

Microstructure-Informed Cloud Computing for Interoperability of Materials Databases and Computational Models: Microtextured Regions in Ti Alloys

Ayman A. Salem¹ · Joshua B. Shaffer¹ · Richard A. Kublik¹ · Luke A. Wuertemberger¹ · Daniel P. Satko¹

Received: 8 December 2016 / Accepted: 17 January 2017 / Published online: 13 March 2017
© The Minerals, Metals & Materials Society 2017

Abstract With the fast global adoption of the Materials Genome Initiative (MGI), scientists and engineers are faced with the need to conduct sophisticated data analytics on large datasets to extract knowledge that can be used in modeling the behavior of materials. This raises a new problem for materials scientists: how to create and foster interoperability and share developed software tools and generated datasets. A microstructure-informed cloud-based platform (MiCloud™) has been developed that addresses this need, enabling users to easily access and insert microstructure informatics into computational tools that predict performance of engineering products by accounting for microstructural dependencies on manufacturing provenance. The platform extracts information from microstructure data by employing algorithms including signal processing, machine learning, pattern recognition, computer vision, predictive analytics, uncertainty quantification, and data visualization. The interoperability capabilities of MiCloud and its various web-based applications are demonstrated in this case study by analyzing Ti6AlV4 microstructure data via automatic identification of various features of interest

and quantifying its characteristics that are used in extracting correlations and causations for the associated mechanical behavior (e.g., yield strength, cold-dwell debit, etc.). The data were recorded by two methods: (1) backscattered electron (BSE) imaging for extracting spatial and morphological information about alpha and beta phases and (2) electron backscatter diffraction (EBSD) for extracting spatial, crystallographic, and morphological information about microtextured regions (MTRs) of the alpha phase. Extracting reliable knowledge from generated information requires data analytics of a large amount of multiscale microstructure data which necessitates the development of efficient algorithms (and the associated software tools) for data recording, analysis, and visualization. The interoperability of these tools and superior effectiveness of the cloud computing approach are validated by featuring several examples of its use in alpha/beta titanium alloys and Ni-based superalloys, reflecting the anticipated computational cost and time savings via the use of web-based applications in implementations of microstructure-informed integrated computational materials engineering (ICME).

✉ Ayman A. Salem
ayman.salem@icmrl.net

Joshua B. Shaffer
joshua.shaffer@icmrl.net

Richard A. Kublik
richard.kublik@icmrl.net

Luke A. Wuertemberger
luke.wuertemberger@icmrl.net

Daniel P. Satko
dan.satko@icmrl.net

Keywords Microtextured regions · Macrozones · Titanium · Cloud computing · ICME · MGI · Material informatics · Higher-order statistics

Background

Successful implementation of the integrated computational materials engineering (ICME) [1] approach is hinged on the integration of microstructure information into computational tools that predict the performance of engineering products and simulate its dependencies on the manufacturing processes. Towards this goal, the Materials Genome Initiative (MGI) [2] called for creation of infrastructures to couple experimental tools,

¹ Materials Resources, LLC, Dayton, OH, USA

computational tools, and digital data to make material science and engineering data more valuable [3]. In pursuit of these goals, there have been recent attempts to build cyberinfrastructure solutions that enable the sharing of materials data [4–8]. Serving the internal needs of a large materials laboratory, the Air Force Research Laboratory (AFRL) developed an integrated collaboration environment (ICE) using a mix of readily available software packages and tailored software solutions [4]. ICE's ability to directly link experimental equipment via the laboratory intranet makes it an attractive solution for well-secured government laboratories. Its use of HUBzero®'s RESTful API, a web-based application program interface, enables sharing data, research notes, tasks, and visualizations among various research groups in the laboratory [9]. While ICE is an excellent solution for AFRL, it is currently available for internal use only with limited information about its availability for global use by the ICME community.

The above example highlights one of many successful activities that focuses on collaboration and efficient sharing of materials datasets (experimental and computational) among researchers working on a variety of materials (e.g., ceramics, polymers, metals, etc.). However, there are limited examples of the interoperability and accessibility of algorithms used to extract information and knowledge from shared materials data. The use of algorithms in manufacturing and materials science (e.g., 10, 11) is not new. However, their use across the materials and manufacturing community is highly dependent on the computer-programming skills of practitioners.

The rise of open-source software [12] combined with web-based hosting services (e.g., GitHub [13]) has facilitated software sharing (e.g., MTEX [14, 15] and DREAM3D [16, 17] for EBSD data analysis). However, the material's practitioner is still expected to download, install, and run the software on the appropriate computational environment for which the software was written leading to potential interoperability and compatibility issues. Moreover, big materials data are also expected to be downloaded from the hosting server and analyzed by the downloaded materials data analytics software which causes bottlenecks with limited bandwidth. With the growth of volume, velocity, and variety of materials datasets, a need emerged to avoid the download, install, and run workflow for both datasets and the associated analytics software (i.e., implemented algorithm). This is accomplished via the use of cloud computing under a *software-as-a-service* and/or *platform-as-a-service* models (i.e., SaaS or PaaS) [18]. Believing in the importance of the underlying structure of metallic alloys (i.e., microstructure) as the linkage between processing and properties, a microstructure-informed cloud computing platform (MiCloud™) [19, 20] has been developed by the authors and deployed on both private and public servers, enabling the co-existence of on-premises and cloud computing. It includes a suite of microstructure informatics and materials data analytics tools that make data and model reuse accessible to both students

and senior scientists across the globe (e.g., [21]) via a *broad-network-access* using heterogeneous thin or thick client platforms (e.g., tablet, laptops, etc.) [18]. Consequently, interoperability, repeatability, and reproducibility of microstructure analytics have been achieved via online access for both bulk metal forming and additive manufacturing activities [21].

Keeping the focus on the importance of the interoperability of algorithms and the associated software, the following section will discuss briefly the essential components of the cloud computing platform followed by a discussion on the ICME workflow and data storage that altogether leads to efficient use of data analytics using various algorithms that are developed either by the authors or by others and the associated software implementations as shown in the case study.

Building Components of MiCloud

Building a microstructure-informed platform that enables ICME implementation in both industrial and academic environments necessitated the development of an ICME workflow. This workflow directed the development of various modules of the platform and then the composition of each module with appropriate components as explained below:

ICME Workflow and MiCloud Modules

A new workflow needed to be developed around microstructure informatics to focus on integrating mesoscale heterogeneities in materials structure with process simulation for accurate prediction of metallic part performance [22]. For horizontal scalability, the workflow utilized (i) data science protocols for efficient analysis of large microstructure datasets (e.g., cluster analysis), (ii) protocols for extracting reduced descriptions of salient microstructure features for insertion into simulations (e.g., regions of homogeneity, ROH), and (iii) protocols for direct and efficient linking of materials models/databases into process/performance simulation codes (e.g., crystal plasticity finite element method) [22]. Following that workflow, MiCloud was built on four main modules, each one containing multiple datasets/models: (1) microstructure/local state, (2) data analytics software, (3) material behavior models, and (4) finite element analysis software (Fig. 1). The workflow guided the development of various components in each module by clearly identifying the inputs and expected output of each module which resulted in identification of data management and storage needs.

The first microstructure/local state module contains a database of location-specific local states (Fig. 1) that includes both experimental datasets (e.g., electron backscatter detection (EBSD), backscatter electron (BSE) imaging, energy-dispersive spectroscopy (EDS), light optical microscopy (LOM), etc.) and simulation results (e.g., phases, stress

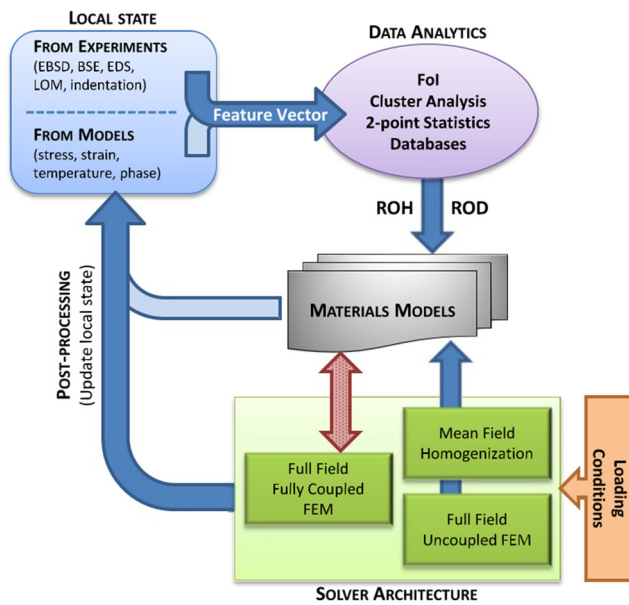


Fig. 1 A data science-centered workflow for integrating mesoscale heterogeneities in materials structure with process simulation within MiCloud [22] (Color figure online)

tensors, strain tensors, etc). The main identifier is the location-specific spatial coordinates (X, Y, Z) with additional identifiers added to the feature vector (FV) such as time or temperature. These location-specific data are used to build location-specific feature vectors. The data are mainly unstructured data which contain images, files, unstructured text, and videos. While the main content being stored does not have a defined structure, it comes packaged in objects that themselves have structure. For example, a location $i_{x,y,z}$ may have multiple BSE images at different resolutions and file formats, multiple EBSD datasets with different step sizes, and FEM simulation results with stress and strain tensors. Thus, a NoSQL database was used for storage.

The second module of MiCloud is a collection of data analytics software that implements various algorithms for pattern recognition in the FV space allowing interoperability across various datasets in the local state database. These tools include computer vision software, dimensionality reduction techniques (e.g., principle component analysis), supervised and unsupervised machine learning for cluster analysis (e.g., k-means clustering), and higher-order statistics (e.g., two-point statistics) and have the ability to implement additional tools as they are developed and added.

The third module includes a collection of materials behavior models such as Taylor-type crystal plasticity models [23], fast Fourier transform-based elastoviscoplastic model [24, 25], and an oxygen ingress in titanium model [26]. The goal of this module is to provide easy access to essential but complicated material behavior models that often require significant time and computational skills to code and compile. The development of a web-based application with a graphical user

interface (GUI) gives access from any thin or thick client platform to use microstructure data that is either already on MiCloud or uploaded by the user.

The fourth module is a collection of commercial off-the-shelf (COTS) finite element modeling software from Simufact (an MSC Software company) [27] for metal forming, metal welding, and additive manufacturing. To complete the vision of ICME discipline, a digital thread integrating microstructure, data analytics, material modeling, and the FEM solvers was established using an ICME workflow [22]. To achieve interoperability in the linking between the various crystal plasticity models and FEM, an uncoupled crystal plasticity FEM calculation (CPFEM) was initially set up with full coupling using HYPELA2, a user material subroutine in MSC Marc, which is under development.

For the uncoupled linking between FEM and crystal plasticity output, input files of FEM simulations account for the crystal plasticity-predicted yield surface that captures the anisotropic material behavior. Additionally, location-specific texture evolution is accomplished by using the evolved deformation gradient based on the particle tracking feature in Simufact. This approach enables cloud-based interoperability of crystal plasticity models with any FEM solver that provides an editable input file for inserting anisotropic yield surfaces and exporting location-specific deformation gradients. It also enables interoperability with FEM solvers running on premises at an organization by downloading the anisotropic yield surface from MiCloud, modifying the FEM input file to include it, and then uploading a location-specific deformation gradient to MiCloud for texture evaluation without revealing any intellectual property related to the setup of the FEM simulations.

To further expand the interoperability to practitioners using FEM solvers that does not export deformation gradients, but use particle-tracking or point-tracking features, the authors developed a crystal-tracking method, in which the user assigns eight particles (or points) at the corners of a cube for each location of interest. Then during the simulation, the FEM solver saves the spatial locations of each particle for each time step and then exports it in an ASCII format. Only the evolved spatial locations for the eight corners (i.e., crystal) are to be exported from the FEM solver. Following the workflow in Fig. 1 and modular programming, a web application was developed, TiCrys™, that enables the prediction of texture evolution and flow behavior of metallic alloys from simulations executed outside the cloud computing platform (Fig. 2).

Organization of Materials Datasets

Handling large materials datasets of different types is becoming one of ICME's challenges. It is typical for an ICME implementation to include microstructure images, FEM simulation results including numerous state variables on hundreds of

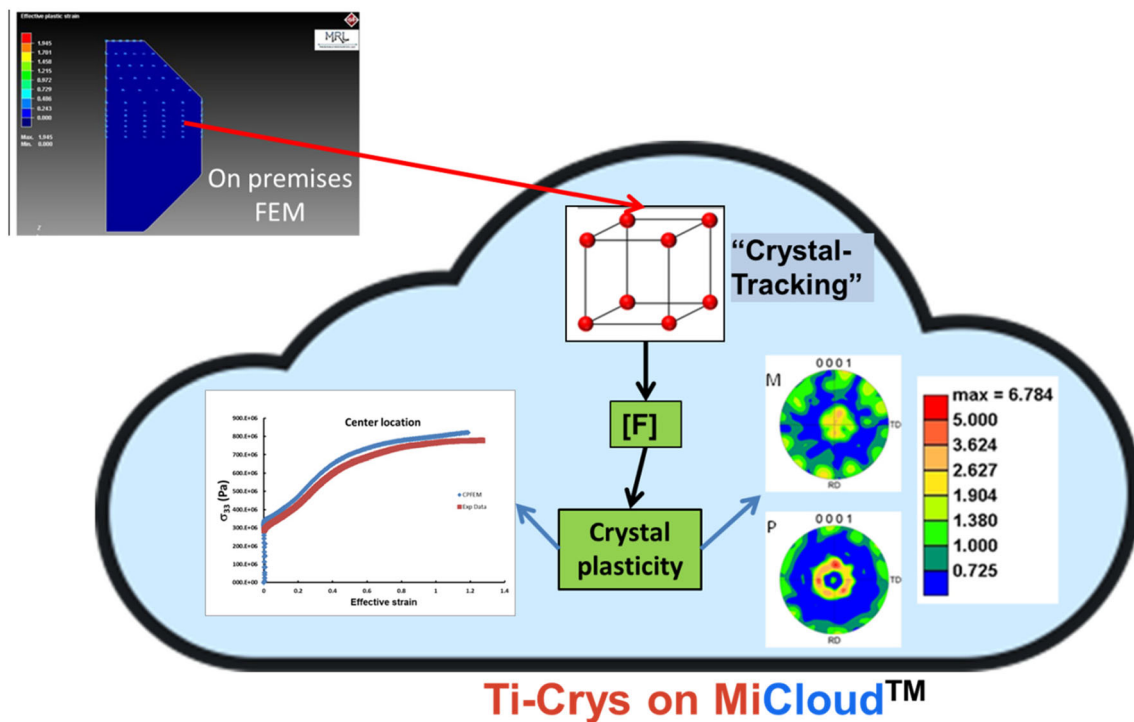


Fig. 2 Workflow for interoperability of MiCloud crystal plasticity with on-premise FEM simulations using the crystal tracking approach for predicting location-specific texture evolution and anisotropic mechanical behavior (Color figure online)

FEM nodes after potentially thousands of time steps, processing videos, and EBSD scans with hundreds of millions of data points. Not only can the volume of data become large, but it also comes with different noise, uncertainty, and biases in experimental data that need to be filtered and interpreted to preserve veracity and give confidence that the results are not distorted. With that in mind, one of the important decisions that had to be made was the approach for data storage, organization, and handling. One often is faced with multiple options, from flat files, structured file formats (e.g., hierarchical data format; HDF5), relational databases (i.e., SQL databases), and non-relational databases (i.e., NoSQL databases). With interoperability of materials datasets and scalability as a top priority, a balance was sought between the advantages and disadvantages of each data organization method with particular emphasis on (i) data containers for efficient portability of large microstructure datasets (e.g., HDF5), (ii) mainstream relational databases (i.e., SQL databases), and (iii) modern non-relational databases (i.e., NoSQL). After testing all three methods, it was determined that all three should be used: (i) data container models to avoid creating ad hoc methods of dealing with multiple streams of input and output within a computational model; (ii) relational database for synthesizing derived data using queries, views, and stored procedures; and (iii) NoSQL databases for creating hierarchical object documents such as data within the microstructure/local-state module. The reasoning for this is expanded in the following paragraphs.

To assure interoperability and reduce the probability of data loss in hierarchical data files, it was decided to store raw data as flat files (data files that contain records with no structured relationships). Both HDF5 files and databases are used for data products from data analysis and simulations. HDF5 is a multipurpose, open-source hierarchical container format capable of storing large numerical datasets with their metadata. Effectively, HDF5 can be seen as a file system within a file, where files are datasets and folders are groups [28]. HDF5 files are portable and extensible which makes them efficient for data sharing among various ICME projects [29]. There are pros and cons to moving forward with this file format. The advantages include open source within a large supportive community [30], portability with metadata included, chunked datasets that can be resized along a given dimension, and possible support for compression.

Some of the reported disadvantages of HDF5 include the following [28]: the file format must be designed, there is a high risk of data corruption, there are bugs and crashes in the HDF5 library and in the wrappers, it could be impossible to explore datasets with standard Unix/Windows tools, there is a hard dependency on a single implementation of the library, and there is a high complexity of the specification and the implementation. During the implementation of HDF5 in MiCloud, the design of HDF5 file format for various material files was time consuming. Furthermore, the implementation of HDF5 to contain thousands of EBSD files failed with the corruption of data in one of those files which often occurs

when collecting EBSD data over multiple days. Consequently, it was decided to use HDF5 only when dealing with multiple streams of input and output within a computational model and when datasets are shared via conventional download process.

For all other activities, flat files were used and managed by a database. This database required the use of a database management system (DBMS) that interacts with the user, other applications, and the database itself to capture and analyze data and allow the definition, creation, querying, update, and administration of databases. Two types of databases were used: SQL and NoSQL databases. Mainly, the NoSQL database was used for managing the unstructured local state datasets, and the SQL database was used elsewhere. Table 1 has an outline of the benefits of each type that directed the implementation in MiCloud. In summary, it was concluded that for interoperability, scalability, and efficient usage of MiCloud, there is no “*one size fits all*” for data storage. Thus, HDF5 files and databases (SQL and NoSQL) were utilized.

The preceding sections focused on describing the building blocks of MiCloud following the ICME workflow. The following case studies demonstrate the utility of MiCloud for interoperability of various datasets and applications, highlighting the platform’s ability to handle multiple types of data, carry out successful data characterization, and integrate between multiple software programs.

Case Description and Discussion

To demonstrate the interoperability benefits of microstructure-informed web-based applications, the following discussion focuses on analytics from microstructure data for Ti6Al4V and modeling data from Ni-based superalloys. Automatic identification and quantification of different phases in BSE images is followed by characterization of microtextured regions (MTRs) in EBSD data that are characterized by the spatial clustering of primary alpha particles with similar orientations. The selection of the case study was motivated by various efforts on the detrimental impact of MTR on the

fatigue life of alpha/beta titanium alloys [32–37]. Researchers seeking quality control via pattern discovery in multiple Ti6Al4V samples [38] can easily generate terabytes of raw data. Storage and exchange of such datasets becomes a challenge for materials scientists and engineers which is compounded as more projects need to be handled across a supply chain or a large organization. Thus, the current case study offers an efficient solution for the interoperability of various algorithms analyzing MTRs with an emphasis on scalability via cloud computing.

MTRs are often reported [32–37] in the bimodal microstructure (Fig. 3) of the alpha/beta titanium alloy (e.g., Ti6Al4V, Ti6242) that is desirable for applications limited by strength, ductility, and/or high cycle fatigue properties. In such a microstructure, the beta phase is shown in BSE images as white/light gray layers or regions surrounding the alpha phase that is revealed as dark gray/black regions due to the segregation of vanadium and aluminum into the two phases, respectively. The alpha phase has two morphologies in the bimodal microstructure: a colony-type morphology with alternating alpha laths laminated in a layer of thin beta in what is called secondary alpha, and islands of dark gray/black globular primary alpha particles [39].

There is a significant range of length scales covered by these features. It has been reported that beta layer thickness can be fractions of a micrometer, secondary alpha laths are measured in a few microns, primary alpha particles cover tens of microns [40], and microtextured regions (MTRs) can cover tens of millimeters [38]. Consequently, various characterization techniques are used to capture the multiscale characteristics of MTRs including (but not limited to) backscattered electron (BSE) imaging to capture the morphology of beta, primary and secondary alphas, and electron backscatter diffraction (EBSD) to capture the crystallography of various phases. Robust MTR identification and quantification may require large-area EBSD scans at a resolution (scan step size) that results in hundreds of millions of pixels. The nature of EBSD data collection by a combined beam/stage control method generates thousands of EBSD tiles (i.e., windows) to characterize an area of 10 mm × 10 mm at 1 μm step size

Table 1 Brief comparison of databases used in MiCloud [31]

	SQL	NoSQL
Data storage	Stored in a relational model, with rows and columns. Rows contain all of the information about one specific entry/entity, and columns are all the separate data points.	The term “NoSQL” encompasses a host of databases, each with different data storage models: document, graph, key value, and columnar.
Schemas and flexibility	Each record conforms to fixed schema, meaning that the columns must be decided and locked before data entry and each row must contain data for each column. Amending involves altering the whole database and going offline.	Schemas are dynamic. Information can be added on the fly, and each “row” (or equivalent) does not have to contain data for each “column.”
Scalability	Scaling is vertical. More data means a bigger server, which can get very expensive. It is possible to scale an RDBMS across multiple servers, but this is a difficult and time-consuming process.	Scaling is horizontal, meaning across servers. Many NoSQL technologies also distribute data across servers automatically.

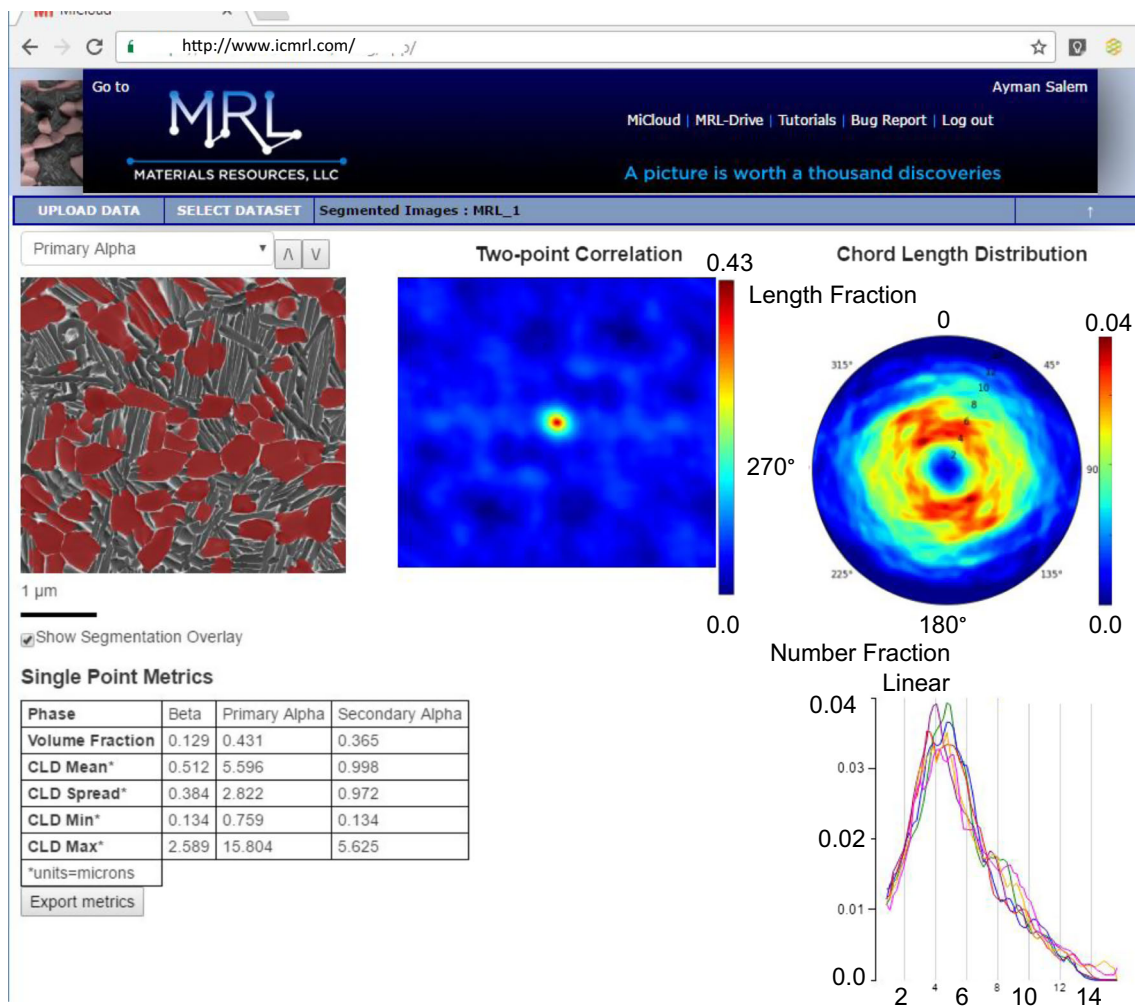


Fig. 3 Screenshot of identification and quantification of primary alpha, secondary alpha, and beta phases in alpha/beta titanium backscattered electron images and optical images using TiSeg™ app running on MiCloud platform (Color figure online)

which results in hundreds of GB of data per sample just for MTR characterization. Since accurate characterization of MTR features requires analysis of both BSE and EBSD datasets, the following sections discussed analytics of BSE first then EBSD datasets.

Feature Detection and Quantification Various Phases in BSE Images

The fact that primary alpha and secondary alpha features have the same Z-Contrast and crystal structure means that they have the same contrast in BSE images [41] and hence require advanced algorithms to identify the features based on chemistry and/or morphology. Since the morphology of both features are dependent on an often unknown processing pedigree, the algorithm used needs to be based on pattern recognition rather than pattern matching (or templates). These techniques are included in a web app (TiSeg™) on MiCloud with repeatable and reproducible results on thousands of BSE images. It

employs various algorithms for computing abstractions of image information based on computer vision and image processing techniques, enabling local decisions at every pixel to determine whether there is primary alpha, secondary alpha, or beta at that pixel. This results in three subsets of the image domain: segmented primary alpha, segmented secondary alpha, and segmented beta (left of Fig. 3).

Segmentation of various phases is only the first step. The crucial second step is the quantification of each phase (i.e., the feature of interest, FOI). To account for the spatial distribution of the FOI, two-point statistics and chord length distribution (CLD) [22] were used (Fig. 3). These techniques enable the user to quantify the heterogeneous spatial variability of the FOI as well as the anisotropic morphology in a manner that is not possible by legacy methods that use single-point metrics (e.g., ASTM E112).

In brief, the two-point statistics captures the probability of finding a vector with the head and the tail from the same FOI for all possible lengths and directions of the vector [42]. The

CLD captures the probability of finding a line segment that falls inside the FOI for all possible lengths and directions of the line segment. Together, they encompass the fingerprint of the whole subdomain image for a specific FOI [22]. As such, further data analytics of the two-point statistics and CLD enables fast, repeatable, and reproducible microstructure-based quality control as will be shown later for MTRs.

Linking to computational models that are based on single-point metrics and the associated distribution is simply done by extracting particular single-point metrics from the higher-order statistics. In particular, the volume fraction of any phase is calculated as two-point statistics for a vector of zero length which is the highest probability for any FOI (e.g., 0.41 for primary alpha, Fig. 3). The average, median, and distribution of an FOI size in any particular direction is calculated from extracting radial data from the CLD plot in Fig. 3. This is done automatically for preselected directions at 10° increments as shown in the bottom right of Fig. 3 or at any direction that is selected by the user via the CLD app. Single-point metrics of FOI size for the whole image are also tabulated in the bottom left corner of Fig. 3. However, any information about morphological anisotropy is lost in this single point metrics data.

Despite the complicated mathematics and computer programming behind the computer vision algorithms and the higher-order statistics, cloud computing made it accessible for materials science practitioners regardless of the computer programming skills [19, 20]. Furthermore, the calculated two-point statistic and CLD are available for various power users to be utilized in various data analytics needs. For example, two Ti6Al4V images showed distinctive microstructure heterogeneity that was not captured by single-point metrics while being revealed by two-point statistic and CLD plots in Fig. 4. Batch processing of tens of thousands of images is also automated for microstructure-based quality control which can be essential for optimizing thermomechanical processing of a material or reducing batch to batch, lot to lot, or supplier to supplier variability in an organization or across a supply chain. It is worth noting that these computational intensive apps handling tens of thousands of images can be accessed via thick or thin web client interface.

In another example, further interoperability of the developed apps is demonstrated by quantification of solidified microstructure predicted by MICRESS®, a MICRostructure Evolution Simulation Software [43, 44] that enables the time- and space-resolved calculation of microstructure formation. The workflow (Figs. 5 and 6) to insert MICRESS results into MiCloud was developed, and data were uploaded in 24 h during the 2016 ICMEg conference. The results of data analytics were obtained and presented in the conference. Specifically, the phase field simulation results by MICRESS (Figs. 6 and 7) for IN738 and IN 738+ 2% Nb were analyzed by TiSeg™ to extract two-point statistics, CLD, and single-point metrics of the dendrites. The higher-order statistics

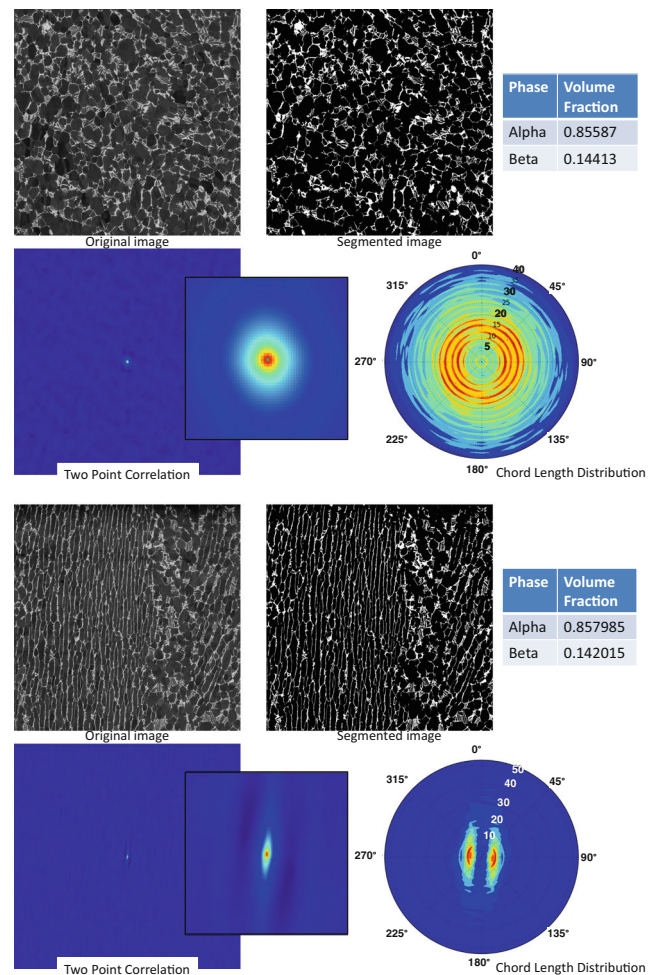
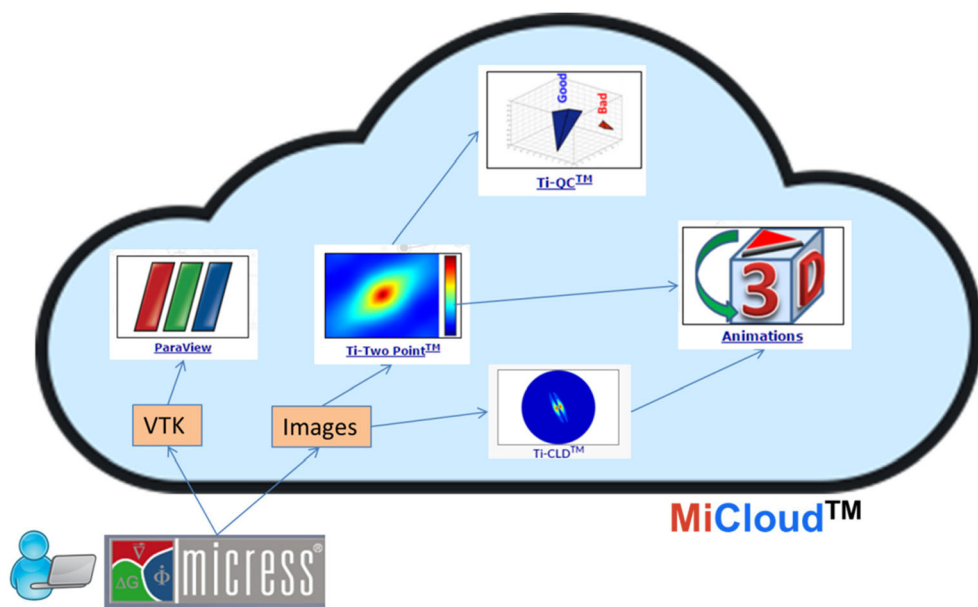


Fig. 4 Quantification of BSE images of the Ti6Al4V sample by two-point correlations and chord length distributions revealed the drastic differences in morphology despite the similarity in the single-point metrics at different spatial locations of the same sample (Color figure online)

revealed a distinct difference due to the addition of Nb as shown in the CLD and two-point statistics (Figs. 8 and 9). It was found that the average size of the dendrites, based on the CLD from TiSeg, was much smaller than the reported size (Fig. 7) values based on the equivalent circle diameter due to the inclusion of the anisotropic morphology of dendrites in the CLD results. Higher magnification of the two-point statistics showed a difference in the alignment in the dendrite arms between both materials as highlighted by the white lines in Figs. 8 and 9. The randomness of the alignment of the dendrite arms can be confirmed by comparing the two-point statistics from multiple simulations of the same material. The difference in the volume fractions, mean, minimum, and maximum CLD metrics was revealed in the tables associated with the results for each material. Hundreds of simulations can be compared easily using this technique to evaluate the effect of alloying elements, modeling parameters, or other objectives that can be set by the user. These can be done in an uncoupled

Fig. 5 MiCloud workflow for quantification of MICRESS [44] results on cloud computing as an example of global interoperability of the platform (Color figure online)



manner where MICRESS runs on premises and then only the results are uploaded to MiCloud for data analytics (Fig. 5). While the above results are not related to Ti6Al4V microstructures, it is included to demonstrate the interoperability across various materials enabling the reuse of the higher-order statistics algorithms.

Feature Recognition of Microtextured Regions in EBSD Datasets

A microtextured region (MTR) is characterized by spatial clustering of a feature of interest (FOI) (e.g., primary alpha) with similar crystallographic orientation. Since the location or size is

unknown prior to conducting EBSD scans, large-area EBSD scans [38] are often conducted covering areas as big as 100 mm². This is accomplished via a special control of the scanning electron microscope that combines beam and stage controls. Such a scan results in thousands of EBSD tiles. Post-processing of these tiles requires the use of software that can handle millions of orientations. Those thousands of EBSD files (in ASCII format) can be batch uploaded to MiCloud with an associated log file that identifies the spatial location of each tile. To assure interoperability of the analytics software with various data generation platforms, both *.ang (from EDAX [45]) and *.ctf files (from Bruker [46] and HKL [47]) can be analyzed. Since the flat files are used without combining them into a HDF5 file, three

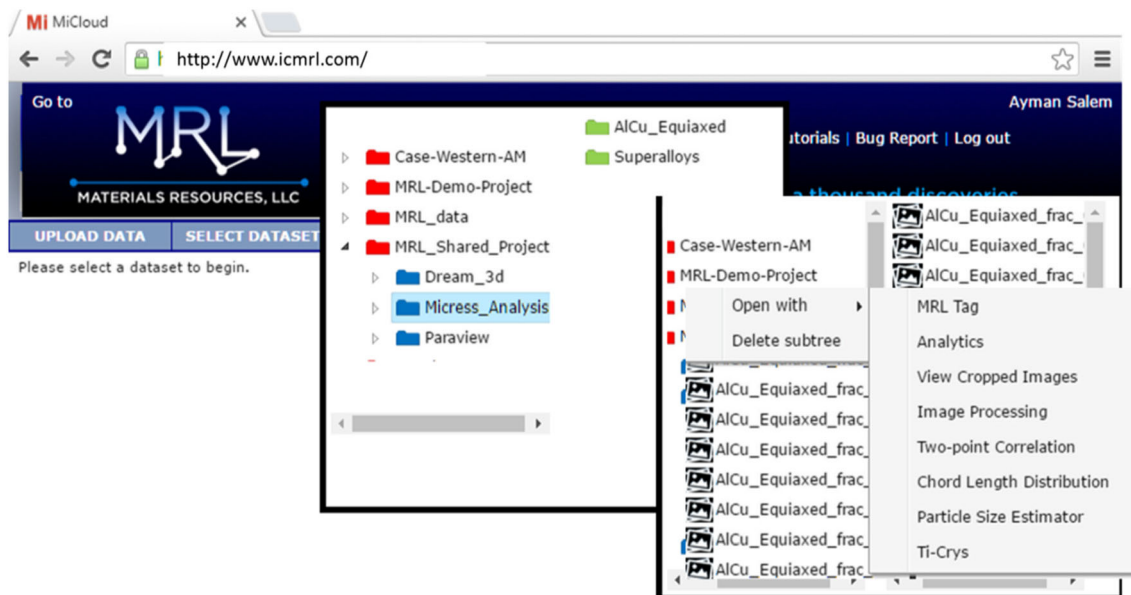
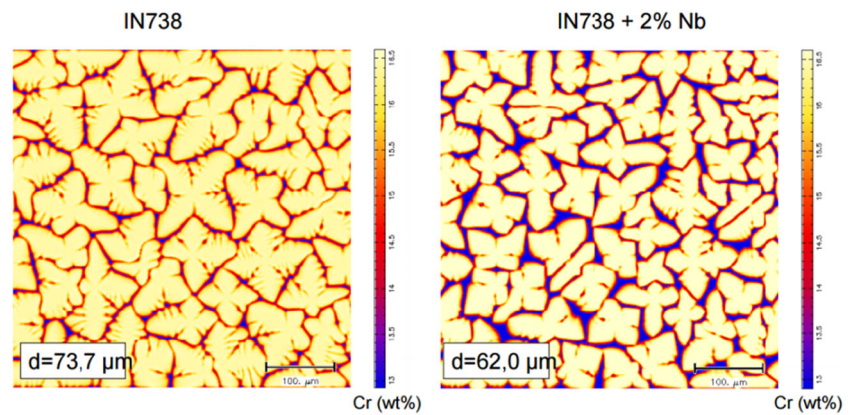


Fig. 6 Screenshot showing the addition of MICRESS [44] results into MiCloud database for analysis (Color figure online)

Fig. 7 MICRESS predictions of IN738 and IN738 + 2% Nb with a comparison of the average grain size [43] (Color figure online)



different apps on MiCloud can be used for analysis, in particular TiZone™, Dream3D, and Woodfield. It is worth noting that Dream3D software is developed and distributed by Bluequartz [17] for on-premise use. However, the web-based implementation in MiCloud and the associated GUI was developed by the authors. The algorithm behind the color coding of orientation information in the Woodfield app was developed by Yamrom [48] and implemented for MTRs by Woodfield [35]. However, the software for implementation on MiCloud was developed by the authors. This demonstrates the ability of inserting algorithms and/or software developed by others in MiCloud platform. A

brief description of each algorithm and the associated software is discussed below:

The TiZone algorithm and the associated software was developed by the authors [49] based on two sequential steps: (1) unsupervised machine learning (e.g., k-means clustering) and pattern recognition in the generalized spherical harmonics (GSH) [50] domain to identify MTR families according to orientation clustering and (2) mapping MTR families to the spatial domain to quantify MTRs based on their spatial clustering [49]. This approach eliminates the need to stitch thousands of EBSD tiles in one large file which offers significant

Fig. 8 MiCloud analytics of MICRESS predictions for IN738 in Fig. 7 (Color figure online)

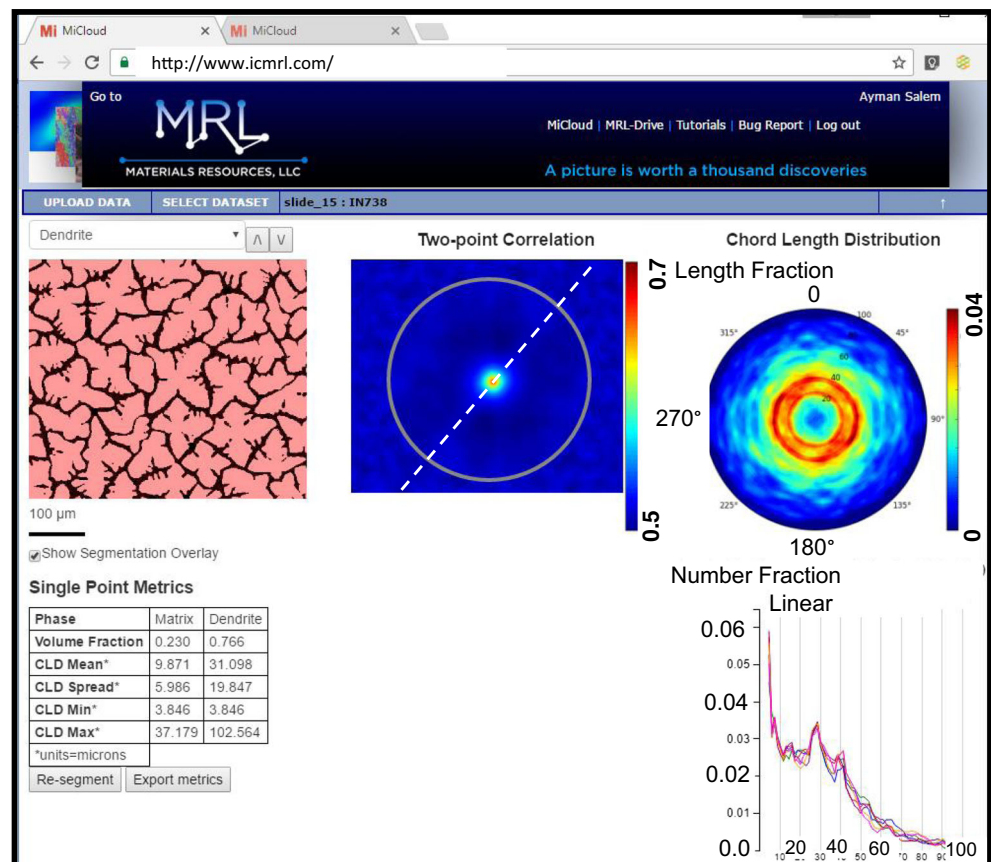
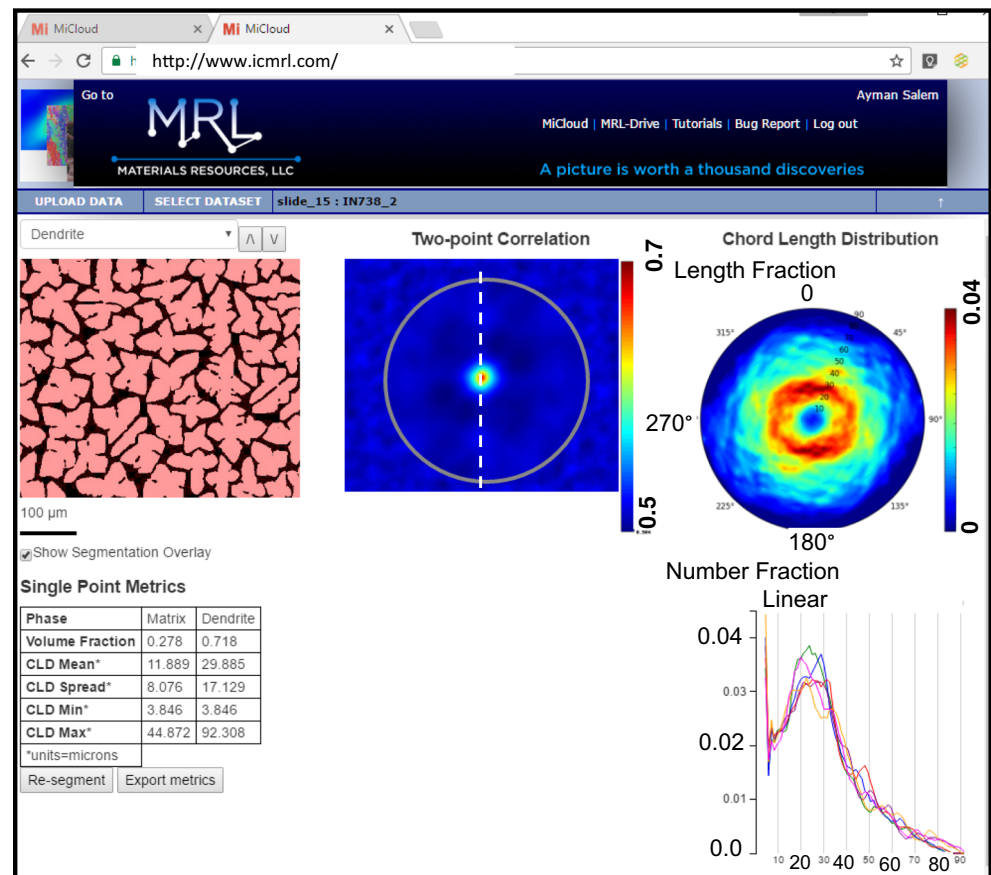


Fig. 9 MiCloud analytics of MICRESS predictions for IN738 + 2% Nb in Fig. 7 (Color figure online)



benefits for handling datasets with corrupt tiles and/or enables MTR recognition without an upper limit on the total number of orientations since tiles are processed individually. Also, TiZone is an example of an algorithm that is built for horizontal scaling by allowing the processing of input datasets in a completely parallel manner with the use of as many computational nodes as needed and then sorting the output and assembling it for representation of the whole dataset.

The Dream3D algorithm and the associated software use an efficient *.hdf5 files to stitch all tiles, and then MTRs are identified based on misorientation of neighboring pixels according to misorientation thresholds identified by the user. The advantage of this approach is that it uses a methodology familiar to the user, as it is similar to the algorithm used in identifying grains in popular commercial EBSD analysis software.

The algorithm of the third approach is based on a coloring technique that was developed by the General Electric (GE) team in 1997 [48], for visualization of a two-dimensional orientation data on a surface by constructing an icosahedron at a point on the surface corresponding to the orientation vector. Orientation vectors are projected onto facets of the icosahedron, and the surface point which is associated with the orientation vector is color-coded with the color of the intersected facet. Colors are assigned to facets of the icosahedron such

that adjacent facets have similar colors. A software for implementing the algorithm was developed in-house and named Woodfield, acknowledging his implementation of the methodology for visual identification of MTRs [35].

The interoperability of the developed apps is shown in the workflow in Fig. 10 including an efficient integration with crystal plasticity and finite element analysis via the generation of representative orientation distribution (ROD) (via Ti-ROD™ app) which saves the crystal plasticity calculation time by more than 99.5% [22].

In particular, upon uploading the individual tiles to TiZone app on the cloud, the three inverse pole figure (IPF) maps for the raw data are plotted (Fig. 11) and the data is colored according to Yamrom method [48] to reveal MTRs visually (Fig. 12).

To demonstrate the pattern discovery in a generalized spherical harmonics (GSH) space for automatic identification of MTR using the C-coefficients, a colored C-Map [49] from TiZone is first used to visualize various MTR families. The C-Map is a spatial representation of the projection of the c-axes of individual crystals to the sample coordinate systems. To confirm interoperability of various developed algorithms processing the same dataset, a simple comparison of the C-Map and Woodfield map for the [0001] axis of the alpha phase in the EBSD is conducted. The maps showed good agreement

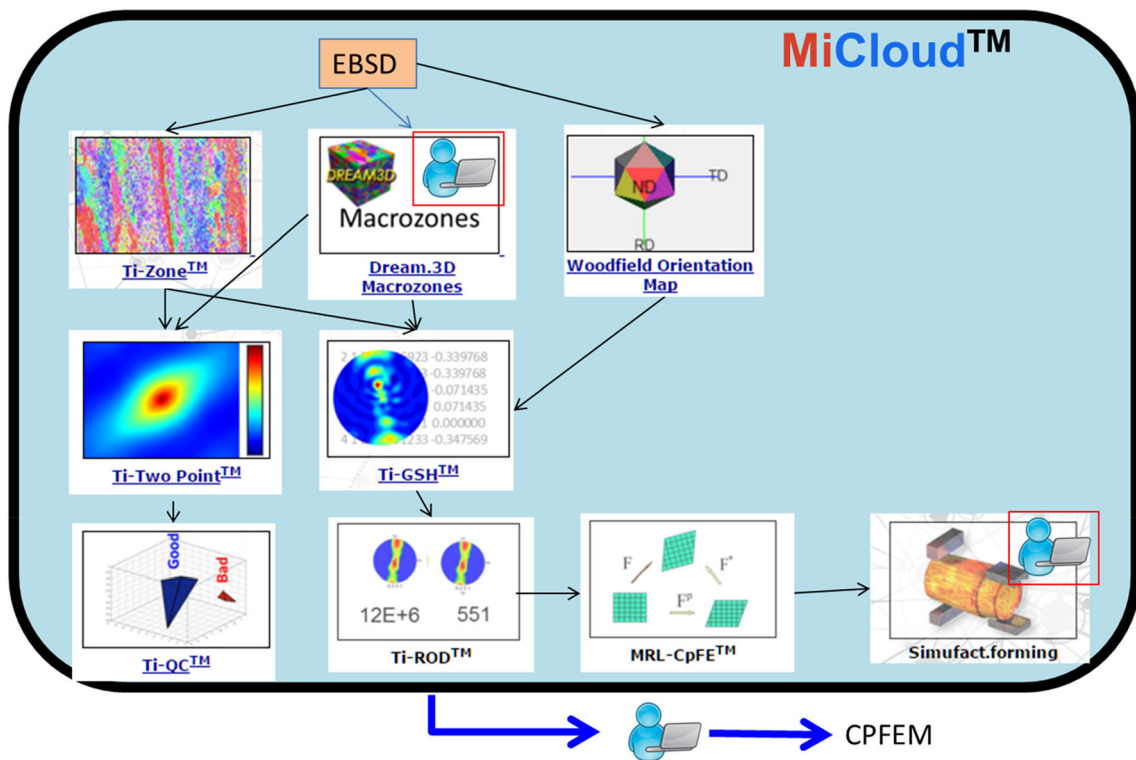


Fig. 10 Workflow for identification and quantification of microtextured regions (MTR) or Macrozones including integration with crystal plasticity and FEM simulations (Color figure online)

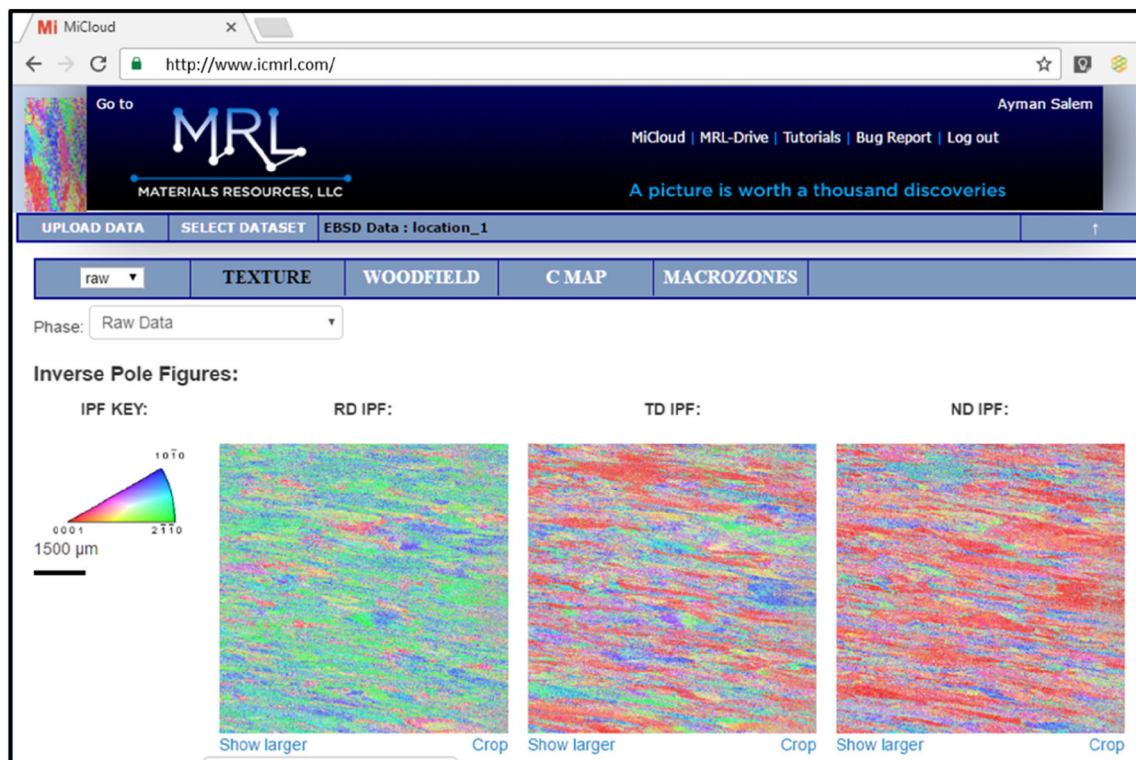


Fig. 11 Inverse pole figure maps for large-area EBSD scans as plotted by TiZone on MiCloud (Color figure online)



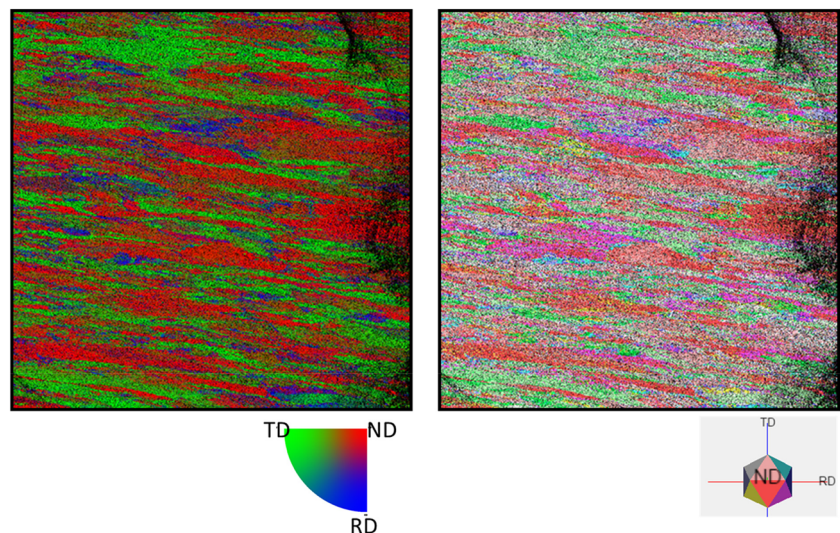
Fig. 12 EBSD data colored according to Yamrom's method [48] to reveal MTR using Woodfield app on MiCloud (Color figure online)

between both methods of coloring the EBSD orientations, revealing the MTRs similarly, but the difference that the C-Map uses continuous coloring scheme while the icosahedron method uses coloring based on preassigned bins (Fig. 13). Both coloring methods showed MTRs regardless of the presence of points with lower confidence index (black pixels in Fig. 13).

A crucial step for MTR identification by TiZone is the segmentation which results in three main families that are color-coded according to the centroid orientation of each

family with the red MTRs having the highest area fraction (close to the ND) and the blue MTR having the lowest area fraction (close to the RD) (Fig. 14). This is achieved by conducting cluster analysis in the GSH domain first followed by mapping clustered data to the real space with each pixel associated with the identified MTR family. This approach enables the use of unsupervised machine learning [49] to identify orientation clusters without prior knowledge of tolerance on the misorientation between neighboring pixels in real space. One of the major benefits of this approach is the ability

Fig. 13 Coloring EBSD dataset by C-Map [49] and icosahedron methods [48] in TiZone (left) and Woodfield apps (right) on MiCloud. Black points are data with low confidence index (CI) from sample preparation (Color figure online)



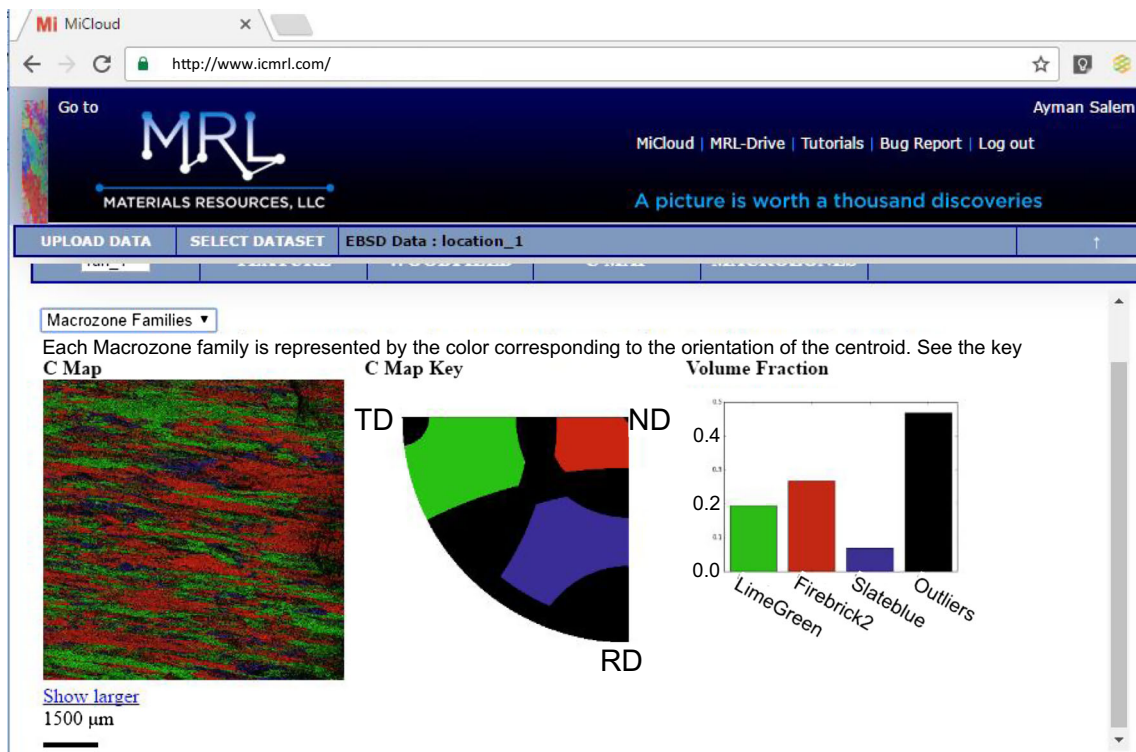


Fig. 14 Identification of MTRs in Ti6Al4V using machine learning in TiZone™ [49] (Color figure online)

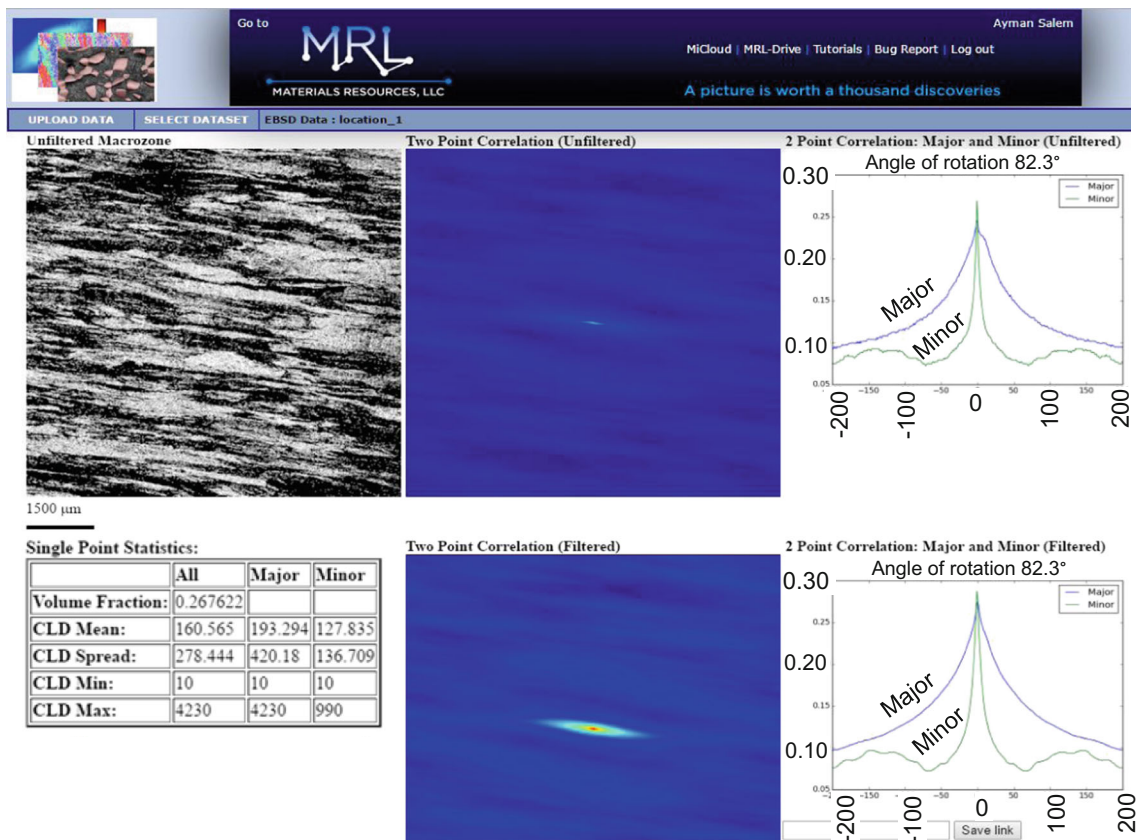


Fig. 15 Quantification of MTRs using two-point correlation and chord length distribution apps of the segmented MTRs in Fig. 14 (Color figure online)

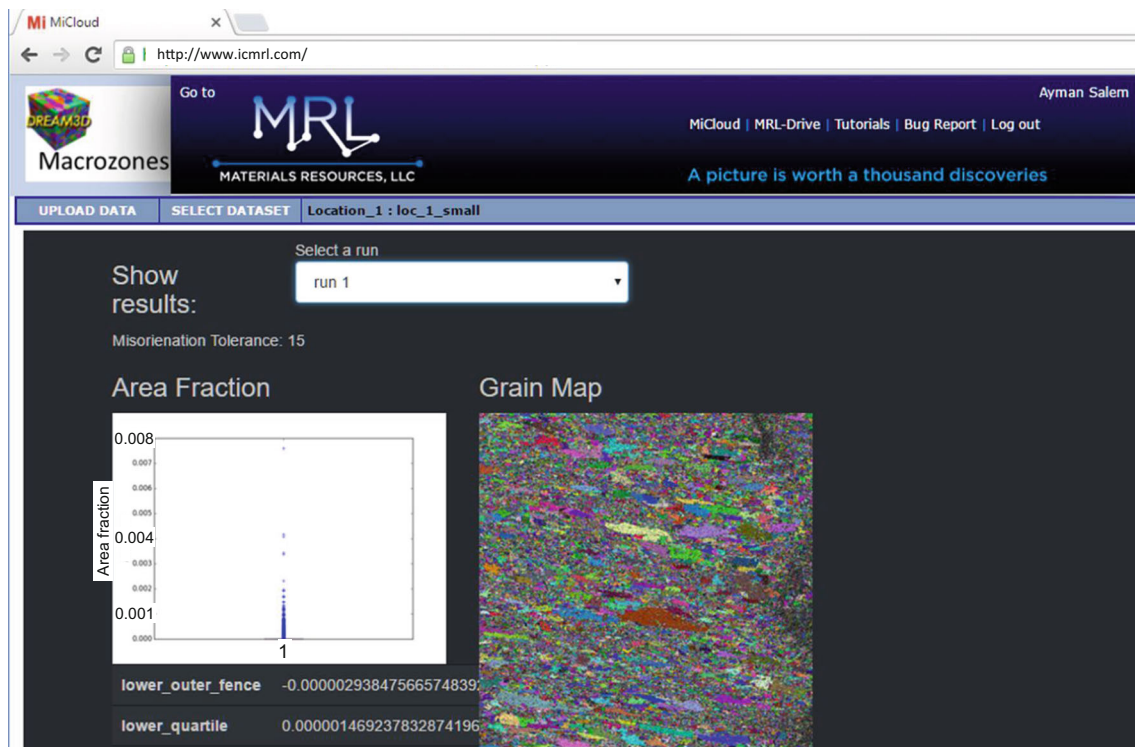


Fig. 16 Identification of MTRs, in Fig. 11, using the Dream3D pipeline deployed on MiCloud. Each *color* defines a single MTR with misorientation tolerance of 15° (Color figure online)

to identify MTRs in 3D EBSD datasets (e.g., serial sectioning or oblique sections) without the constraints of the proximity of the dataset as long as each pixel is associated with an x, y, z spatial location in the sample coordinate system. The ability to analyze non-rectangular scans is another benefit.

The quantification of the segmented MTR is then achieved by the use of higher-order statistics (i.e., two-point correlation) by integration with the two-point correlation app and the chord length distribution app (Fig. 15). These two apps are the same ones used to quantify TiSeg segmented features.

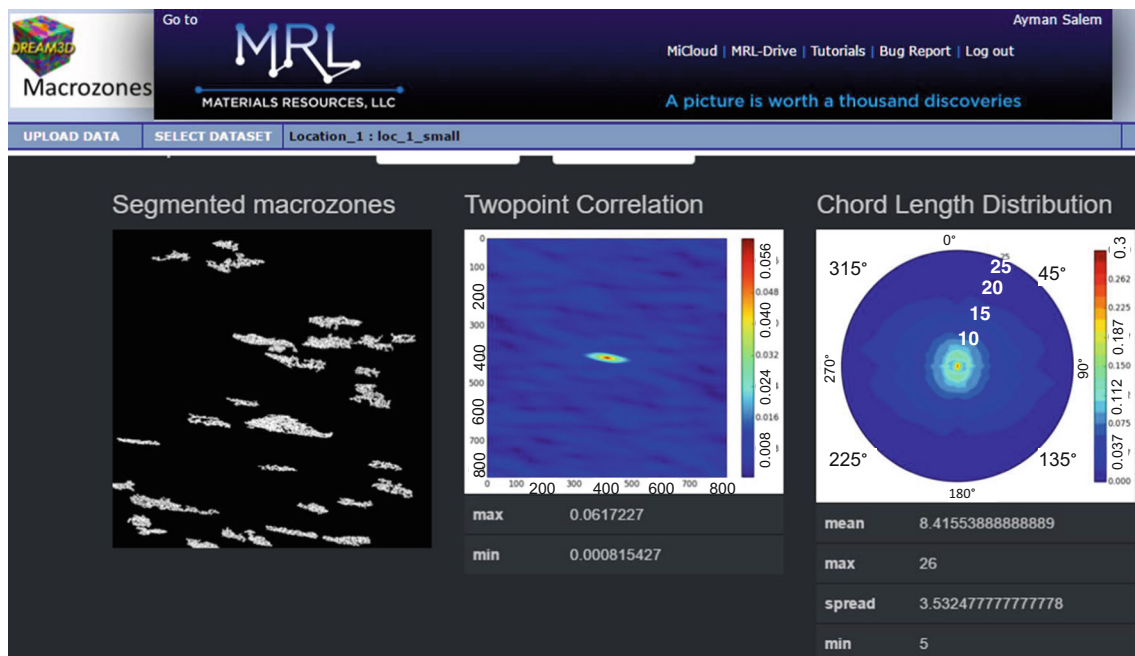


Fig. 17 Quantifications of segmented MTRs by Dream3D using two-point correlation and CLD on MiCloud (Color figure online)

This confirms the interoperability of both apps since they are built to quantify the spatial distribution of an FOI regardless of what is that feature (e.g., primary alpha phase, dendrites, or MTR family). Since the FOI is identified with algorithms acting on input data in parallel, the quantification is also done in parallel, enabling horizontal scaling with the size of the input dataset. An example of conducting MTR identification and quantification on many samples was demonstrated by Pilchak et al. [38] using microstructure-based quality control to identify the effect of billet size on MTR for Ti6242Si.

The interoperability of microstructure datasets stored in flat files is further proven by the authors' web-based implementation of the open-source application Dream3D [16] that was developed around the Macrozone pipeline [17] to identify MTR. It was then integrated with MiCloud for quantification of the segmented MTR by two-point statistics and CLD apps that were previously used with TiZone and TiSeg. Using the same EBSD dataset demonstrated above by the TiZone and Woodfield apps, a grain map was generated in which MTR was identified by randomly selected colors with area fraction of each grain plotted in a boxplot (Fig. 16) to be used for further segmentation of MTRs based on area fraction threshold selected by the user which enabled quantification of segmented MTRs by the two-point correlations and CLD apps (Fig. 17).

Using the same EBSD dataset across various apps showed some of the many benefits of adopting a cloud computing platform in ICME implementations, allowing for data and model reuse by users across the globe.

Conclusions

Interoperability of materials datasets, data analytics tools, and computational software are crucial elements for practical implementation of ICME and MGI visions. Microstructure-informed cloud computing (MiCloud) was developed to offer a practical platform for interoperability of materials datasets and data analytics tools. Global sharing of microstructure data analytics and collaboration among academic researchers and across industrial supply chains was demonstrated on a case study using BSE and EBSD datasets for Ti6Al4V that were analyzed for identification and quantification of various features of interest (FOIs) including primary alpha particles, beta phase, secondary alpha laths, and microtextured regions (MTRs). The case study showed that for interoperability, scalability, and efficient reuse of big material datasets and data analytics, there was no “one size fits all” for data storage. Thus, flat files, HDF5 files, and databases (SQL and NoSQL) were all utilized in MiCloud.

Acknowledgements Discussions with Dr. A. Pilchak, Dr. S.L. Semiatin, Dr. J. Calcaterra, and Dr. C. Ward of the Air Force Research

Laboratory; Dr. T. Broderick and Dr. A. Woodfield of GE Aviation; Dr. M.G. Glavicic of Rolls Royce; Dr. V. Venkatesh of Pratt & Whitney; and Dr. S. Tamirisakandala of Arconic are gratefully acknowledged. Use and supply of software and datasets for MiCloud by Prof. S. Niezgodá (OSU), Prof. R. Srinivasan (WSU), Prof. J. Lewandowski (CWRU), and Dr. M. Seifi (ASTM International) are highly appreciated.

Authors' Contributions AAS conceived MiCloud concept and its interoperability and created the initial draft. JBS, RAK, LAW, and DPS helped with manuscript writing and analysis of exemplary datasets. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Competing Interests The authors are employees of Materials Resources. The TiZone algorithm is covered under US patent US9070203 B2.

References

1. Committee on Integrated Computational Materials Engineering National Materials Advisory Board Division on Engineering and Physical Sciences National Research Council (2008) Integrated Computational Materials Engineering. doi: 10.17226/12199
2. National Science and Technology Council (2011) Materials Genome Initiative for global competitiveness. Executive Office of the President
3. Ward CH, Warren JA, Hanisch RJ (2014) Making materials science and engineering data more valuable research products. *Integr Mater Manuf Innov* 3:1–17. doi:10.1186/s40192-014-0022-8
4. Jacobsen MD, Fourman JR, Porter KM et al (2016) Creating an integrated collaborative environment for materials research. *Integr Mater Manuf Innov* 5:12. doi:10.1186/s40192-016-0055-2
5. University of Michigan (2015) PRISM project. <https://wiki.umms.med.umich.edu/display/UMHSHELPDESK/Prism>
6. University of Illinois T2C2: Timely and trusted curation and coordination. In: 2015. <http://t2c2.csl.illinois.edu/>
7. van Dam KK, Carson J, Corrigan A et al. (2012) Velo and REXAN 2014; Integrated data management and high speed analysis for experimental facilities. In: 2012 I.E. 8th Int. Conf. E-Science. IEEE, pp 1–9
8. Carey NS, Budavári T, Daphalapurkar N, Ramesh KT (2016) Data integration for materials research. *Integr Mater Manuf Innov* 5:7. doi:10.1186/s40192-016-0049-0
9. McLennan M, Kennell R (2010) HUBzero: a platform for dissemination and collaboration in computational science and engineering. *Comput Sci Eng* 12:48–53. doi:10.1109/MCSE.2010.41
10. Wallack AS (1995) Algorithms and techniques for manufacturing. (Ph.D. thesis, University of California at Berkeley)
11. Bonte MHA, van den Boogaard AH, Huétink J (2007) A metamodel based optimisation algorithm for metal forming processes. In: *Adv Methods Mater. Form.* Springer Berlin Heidelberg, Berlin, Heidelberg, pp 55–72
12. Raymond E (1999) The cathedral and the bazaar. *Knowledge, Technol Policy* 12:23–49. doi:10.1007/s12130-999-1026-0
13. GitHub I GitHub. <https://github.com/>. Accessed 1 Jan 2016
14. Hielscher R, Bachmann F (2017) MTEX-A texture calculation toolbox. <http://mtex-toolbox.github.io/>. Accessed 1 Jan
15. Bachmann F, Hielscher R, Schaeben H (2011) Grain detection from 2d and 3d EBSD data—specification of the MTEX algorithm. *Ultramicroscopy* 111:1720–1733. doi:10.1016/j.ultramic.2011.08.002

16. BlueQuartz Software LLC DREAM3D. <https://github.com/BlueQuartzSoftware/DREAM3D>. Accessed 1 Jan 2017
17. Groeber M, Jackson M (2014) DREAM.3D: a digital representation environment for the analysis of microstructure in 3D. *Integr Mater Manuf Innov* 3:5. doi:10.1186/2193-9772-3-5
18. Mell PM, Grance T (2011) The NIST definition of cloud computing. doi: 10.6028/NIST.SP.800-145
19. Materials Resources LLC (2015) MiCloud. <http://www.icmrl.com>. Accessed 1 Jan 2017
20. Materials Resources LLC (2016) MiCloud.AM for additive manufacturing. www.MiCloud.AM. Accessed 1 Jan 2017
21. Seifi M, Salem A, Beuth J et al. (2016) Overview of materials qualification needs for metal additive manufacturing. *JOM* 68: 747–764. doi: 10.1007/s11837-015-1810-0
22. Salem AA, Shaffer JB, Satko DP et al (2014) Workflow for integrating mesoscale heterogeneities in materials structure with process simulation of titanium alloys. *Integr Mater Manuf Innov* 3:24. doi:10.1186/s40192-014-0024-6
23. Salem AA, Kalidindi SR, Doherty RD, Semiati SL (2006) Strain hardening due to deformation twinning in α -titanium: mechanisms. *Acta Mater* 37:259–268
24. Zhao P, Song En Low T, Wang Y, Niezgod SR (2016) An integrated full-field model of concurrent plastic deformation and microstructure evolution: application to 3D simulation of dynamic recrystallization in polycrystalline copper. *Int J Plast* 80:38–55. doi:10.1016/j.ijplas.2015.12.010
25. Lebensohn RA, Kanjarla AK, Eisenlohr P (2012) An elastoviscoplastic formulation based on fast Fourier transforms for the prediction of micromechanical fields in polycrystalline materials. *Int J Plast* 32–33:59–69. doi:10.1016/j.ijplas.2011.12.005
26. Satko DP, Shaffer JB, Tiley JS et al (2016) Effect of microstructure on oxygen rich layer evolution and its impact on fatigue life during high-temperature application of α/β titanium. *Acta Mater* 107:377–389. doi:10.1016/j.actamat.2016.01.058
27. MSC software simufact—simulating manufacturing. <http://www.simufact.de/en/index.html>. Accessed 4 Jun 2015
28. Rossant C (2016) Moving away from HDF5. <http://cyrille.rossant.net/moving-away-hdf5/>
29. Schmitz GJ, Böttger B, Apel M et al (2016) Towards a metadata scheme for the description of materials—the description of microstructures. *Sci Technol Adv Mater*. doi:10.1080/14686996.2016.1194166
30. The HDF group (2016) HDF5 HOME PAGE. <https://support.hdfgroup.org/HDF5/>
31. MCNULTY E (2014) SQL VS. NOSQL- What you need to know. <http://dataconomy.com/sql-vs-nosql-need-know/>
32. Germain L, Gey N, Humbert M et al (2005) Analysis of sharp microtexture heterogeneities in a bimodal IMI 834 billet. *Acta Mater* 53:3535–3543. doi:10.1016/j.actamat.2005.03.043
33. Pilchak AL (2014) A simple model to account for the role of microtexture on fatigue and dwell fatigue lifetimes of titanium alloys. *Scr Mater* 74:68–71. doi:10.1016/j.scriptamat.2013.10.024
34. Pilchak AL, Bhattacharjee A, Williams REA, Williams JC (2009) The effect of microstructure on fatigue crack initiation in Ti-6Al-4V. *ICF12*
35. Woodfield AP, Gorman MD, Sutliff JA, Corderman RR (1995) Effect of microstructure on dwell fatigue behavior of Ti-6242. In: *Titanium'95 Sci. Technol.* Birmingham, UK, pp 1116–1123
36. Venkatesh V, Tamirisa S, Sartkulvanich J et al. (2016) Icme of microtexture evolution in dual phase titanium alloys. In: *Proc. 13th World Conf. Titan.* Wiley, Inc., Hoboken, NJ, USA, pp 1907–1912
37. Qiu J, Ma Y, Lei J et al (2014) A comparative study on dwell fatigue of Ti-6Al-2Sn-4Zr-xMo (x = 2 to 6) alloys on a microstructure-normalized basis. *Metall Mater Trans A* 45:6075–6087. doi:10.1007/s11661-014-2541-5
38. Pilchak AL, Szczepanski CJ, Shaffer JA et al (2013) Characterization of microstructure, texture, and microtexture in near-alpha titanium mill products. *Metall Mater Trans A* 44:4881–4890
39. Semiati SL, Seetharaman V, Weiss I (1996) Hot working of titanium alloys—an overview. *Adv Sci Technol Titan Alloy Process* 3–73
40. Semiati SL, Knisley SL, Fagin PN et al (2003) Microstructure evolution during alpha-beta heat treatment of Ti-6Al-4V. *Metall Mater Trans A* 34:2377–2386. doi:10.1007/s11661-003-0300-0
41. Salem AA, Glavicic MG, Semiati SL (2008) A coupled EBSD/EDS method to determine the primary- and secondary-alpha textures in titanium alloys with duplex microstructures. *Mater Sci Eng A* 494:350–359. doi:10.1016/j.msea.2008.06.022
42. Kalidindi SR, Niezgod SR, Salem AA (2011) Microstructure informatics using higher-order statistics and efficient data-mining protocols. *JOM* 63:34–41
43. ACCESS (2009) Microstructure simulation using the software MICRESS
44. ACCESS MICRESS. <http://web.micress.de/>
45. EDAX Orientation Imaging Microscopy (OIM) Analysis. <http://www.edax.com/Products/EBSD/OIM-Data-Analysis-Microstructure-Analysis.aspx>. Accessed 1 Jan 2017
46. Bruker Quantax EBSD. <https://www.bruker.com/products/x-ray-diffraction-and-elemental-analysis/eds-wds-ebds-sem-micro-xrf-and-sem-micro-ct/quantax-ebds/overview.html>. Accessed 1 Jan 2017
47. Oxford AZtecHKL. <https://www.oxford-instruments.com/products/microanalysis/ebds/aztechkl-ebds-software>. Accessed 1 Jan 2017
48. Yamrom B (1997) Method of color coding orientation information. 5
49. Salem AA, Shaffer JB (2013) Identification and quantification of microtextured regions in materials with ordered crystal structure
50. Bunge HJ (1982) *Texture analysis in materials science: mathematical methods.* Butterworths, London