen, Z.; Daly, S. Automated identification of deformation twin systems in Mg WE43 from SEM DIC. *Materials Characterization* **2020**, *169*, 110628.
59. Ding, Z.; Pascal, E.; De Graef, M. Indexing of electron back-scatter diffraction patterns using a convolutional neural network. *Acta Materialia* **2020**, *199*, 370–382.
60. Kaufmann, K.; Zhu, C.; Rosengarten, A.S.; Vecchio, K.S. Deep neural network enabled space group identification in EBSD. *Microscopy and Microanalysis* **2020**, *26*, 447–457.
61. Kaufmann, K.; Zhu, C.; Rosengarten, A.S.; Maryanovsky, D.; Wang, H.; Vecchio, K.S. Phase Mapping in EBSD using convolutional neural networks. *Microscopy and Microanalysis* **2020**, *26*, 458–468.
62. Kaufmann, K.; Lane, H.; Liu, X.; Vecchio, K.S. Efficient few-shot machine learning for classification of EBSD patterns. *Scientific reports* **2021**, *11*, 1–12.
63. Orme, A.D.; Chelladurai, I.; Rampton, T.M.; Fullwood, D.T.; Khosravani, A.; Miles, M.P.; Mishra, R.K. Insights into twinning in Mg AZ31: A combined EBSD and machine learning study. *Computational Materials Science* **2016**, *124*, 353–363.
64. Zhang, Y.; Ngan, A.H. Extracting dislocation microstructures by deep learning. *Int. J. Plast.* **2019**, *115*, 18–28.
65. Madsen, J.; Liu, P.; Kling, J.; Wagner, J.B.; Hansen, T.W.; Winther, O.; Schiøtz, J. A deep learning approach to identify local structures in atomic-resolution transmission electron microscopy images. *Advanced Theory and Simulations* **2018**, *1*, 1800037.
66. Lee, B.; Yoon, S.; Lee, J.W.; Kim, Y.; Chang, J.; Yun, J.; Ro, J.C.; Lee, J.S.; Lee, J.H. Statistical characterization of the morphologies of nanoparticles through machine learning based electron microscopy image analysis. *ACS nano* **2020**, *14*, 17125–17133.
67. Horwath, J.P.; Zakharov, D.N.; Mégrét, R.; Stach, E.A. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *npj Computational Materials* **2020**, *6*, 1–9.
68. Ziatdinov, M.; Dyck, O.; Maksov, A.; Li, X.; Sang, X.; Xiao, K.; Unocic, R.R.; Vasudevan, R.; Jesse, S.; Kalinin, S.V. Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations. *ACS nano* **2017**, *11*, 12742–12752.
69. Cho, P.; Wood, A.; Mahalingam, K.; Eyink, K. Defect Detection in Atomic Resolution Transmission Electron Microscopy Images Using Machine Learning. *Mathematics* **2021**, *9*, 1209.
70. Papanikolaou, S.; Cui, Y.; Ghoniem, N. Avalanches and plastic flow in crystal plasticity: an overview. *Modelling and Simulation in Materials Science and Engineering* **2017**, *26*, 013001.
71. Yamamoto, Y.; Hattori, M.; Ohyama, J.; Satsuma, A.; Tanaka, N.; Muto, S. Twinned/untwinned catalytic gold nanoparticles identified by applying a convolutional neural network to their Hough transformed Z-contrast images. *Microscopy* **2018**, *67*, 321–330.
72. Dan, J.; Zhao, X.; Pennycook, S.J. A machine perspective of atomic defects in scanning transmission electron microscopy. *InfoMat* **2019**, *1*, 359–375.
73. Kalinin, S.V.; Lupini, A.R.; Dyck, O.; Jesse, S.; Ziatdinov, M.; Vasudevan, R.K. Lab on a beam—Big data and artificial intelligence in scanning transmission electron microscopy. *MRS Bulletin* **2019**, *44*, 565–575.
74. Was, G.S. *Fundamentals of radiation materials science: metals and alloys*; Springer, 2017.
75. Li, W.; Field, K.G.; Morgan, D. Automated defect analysis in electron microscopic images. *npj Computational Materials* **2018**, *4*, 1–9.
76. Shen, M.; Li, G.; Wu, D.; Yaguchi, Y.; Haley, J.C.; Field, K.G.; Morgan, D. A deep learning based automatic defect analysis framework for In-situ TEM ion irradiations. *Comput. Mater. Sci.* **2021**, *197*, 110560.
77. Wu, Z.H.; Bai, J.J.; Zhang, D.D.; Huang, G.; Zhu, T.B.; Chang, X.J.; Liu, R.D.; Lin, J.; Sun, J.A. Statistical analysis of helium bubbles in transmission electron microscopy images based on machine learning method. *Nuclear Science and Techniques* **2021**, *32*, 1–11. doi:https://doi.org/10.1007/s41365-021-00886-y.
78. Chun, M.; Biglou, J.; Lenard, J.; Kim, J. Using neural networks to predict parameters in the hot working of aluminum alloys. *Journal of Materials Processing Technology* **1999**, *86*, 245–251.
79. Haghdadi, N.; Zarei-Hanzaki, A.; Khalesian, A.; Abedi, H. Artificial neural network modeling to predict the hot deformation behavior of an A356 aluminum alloy. *Materials & Design* **2013**, *49*, 386–391.

80. Ji, G.; Li, F.; Li, Q.; Li, H.; Li, Z. A comparative study on Arrhenius-type constitutive model and artificial neural network model to predict high-temperature deformation behaviour in Aermet100 steel. *Materials Science and Engineering: A* **2011**, 528, 4774–4782.
81. Mandal, S.; Sivaprasad, P.; Venugopal, S.; Murthy, K. Artificial neural network modeling to evaluate and predict the deformation behavior of stainless steel type AISI 304L during hot torsion. *Applied Soft Computing* **2009**, 9, 237–244.
82. Sun, Y.; Zeng, W.; Zhao, Y.; Qi, Y.; Ma, X.; Han, Y. Development of constitutive relationship model of Ti600 alloy using artificial neural network. *Computational Materials Science* **2010**, 48, 686–691.
83. Ozerdem, M.S.; Kolukisa, S. Artificial Neural Network approach to predict mechanical properties of hot rolled, nonresulfurized, AISI 10xx series carbon steel bars. *Journal of Materials Processing Technology* **2008**, 199, 437–439.
84. Ozerdem, M.S.; Kolukisa, S. Artificial neural network approach to predict the mechanical properties of Cu–Sn–Pb–Zn–Ni cast alloys. *Materials & Design* **2009**, 30, 764–769.
85. Guo, Z.; Sha, W. Modelling the correlation between processing parameters and properties of maraging steels using artificial neural network. *Computational Materials Science* **2004**, 29, 12–28.
86. Malinov, S.; Sha, W.; McKeown, J. Modelling the correlation between processing parameters and properties in titanium alloys using artificial neural network. *Computational materials science* **2001**, 21, 375–394.
87. Thankachan, T.; Prakash, K.S.; Pleass, C.D.; Rammasamy, D.; Prabakaran, B.; Jothi, S. Artificial neural network to predict the degraded mechanical properties of metallic materials due to the presence of hydrogen. *international journal of hydrogen energy* **2017**, 42, 28612–28621.
88. Chatterjee, S.; Muruganath, M.; Bhadeshia, H. δ TRIP steel. *Materials Science and Technology* **2007**, 23, 819–827.
89. Mahfouf, M.; Jamei, M.; Linkens, D. Optimal design of alloy steels using multiobjective genetic algorithms. *Materials and Manufacturing processes* **2005**, 20, 553–567.
90. Arceri, F.; Landes, F.P.; Berthier, L.; Biroli, G. Glasses and aging: a statistical mechanics perspective. *arXiv preprint arXiv:2006.09725* **2020**.
91. Cubuk, E.D.; Schoenholz, S.S.; Kaxiras, E.; Liu, A.J. Structural properties of defects in glassy liquids. *The Journal of Physical Chemistry B* **2016**, 120, 6139–6146.
92. Royall, C.P.; Kob, W. Locally favoured structures and dynamic length scales in a simple glass-former. *Journal of Statistical Mechanics: Theory and Experiment* **2017**, 2017, 024001.
93. Boattini, E.; Marín-Aguilar, S.; Mitra, S.; Foffi, G.; Smallenburg, F.; Filion, L. Autonomously revealing hidden local structures in supercooled liquids. *Nature communications* **2020**, 11, 1–9.
94. Schoenholz, S.S.; Cubuk, E.D.; Kaxiras, E.; Liu, A.J. Relationship between local structure and relaxation in out-of-equilibrium glassy systems. *Proceedings of the National Academy of Sciences* **2017**, 114, 263–267.
95. Hufnagel, T.C.; Schuh, C.A.; Falk, M.L. Deformation of metallic glasses: Recent developments in theory, simulations, and experiments. *Acta Materialia* **2016**, 109, 375–393.
96. Ding, J.; Cheng, Y.; Ma, E. Correlating local structure with inhomogeneous elastic deformation in a metallic glass. *Applied Physics Letters* **2012**, 101, 121917.
97. Fan, Y.; Iwashita, T.; Egami, T. Evolution of elastic heterogeneity during aging in metallic glasses. *Physical Review E* **2014**, 89, 062313.
98. Ding, J.; Cheng, Y.Q.; Ma, E. Full icosahedra dominate local order in Cu₆₄Zr₃₄ metallic glass and supercooled liquid. *Acta materialia* **2014**, 69, 343–354.
99. Ding, J.; Patinet, S.; Falk, M.L.; Cheng, Y.; Ma, E. Soft spots and their structural signature in a metallic glass. *Proceedings of the National Academy of Sciences* **2014**, 111, 14052–14056.
100. Ma, E.; Ding, J. Tailoring structural inhomogeneities in metallic glasses to enable tensile ductility at room temperature. *Materials Today* **2016**, 19, 568–579.
101. Schoenholz, S.S. Combining machine learning and physics to understand glassy systems. *Journal of Physics: Conference Series*. IOP Publishing, 2018, Vol. 1036, p. 012021.
102. Schoenholz, S.S.; Cubuk, E.D.; Sussman, D.M.; Kaxiras, E.; Liu, A.J. A structural approach to relaxation in glassy liquids. *Nature Physics* **2016**, 12, 469–471.
103. Cubuk, E.D.; Liu, A.J.; Kaxiras, E.; Schoenholz, S.S. Unifying framework for strong and fragile liquids via machine learning: a study of liquid silica. *arXiv preprint arXiv:2008.09681* **2020**.
104. Cubuk, E.D.; Schoenholz, S.S.; Rieser, J.M.; Malone, B.D.; Rottler, J.; Durian, D.J.; Kaxiras, E.; Liu, A.J. Identifying structural flow defects in disordered solids using machine-learning methods. *Physical review letters* **2015**, 114, 108001.
105. Cubuk, E.D.; Ivancic, R.; Schoenholz, S.S.; Strickland, D.; Basu, A.; Davidson, Z.; Fontaine, J.; Hor, J.L.; Huang, Y.R.; Jiang, Y.; others. Structure-property relationships from universal signatures of plasticity in disordered solids. *Science* **2017**, 358, 1033–1037.
106. Rocks, J.W.; Ridout, S.A.; Liu, A.J. Learning-based approach to plasticity in athermal sheared amorphous packings: Improving softness. *APL Materials* **2021**, 9, 021107.
107. Wang, Q.; Jain, A. A transferable machine-learning framework linking interstice distribution and plastic heterogeneity in metallic glasses. *Nature communications* **2019**, 10, 1–11.
108. Lim, H.; Carroll, J.; Battaile, C.C.; Buchheit, T.; Boyce, B.; Weinberger, C. Grain-scale experimental validation of crystal plasticity finite element simulations of tantalum oligocrystals. *Int. J. Plast.* **2014**, 60, 1–18.

109. Guan, Y.; Chen, B.; Zou, J.; Britton, T.B.; Jiang, J.; Dunne, F.P. Crystal plasticity modelling and HR-DIC measurement of slip activation and strain localization in single and oligo-crystal Ni alloys under fatigue. *Int. J. Plast.* **2017**, *88*, 70–88.
110. Kapoor, K.; Sangid, M.D. Initializing type-2 residual stresses in crystal plasticity finite element simulations utilizing high-energy diffraction microscopy data. *Materials Science and Engineering: A* **2018**, *729*, 53–63.
111. Bandyopadhyay, R.; Mello, A.W.; Kapoor, K.; Reinhold, M.P.; Broderick, T.F.; Sangid, M.D. On the crack initiation and heterogeneous deformation of Ti-6Al-4V during high cycle fatigue at high R ratios. *J. Mech. Phys. Solids* **2019**, *129*, 61–82.
112. Zhang, T.; Collins, D.M.; Dunne, F.P.; Shollock, B.A. Crystal plasticity and high-resolution electron backscatter diffraction analysis of full-field polycrystal Ni superalloy strains and rotations under thermal loading. *Acta materialia* **2014**, *80*, 25–38.
113. Kartal, M.; Kiwanuka, R.; Dunne, F. Determination of sub-surface stresses at inclusions in single crystal superalloy using HR-EBSD, crystal plasticity and inverse eigenstrain analysis. *International Journal of Solids and Structures* **2015**, *67*, 27–39.
114. Rovinelli, A.; Guilhem, Y.; Proudhon, H.; Lebensohn, R.A.; Ludwig, W.; Sangid, M.D. Assessing reliability of fatigue indicator parameters for small crack growth via a probabilistic framework. *Modelling and Simulation in Materials Science and Engineering* **2017**, *25*, 045010.
115. Rovinelli, A.; Sangid, M.D.; Proudhon, H.; Guilhem, Y.; Lebensohn, R.A.; Ludwig, W. Predicting the 3D fatigue crack growth rate of small cracks using multimodal data via Bayesian networks: In-situ experiments and crystal plasticity simulations. *Journal of the Mechanics and Physics of Solids* **2018**, *115*, 208–229.
116. Miyazawa, Y.; Briffod, F.; Shiraiwa, T.; Enoki, M. Prediction of cyclic stress-strain property of steels by crystal plasticity simulations and machine learning. *Materials* **2019**, *12*, 3668.
117. Ali, U.; Muhammad, W.; Brahme, A.; Skiba, O.; Inal, K. Application of artificial neural networks in micromechanics for polycrystalline metals. *Int. J. Plast.* **2019**, *120*, 205–219.
118. Kalidindi, S.R.; Niezgoda, S.R.; Landi, G.; Vachhani, S.; Fast, T. A Novel Framework for Building Materials Knowledge Systems. *Computers, Materials & Continua* **2010**, *17*, 103–126.
119. Brough, D.B.; Wheeler, D.; Kalidindi, S.R. Materials knowledge systems in python—a data science framework for accelerated development of hierarchical materials. *Integrating materials and manufacturing innovation* **2017**, *6*, 36.
120. Paulson, N.H.; Priddy, M.W.; McDowell, D.L.; Kalidindi, S.R. Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics. *Acta Mat.* **2017**, *129*, 428–438.
121. Priddy, M.W.; Paulson, N.H.; Kalidindi, S.R.; McDowell, D.L. Strategies for rapid parametric assessment of microstructure-sensitive fatigue for HCP polycrystals. *International Journal of Fatigue* **2017**, *104*, 231–242.
122. Paulson, N.H.; Priddy, M.W.; McDowell, D.L.; Kalidindi, S.R. Data-driven reduced-order models for rank-ordering the high cycle fatigue performance of polycrystalline microstructures. *Materials & Design* **2018**, *154*, 170–183.
123. Paulson, N.H.; Priddy, M.W.; McDowell, D.L.; Kalidindi, S.R. Reduced-order microstructure-sensitive protocols to rank-order the transition fatigue resistance of polycrystalline microstructures. *International Journal of Fatigue* **2019**, *119*, 1–10.
124. Kubin, L.; Canova, G. The modelling of dislocation patterns. *Scripta Metallurgica et Materialia* **1992**, *27*, 957–962. doi:https://doi.org/10.1016/0956-716X(92)90456-O.
125. Devincere, B.; Kubin, L. Mesoscopic simulations of dislocations and plasticity. *Mater. Sci. Eng. A* **1997**, *234-236*, 8–14. doi:10.1016/S0921-5093(97)00146-9.
126. Hirth, J.; Lothe, J. *Theory of Dislocations*; Krieger Publishing Company, 1982.
127. Ananthakrishna, G. Chapter 73 Statistical and Dynamical Approaches to Collective Behavior of Dislocations. In *Dislocations in Solids*; Nabarro, F.; Hirth, J., Eds.; Elsevier, 2007; Vol. 13, pp. 81–223. doi:https://doi.org/10.1016/S1572-4859(07)80005-2.
128. Fivel, M.C. Discrete dislocation dynamics: an important recent break-through in the modelling of dislocation collective behaviour. *Comptes Rendus Physique* **2008**, *9*, 427–436. doi:10.1016/j.crhy.2007.11.005.
129. Lefebvre, S.; Devincere, B.; Hoc, T. Yield stress strengthening in ultrafine-grained metals: A two-dimensional simulation of dislocation dynamics. *Journal of the Mechanics and Physics of Solids* **2007**, *55*, 788–802. doi:https://doi.org/10.1016/j.jmps.2006.10.002.
130. Kumar, R.; Nicola, L.; Van der Giessen, E. Density of grain boundaries and plasticity size effects: A discrete dislocation dynamics study. *Materials Science and Engineering A* **2009**, *527*, 7–15. doi:10.1016/j.msea.2009.08.072.
131. Zhou, C.; Biner, S.B.; LeSar, R. Discrete dislocation dynamics simulations of plasticity at small scales. *Acta Materialia* **2010**, *58*, 1565–1577. doi:https://doi.org/10.1016/j.actamat.2009.11.001.
132. Han, C.S.; Hartmaier, A.; Gao, H.; Huang, Y. Discrete dislocation dynamics simulations of surface induced size effects in plasticity. *Materials Science and Engineering: A* **2006**, *415*, 225–233. doi:https://doi.org/10.1016/j.msea.2005.09.075.
133. Papanikolaou, S.; Dimiduk, D.M.; Choi, W.; Sethna, J.P.; Uchic, M.D.; Woodward, C.F.; Zapperi, S. Quasi-periodic events in crystal plasticity and the self-organized avalanche oscillator. *Nature* **2012**, *490*, 517–521.
134. Song, H.; Papanikolaou, S. From statistical correlations to stochasticity and size effects in sub-micron crystal plasticity. *Metals* **2019**, *9*, 835.
135. Song, H.; Dimiduk, D.; Papanikolaou, S. Universality class of nanocrystal plasticity: localization and self-organization in discrete dislocation dynamics. *Physical review letters* **2019**, *122*, 178001.
136. Madec, R.; Devincere, B.; Kubin, L.; Hoc, T.; Rodney, D. The role of collinear interaction in dislocation-induced hardening. *Science* **2003**, *301*, 1879–1882. doi:10.1126/science.1085477.
137. Bulatov, V.; Abraham, F.F.; Kubin, L.; Devincere, B.; Yip, S. Connecting atomistic and mesoscale simulations of crystal plasticity. *Nature* **1998**, *391*, 669–672. doi:10.1038/35577.

138. Papanikolaou, S.; Po, G. Λ -Invariant and Topological Pathways to Influence the Strength of Submicron Crystals. *Physical Review Letters* **2020**, *124*, 205502.
139. Akhondzadeh, S.; Sills, R.; Papanikolaou, S.; Van der Giessen, E.; Cai, W. Geometrically projected discrete dislocation dynamics. *Modelling and Simulation in Materials Science and Engineering* **2018**, *26*, 065011.
140. Csikor, F.F.; Motz, C.; Weygand, D.; Zaiser, M.; Zapperi, S. Dislocation avalanches, strain bursts, and the problem of plastic forming at the micrometer scale. *Science* **2007**, *318*, 251–254. doi:10.1126/science.1143719.
141. Devincre, B.; Hoc, T.; Kubin, L. Dislocation Mean Free Paths and Strain Hardening of Crystals. *Science* **2008**, *320*, 1745–1748, [<https://science.sciencemag.org/content/320/5884/1745.full.pdf>]. doi:10.1126/science.1156101.
142. Chen, Y.J.; Papanikolaou, S.; Sethna, J.P.; Zapperi, S.; Durin, G. Avalanche spatial structure and multivariable scaling functions: Sizes, heights, widths, and views through windows. *Physical Review E* **2011**, *84*, 061103.
143. Bolin, R.; Yavas, H.; Song, H.; Hemker, K.J.; Papanikolaou, S. Bending nanoindentation and plasticity noise in FCC single and polycrystals. *Crystals* **2019**, *9*, 652.
144. Song, H.; Yavas, H.; Van der Giessen, E.; Papanikolaou, S. Discrete dislocation dynamics simulations of nanoindentation with pre-stress: Hardness and statistics of abrupt plastic events. *Journal of the Mechanics and Physics of Solids* **2019**, *123*, 332–347.
145. Ni, X.; Papanikolaou, S.; Vajente, G.; Adhikari, R.X.; Greer, J.R. Probing microplasticity in small-scale fcc crystals via dynamic mechanical analysis. *Physical review letters* **2017**, *118*, 155501.
146. Bertin, N.; Sills, R.B.; Cai, W. Frontiers in the Simulation of Dislocations. *Annual Review of Materials Research* **2020**, *50*, 437–464. doi:10.1146/annurev-matsci-091819-015500.
147. Prakash, A.; Sandfeld, S. Chances and Challenges in Fusing Data Science with Materials Science. *Practical Metallography* **2018**, *55*, 493–514. doi:doi:10.3139/147.110539.
148. Salmenjoki, H.; Alava, M.J.; Laurson, L. Machine learning plastic deformation of crystals. *Nature Communications* **2018**, *9*, 5307. doi:10.1038/s41467-018-07737-2.
149. Chen, Y.S.; Choi, W.; Papanikolaou, S.; Sethna, J.P. Bending crystals: emergence of fractal dislocation structures. *Physical review letters* **2010**, *105*, 105501.
150. Chen, Y.S.; Choi, W.; Papanikolaou, S.; Bierbaum, M.; Sethna, J.P. Scaling theory of continuum dislocation dynamics in three dimensions: Self-organized fractal pattern formation. *International Journal of Plasticity* **2013**, *46*, 94–129.
151. Yang, Z.; Papanikolaou, S.; Reid, A.C.; keng Liao, W.; Choudhary, A.N.; Campbell, C.; Agrawal, A. Learning to Predict Crystal Plasticity at the Nanoscale: Deep Residual Networks and Size Effects in Uniaxial Compression Discrete Dislocation Simulations. *Scientific Reports* **2020**, *10*, 1–14. doi:10.1038/s41598-020-65157-z.
152. Steinberger, D.; Song, H.; Sandfeld, S. Machine Learning-Based Classification of Dislocation Microstructures. *Frontiers in Materials* **2019**, *6*, 141. doi:10.3389/fmats.2019.00141.
153. Messner, M.C.; Rhee, M.; Arsenlis, A.; Barton, N.R. A crystal plasticity model for slip in hexagonal close packed metals based on discrete dislocation simulations. *Modelling and Simulation in Materials Science and Engineering* **2017**, *25*, 044001. doi:10.1088/1361-651x/aa687a.
154. Espinosa, H.D.; Panico, M.; Berbenni, S.; Schwarz, K.W. Discrete dislocation dynamics simulations to interpret plasticity size and surface effects in freestanding FCC thin films. *International Journal of Plasticity* **2006**, *22*, 2091–2117. doi:10.1016/j.ijplas.2006.01.007.
155. Wang, Z.; McCabe, R.J.; Ghoniem, N.M.; LeSar, R.; Misra, A.; Mitchell, T.E. Dislocation motion in thin Cu foils: A comparison between computer simulations and experiment. *Acta Materialia* **2004**, *52*, 1535–1542. doi:10.1016/j.actamat.2003.12.005.
156. Weygand, D.; Poignant, M.; Gumbsch, P.; Kraft, O. Three-dimensional dislocation dynamics simulation of the influence of sample size on the stress-strain behavior of fcc single-crystalline pillars. *Materials Science and Engineering A* **2008**, *483–484*, 188–190. doi:10.1016/j.msea.2006.09.183.
157. Senger, J.; Weygand, D.; Gumbsch, P.; Kraft, O. Discrete dislocation simulations of the plasticity of micro-pillars under uniaxial loading. *Scripta Materialia* **2008**, *58*, 587–590. doi:10.1016/j.scriptamat.2007.11.031.
158. Akarapu, S.; Zbib, H.M.; Bahr, D.F. Analysis of heterogeneous deformation and dislocation dynamics in single crystal micropillars under compression. *International Journal of Plasticity* **2010**, *26*, 239–257. doi:10.1016/j.ijplas.2009.06.005.
159. Zhang, J.L.; Zaefferer, S.; Raabe, D. A study on the geometry of dislocation patterns in the surrounding of nanoindentations in a TWIP steel using electron channeling contrast imaging and discrete dislocation dynamics simulations. *Materials Science and Engineering A* **2015**, *636*, 231–242. doi:10.1016/j.msea.2015.03.078.
160. Sangid, M.D. Coupling in situ experiments and modeling—Opportunities for data fusion, machine learning, and discovery of emergent behavior. *Current Opinion in Solid State and Materials Science* **2020**, *24*, 100797.
161. Vishnubhotla, S.B.; Chen, R.; Khanal, S.R.; Hu, X.; Martini, A.; Jacobs, T.D.B. Matching Atomistic Simulations and In Situ Experiments to Investigate the Mechanics of Nanoscale Contact. *Tribology Letters* **2019**, *67*, 97.
162. Wang, J.; Cao, G.; Zhang, Z.; Sansoz, F. Size-dependent dislocation-twin interactions. *Nanoscale* **2019**, *11*, 12672–12679. doi:10.1039/C9NR03637G.
163. Hata, S.; Honda, T.; Saito, H.; Mitsuhashi, M.; Petersen, T.; Murayama, M. Electron tomography: An imaging method for materials deformation dynamics. *Current Opinion in Solid State and Materials Science* **2020**, *24*, 100850. doi:<https://doi.org/10.1016/j.cossms.2020.100850>
164. Chowdhury, A.; Kautz, E.; Yener, B.; Lewis, D. Image driven machine learning methods for microstructure recognition. *Computational Materials Science* **2016**, *123*, 176–187. doi:<https://doi.org/10.1016/j.commatsci.2016.05.034>.

165. Graham, R.; Simon, Y.H.; Sainju, R.; Danny J. Edwards, B.H.Y.Z. Deep Learning for Semantic Segmentation of Defects in Advanced STEM Images of Steels. *Sci. Rep.* **2019**, *9*, 12744.
166. Yeh, J.W.; Chen, S.K.; Lin, S.J.; Gan, J.Y.; Chin, T.S.; Shun, T.T.; Tsau, C.H.; Chang, S.Y. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Advanced Engineering Materials* **2004**, *6*, 299–303.
167. Cantor, B.; Chang, I.; Knight, P.; Vincent, A. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **2004**, *375*, 213–218.
168. Flory, P.J. Phase equilibria in solutions of rod-like particles. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1956**, *234*, 73–89.
169. Onsager, L. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences* **1949**, *51*, 627–659.
170. Bernal, J.D. A geometrical approach to the structure of liquids. *Nature* **1959**, *183*, 141–147.
171. Tsuzuki, H.; Branicio, P.S.; Rino, J.P. Structural characterization of deformed crystals by analysis of common atomic neighborhood. *Computer physics communications* **2007**, *177*, 518–523.
172. Honeycutt, J.D.; Andersen, H.C. Molecular dynamics study of melting and freezing of small Lennard-Jones clusters. *Journal of Physical Chemistry* **1987**, *91*, 4950–4963.
173. Blaisten-Barojas, E.; Andersen, H.C. Effects of three-body interactions on the structure of clusters. *Surface Science* **1985**, *156*, 548–555.
174. Kelchner, C.L.; Plimpton, S.; Hamilton, J. Dislocation nucleation and defect structure during surface indentation. *Physical review B* **1998**, *58*, 11085.
175. Li, J. AtomEye: an efficient atomistic configuration viewer. *Modelling and Simulation in Materials Science and Engineering* **2003**, *11*, 173.
176. Li, J. Central symmetry parameter, 2003.
177. Posada-Amarillas, A.; Garzón, I.L. Microstructural analysis of simulated liquid and amorphous Ni. *Physical Review B* **1996**, *53*, 8363.
178. Schiötz, J.; Di Tolla, F.D.; Jacobsen, K.W. Softening of nanocrystalline metals at very small grain sizes. *Nature* **1998**, *391*, 561–563.
179. Schiötz, J.; Vegge, T.; Di Tolla, F.; Jacobsen, K.W. Atomic-scale simulations of the mechanical deformation of nanocrystalline metals. *Physical Review B* **1999**, *60*, 11971.
180. Yavari, A.R. A new order for metallic glasses. *Nature* **2006**, *439*, 405–406.
181. Van Swygenhoven, H.; Weertman, J.R. Deformation in nanocrystalline metals. *Materials today* **2006**, *9*, 24–31.
182. Monk, J.; Farkas, D. Strain-induced grain growth and rotation in nickel nanowires. *Physical Review B* **2007**, *75*, 045414.
183. Bringa, E.; Rosolankova, K.; Rudd, R.; Remington, B.; Wark, J.; Duchaineau, M.; Kalantar, D.; Hawreliak, J.; Belak, J. Shock deformation of face-centred-cubic metals on subnanosecond timescales. *Nature materials* **2006**, *5*, 805–809.
184. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les paralléloèdres primitifs. *Journal für die reine und angewandte Mathematik (Crelles Journal)* **1908**, *1908*, 198–287.
185. Stukowski, A. Structure identification methods for atomistic simulations of crystalline materials. *Modelling and Simulation in Materials Science and Engineering* **2012**, *20*, 045021.
186. Lazar, E.A. VoroTop: Voronoi cell topology visualization and analysis toolkit. *Modelling and Simulation in Materials Science and Engineering* **2017**, *26*, 015011.
187. Lazar, E.A.; Han, J.; Srolovitz, D.J. Topological framework for local structure analysis in condensed matter. *Proceedings of the National Academy of Sciences* **2015**, *112*, E5769–E5776.
188. Trybula, M.E. Structure and transport properties of the liquid Al₈₀Cu₂₀ alloy—a molecular dynamics study. *Computational Materials Science* **2016**, *122*, 341–352.
189. Trybula, M.E.; Szafranski, P.W.; Korzhavyi, P.A. Structure and chemistry of liquid Al–Cu alloys: molecular dynamics study versus thermodynamics-based modelling. *Journal of materials science* **2018**, *53*, 8285–8301.
190. Wang, W.; Han, J.; Fang, H.; Wang, J.; Liang, Y.; Shang, S.; Wang, Y.; Liu, X.; Kecskes, L.; Mathaudhu, S.; others. Anomalous structural dynamics in liquid Al₈₀Cu₂₀: An ab initio molecular dynamics study. *Acta Materialia* **2015**, *97*, 75–85.
191. Du, Q.; Liu, X.; Fan, H.; Zeng, Q.; Wu, Y.; Wang, H.; Chatterjee, D.; Ren, Y.; Ke, Y.; Voyles, P.M.; others. Reentrant glass transition leading to ultrastable metallic glass. *Materials today* **2020**, *34*, 66–77.
192. Nagamanasa, K.H.; Gokhale, S.; Ganapathy, R.; Sood, A. Confined glassy dynamics at grain boundaries in colloidal crystals. *Proceedings of the National Academy of Sciences* **2011**, *108*, 11323–11326.
193. LaViolette, R.A.; Stillinger, F.H. Enumeration of random packings for atomic substances. *Physical Review B* **1987**, *35*, 5446.
194. Stillinger, F.H.; Weber, T.A. Hidden structure in liquids. *Physical Review A* **1982**, *25*, 978.
195. Oh, H.S.; Kim, S.J.; Obadraakh, K.; Ryu, W.H.; Yoon, K.N.; Mu, S.; Körmann, F.; Ikeda, Y.; Tasan, C.C.; Raabe, D.; others. Engineering atomic-level complexity in high-entropy and complex concentrated alloys. *Nature communications* **2019**, *10*, 1–8.
196. Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Physical Review B* **2011**, *83*, 153101.
197. Artrith, N.; Urban, A.; Ceder, G. Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. *Physical Review B* **2017**, *96*, 014112.

198. Hume-Rothery, W.; Powell, H.M. On the theory of super-lattice structures in alloys. *Zeitschrift für Kristallographie-Crystalline Materials* **1935**, *91*, 23–47.
199. Hume-Rothery, W.; Raynor, G.V. *The structure of metals and alloys*; Institute of metals, 1936.
200. Zhang, Y.; Zhou, Y.J.; Lin, J.P.; Chen, G.L.; Liaw, P.K. Solid-solution phase formation rules for multi-component alloys. *Advanced engineering materials* **2008**, *10*, 534–538.
201. Troparevsky, M.C.; Morris, J.R.; Kent, P.R.; Lupini, A.R.; Stocks, G.M. Criteria for predicting the formation of single-phase high-entropy alloys. *Physical Review X* **2015**, *5*, 011041.
202. George, E.P.; Raabe, D.; Ritchie, R.O. High-entropy alloys. *Nature Reviews Materials* **2019**, *4*, 515–534.
203. Inoue, A. Stabilization of metallic supercooled liquid and bulk amorphous alloys. *Act. Mater.* **2000**, *48*, 279–306.
204. Wang, D.; Li, Y.; Sun, B.; Sui, M.; Lu, K.; Ma, E. Bulk metallic glass formation in the binary Cu–Zr system. *Applied Physics Letters* **2004**, *84*, 4029–4031.
205. Xu, D.; Lohwongwatana, B.; Duan, G.; Johnson, W.L.; Garland, C. Bulk metallic glass formation in binary Cu-rich alloy series–Cu100–xZrx (x= 34, 36, 38.2, 40 at.%) and mechanical properties of bulk Cu64Zr36 glass. *Acta Materialia* **2004**, *52*, 2621–2624.
206. Pei, Z.; Yin, J.; Hawk, J.A.; Alman, D.E.; Gao, M.C. Machine-learning informed prediction of high-entropy solid solution formation: Beyond the Hume-Rothery rules. *npj Computational Materials* **2020**, *6*, 1–8.
207. Xiong, J.; Shi, S.; Zhang, T. Machine learning prediction of glass-forming ability in bulk metallic glasses. *Computational Materials Science* **2021**, *192*, 110362.
208. Ren, R.; Hung, T.; Tan, K.C. A generic deep-learning-based approach for automated surface inspection. *IEEE transactions on cybernetics* **2017**, *48*, 929–940.
209. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; others. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials* **2013**, *1*, 011002.
210. Pei, Z.; Yin, J. Machine learning as a contributor to physics: Understanding Mg alloys. *Materials & Design* **2019**, *172*, 107759.
211. Kostuchenko, T.; Körmann, F.; Neugebauer, J.; Shapeev, A. Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials. *npj Computational Materials* **2019**, *5*, 1–7.
212. Abu-Odeh, A.; Galvan, E.; Kirk, T.; Mao, H.; Chen, Q.; Mason, P.; Malak, R.; Arróyave, R. Efficient exploration of the High Entropy Alloy composition-phase space. *Acta Materialia* **2018**, *152*, 41–57.
213. Huang, W.; Martin, P.; Zhuang, H.L. Machine-learning phase prediction of high-entropy alloys. *Acta Materialia* **2019**, *169*, 225–236.
214. Kajita, S.; Ohba, N.; Jinnouchi, R.; Asahi, R. A universal 3D voxel descriptor for solid-state material informatics with deep convolutional neural networks. *Scientific reports* **2017**, *7*, 1–9.
215. Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M.E.; Burke, K.; Müller, K.R. Bypassing the Kohn-Sham equations with machine learning. *Nature communications* **2017**, *8*, 1–10.
216. Nagai, R.; Akashi, R.; Sasaki, S.; Tsuneyuki, S. Neural-network Kohn-Sham exchange-correlation potential and its out-of-training transferability. *The Journal of chemical physics* **2018**, *148*, 241737.
217. Snyder, J.C.; Rupp, M.; Hansen, K.; Müller, K.R.; Burke, K. Finding density functionals with machine learning. *Physical review letters* **2012**, *108*, 253002.
218. Nelson, J.; Tiwari, R.; Sanvito, S. Machine learning density functional theory for the Hubbard model. *Physical Review B* **2019**, *99*, 075132.
219. Mills, K.; Spanner, M.; Tamblyn, I. Deep learning and the Schrödinger equation. *Physical Review A* **2017**, *96*, 042113.
220. Nie, X.; Chien, P.; Morgan, D.; Kaczmarowski, A. A statistical method for emulation of computer models with invariance-preserving properties, with application to structural energy prediction. *Journal of the American Statistical Association* **2020**, *115*, 1798–1811.
221. Horsch, M.T.; Chiacchiera, S.; Cavalcanti, W.L.; Schembera, B. *Data Technology in Materials Modelling*; Springer, 2021.
222. Forte, E.; Jirasek, F.; Bortz, M.; Burger, J.; Vrabec, J.; Hasse, H. Digitalization in thermodynamics. *Chem. Ing. Tech* **2019**, *91*, 201–214.
223. Tchipev, N.; Seckler, S.; Heinen, M.; Vrabec, J.; Gratl, F.; Horsch, M.; Bernreuther, M.; Glass, C.; Niethammer, C.; Hammer, N.; Krischok, B.; Resch, M.; Kranzlmüller, D.; Hasse, H.; Bungartz, H.; Neumann, P. TweTriS: twenty trillion-atom simulation. *Int. J. HPC Appl* **2019**, *33*, 838–854.
224. Stephan, S.; Horsch, M.; Vrabec, J.; Hasse, H. MolMod - an open access database of force fields for molecular simulations of fluids. *Mol. Sim.* **2019**, *45*, 806–814.
225. Schmitz, G. Microstructure modeling in integrated computational materials engineering (ICME) settings: can HDF5 provide the basis for an emerging standard for describing microstructures? *JOM* **2016**, *68*, 77–83.
226. Morbach, J.; Wiesner, A.; Marquardt, W. OntoCAPE 2.0: a (re-)usable ontology for computeraided process engineering. *Comput. Aid. Chem. Eng.* **2008**, *25*, 991–996.
227. Morbach, J.; Wiesner, A.; Marquardt, W. Shedding light on the dark data in the long tail of science. *Libr. Trends* **2008**, *57*, 280–299.
228. Schembera, B.; Durán, J. Dark data as the new challenge for big data science and the introduction of the scientific data officer. *Philos. Technol.* **2020**, *33*, 93–115.
229. Bicarregui, J. *Building and sustaining data infrastructures: putting policy into practice. Policy document*; Welcome, 2016.
230. Ribes, A.; Caremoli, C. Salomé platform component model for numerical simulation, in Proceedings of COMPSAC 2007. *IEEE Computer Society*, vol. 2, ed. by C.K. Chang (Los Alamitos, California, USA, 2007) **2017**, pp. 553–564.

-
231. Schneider, T.; Simkus, M. Ontologies and data management: a brief survey. *Künstl. Intell.* **2020**, *34*, 329–353.
 232. Mons, B. *Data Stewardship for Open Science*; CRC Press, Boca Raton, 2018.
 233. Torres, B.M.; Völker, C.; Nagel, S.M.; Hanke, T.; Kruschwitz, S. *An Ontology-Based Approach to Enable Data-Driven Research in the Field of NDT in Civil Engineering*; ., 2021.
 234. M. Schoop, A. de Moor, J.D. The pragmatic web: a manifesto. *Commun. ACM* **2006**, *49*, 75–76.
 235. Neiva, F.W.; David, J.; Braga, R.; Borges, M.; Campos, F. SM2PIA: a model to support the development of pragmatic interoperability requirements. *Proceedings of ICGSE 2016, ed. by D. Redmiles, M.A. Gerosa, T. Hildenbrand (IEEE, New York, USA)* **2016**, pp. 119–128.
 236. Neiva, F.W.; David, J.; Braga, R.; Campos, F. Towards pragmatic interoperability to support collaboration: a systematic review and mapping of the literature. *Inf. Softw. Technol.* **2016**, *72*, 137–150.
 237. Hey, A.; Trefethen, A. *The data deluge: an e-science perspective, in Grid Computing: Making the Global Infrastructure a Reality, ed. by F. Berman, G.C. Fox, A.J.G. Hey; Vol. 72, Wiley, 2003; pp. 809–824.*
 238. Edwards, P.; Mayernik, M.; Batcheller, A.; Bowker, G.; Borgman, C. Science friction: data, metadata, and collaboration. *Soc. Stud. Sci.* **2011**, *41*, 667–690.
 239. Greenberg, J. g, Metadata and the world wide web. *Encycl. Libr. Inf. Sci.* **2003**, *3*, 1876–1888.
 240. Li, H.; Armiento, R.; Lambrix, P. *A Method for Extending Ontologies with Application to the Materials Science Domain*; ., 2019.
 241. Neumann, J.; Brase, J. DataCite. Names for research data. *J. Comp.-Aid. Mol. Des.* **2014**, *28*, 1035–1041.
 242. Jones, M.; Boettiger, C.; Mayes, A.C.; Smith, A.; Slaughter, P.; Niemeyer, K.; Gil, Y.; Fenner, M.; Nowak, K.; Hahnel, M.; Coy, L.; Allen, A.; Crosas, M.; Sands, A.; Hong, N.; Cruse, P.; Katz, D.; Goble, C. *CodeMeta: an exchange schema for software metadata. Version 2.0. Technical report, KNB Data Repository*; ., 2017.
 243. Druskat, S.; Hong, N.; Haines, R.; Baker, J. *ation file format (CFF): specifications. Technical report, Zenodo*; ., 2018.