

Accessing Materials Data: Challenges and Directions in the Digital Era

John R. Rumble Jr¹

Received: 16 February 2017 / Accepted: 7 May 2017 / Published online: 5 June 2017
© The Author(s) 2017. This article is an open access publication

Abstract Providing better availability to materials data has recently gained new momentum. Many successes abound—large numbers of individual materials databases exist, powerful modeling and data analysis approaches have been developed, and Web-based technologies are available. At the same time, challenges remain: one-stop access is lacking, use of multiple databases at the same time is virtually impossible, using shared data is difficult, and understanding data quality is very hard. In this paper, we review the successes and challenges of accessing digital materials data, especially as new initiatives are starting. We also identify insights from previous work that provide guidance to future progress, including adherence to the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles, in achieving this dream.

Keywords Materials data · Online systems · Informatics · Databases · Data accessibility

Introduction

Our physical world is made of materials that, with few exceptions, have been processed from naturally occurring substances into products and structures that enable life as we know it today. The volume of materials produced each year is very large, both in terms of quantity and diversity. For example, in 2014, industrial production of primary metals (iron, steel, aluminum, and other non-ferrous metals)

contributed over \$281 B to the U.S. GDP [1]. Materials such as plastics, polymers, ceramics, and composites added similar substantial amounts. Worldwide, the numbers are staggering. As these materials are converted into products, it is clear how important materials are to our society and economy.

The measurement and availability of materials property data are crucial to successful design, manufacture, utilization, and disposal of products and structures. Today these data are generated, collected, evaluated, managed, analyzed, exploited, and disseminated using typical modern informatics tools. While materials informatics has resulted in large collections of high-quality materials property data, these collections are dispersed, often incomplete, difficult to access concurrently and integrate together, and of limited availability. Work during the last four decades has created and advanced materials informatics and increased the accessibility of computerized materials data. New initiatives, such as the Materials Genome Initiative [2], “Big Data” [3–5], and semantic Web technology [6], have opened the door to faster progress. In this paper, we review the many facets of materials data and how they impact present and future computerized access.

To begin, a twenty-first century vision for access to materials data was articulated several years ago [7, 8]:

The ability to locate and use all property data on all engineering materials easily, regardless of where those data are stored and maintained, through one or a small number of data portals (Web interfaces), noting that different data sets may have different data use restrictions including fees and proprietary control.

In this paper, we address the growing needs for access to materials data from the perspective of supporting the design and optimization of advanced materials, noting issues specific

✉ John R. Rumble, Jr
john.rumble@randrdata.com

¹ R&R Data Services, 11 Montgomery Avenue,
Gaithersburg, MD 20877, USA

to materials data that affect access as defined above. The remainder of the paper is structured as follows.

- “[Why Digital Access to Materials Data is Becoming More Important](#)” discusses motivation for the growing need for better access to digital materials data.
- “[Brief Review of Materials Data and Databases](#)” provides a brief review of materials databases from several important perspectives.
- “[Comprehensive Online Materials Data Systems](#)” reviews comprehensive online materials data systems, including their characteristics and challenges.
- New activities related to accessing materials data and why they have emerged.
- “[Contemporary Efforts](#)” deals with the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles in the context of materials databases.
- “[Thoughts on the Future of Materials Data Access](#)” presents concluding remarks.

We hope that this comprehensive review of access to materials data can provide guidance for future progress in improving their accessibility and use.

Why Digital Access to Materials Data Is Becoming More Important

We can identify five major reasons, as shown in Fig. 1, why digital access to materials has become more important in recent years. Each of the reasons is discussed in the subsequent paragraphs below.

Automation of Product Design and Engineering

Computer-assisted engineering (CAE) is now essentially complete, to the extent that each individual engineering activity, from planning to design to manufacturing to distribution, has been computerized and is now executed using software, middleware, and hardware of increasing sophistication [9, 10]. There has been significant success in integrating the individual tools for one activity into comprehensive systems in which information and data from one activity can be passed to another activity with little or no loss of fidelity and quality, e.g., the integration of computer-aided design with computer-aided manufacturing. The engineering integration process is not yet complete, but in today’s environment of global manufacturing concerns and multiple suppliers, production engineering and manufacturing are truly approaching a totally integrated and diversified enterprise.

The role of information and data on engineering materials is a critical component of the entire production cycle (all physical products are made from materials!), but in some ways it

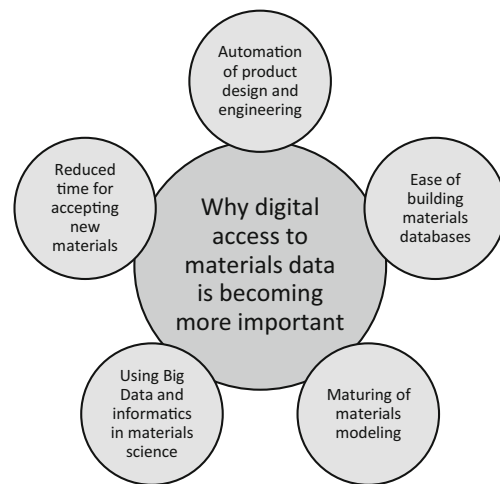


Fig. 1 Motivating factors driving better access digital materials data and databases

remains one of the least successful in terms of computerization and integration. It is apparent that any activity related to engineering materials, whether product design, materials selection, or manufacturing process planning, stands to benefit from access to computerized materials data. Yet the availability of materials data is both fragmented and incomplete. Within companies, individual departments often maintain and access different materials databases. Rarely is there a cohesive and comprehensive plan for ensuring access to needed materials data in support of CAE. The limited access to materials data inhibits extension of CAE and its associated business processes and acts as an obstacle to the capturing full benefits of the computer era.

Ease of Building Materials Databases

The information revolution—that is the combination of computers, telecommunications, software, and databases developed in the second half of the twentieth century—has produced a remarkable set of informatics tools that has made the computerization of materials data (in a manner similar to most other types of data) possible. Test equipment collects property measurements and not only stores those data in databases but also provides a suite of analytical tools to transform bits and bytes into meaningful physical quantities. Personal computers come with database management systems that allow individual scientists to manage, analyze, visualize, and store data efficiently with minimal training. The Internet, Web, and networking provide tools with which to share data with users and colleagues throughout the world almost trivially. Large data repositories gather data produced by diverse groups and published in a multitude of journals, allowing access to complete sets of related data. The task of building a materials database has never been easier [11, 12].

The very ease with which these tools can be created is a mixed blessing, however, because of the large number of similar yet not quite compatible tools. While there are many materials property databases, users are confronted with great difficulty in using them as a premeditated integrated resource. As an example, a recent survey of ceramic property data resources identified over 100 individual separate resources; none of which are integrated together [13] and no actual directory pointing to those databases exists. The same holds true for databases for metals, plastics, composites, nanomaterials, and other engineering materials.

Maturing of Modeling and the Need for Supporting Data

Physics-based modeling, which aims to describe and link materials behavior at different length scales, has huge potential for the design and development of materials that are required for evermore challenging applications [14–17]. This work is the basis of integrated computational materials engineering (ICME) [18], which holds great promise for better materials adapted into commercial use more quickly. These models require data for development and validation. They then require data sharing standards so that results at one length scale can be passed routinely to the next scale. Use of these models has been delayed in the absence of the needed benchmarking against experimental data collections, which itself depends on effective integration of modeling tools and databases [2]. We will not review the models in any depth but will briefly describe general characteristics at several length scales.

Material modeling at all scales both uses and generates large amounts of data. Some comprehensive collections of “fundamental” data are available (e.g., crystallographic data and potential energy curves), but other data important for modeling (e.g., elastic constants and electric and magnetic properties) are not. It should also be noted that with a few exceptions, the data generated by modeling usually are not made available through materials databases. The full value of materials modeling will not be realized until the data used by and generated during modeling have greater availability.

Emergence of New Materials and the Need to Speed Up Their Acceptance

The world of materials is exploding with new materials and new applications. Nanomaterials are entering the phase of their commercial adoption. Engineered biomaterials are close to that stage. The demand for higher performing electronic materials is growing. Structural materials that perform better under more extreme temperature, force, energy, and load conditions are in constant demand [19].

The flow of data and information on these advanced materials to designers and manufacturers is crucial to their acceptance, yet comprehensive data sources are lacking. One

negative impact of this situation is that information on emerging materials can be hard to find, resulting in significant delays in their adoption into products. The potential for modeling and integrated manufacturing to reduce time of adoption is significant, and the poor availability of data for new materials reduces that potential [20]. Improving this situation is a major goal of the U.S. Materials Genome Initiative [2].

Big Data and Informatics Tools That Allow Development of New Knowledge from Data

The rapid emergence of Big Data [5] as a hot topic has impacted materials data activities, and a number of workshops on the intersection of the two subjects have been held [21]. Big Data is often defined by the four data “Vs”: volume, velocity, variety, and veracity. While volume and velocity (speed of data acquisition) are less relevant for materials data, variety and veracity (or data quality) are, and they have been the subject of previous sections of this paper. What is especially important to note with respect to the impact of Big Data on materials data activities is the publicity that is being brought to all data activities. In particular, there is a new recognition that data collections have an importance beyond just archiving existing measurements and that data collections have the potential of supporting knowledge discovery activities [3, 22–24].

In parallel with Big Data, the field of scientific informatics has advanced in the last two decades. New tools to model, visualize, organize, and manage data have emerged that greatly aid materials data management [25–28]. Among these are ontology development and its support tools [29–31]. The complexity of materials metadata issues such as materials nomenclature, description of test procedures, and understanding analysis techniques means that successful use of ontologies must include materials experts who, unfortunately, are mostly unfamiliar with ontological approaches.

One feature in the development of Big Data and informatics is the maturing of tools to analyze data collections to extract new knowledge. Tools such as machine learning [32], deep learning [33], and other data-driven approaches [22] are becoming more common with increasing sophistication. It is particularly critical that for these approaches to work and produce meaningful results in materials science, complete and accurate materials data sets are available.

Brief Review of Materials Data and Databases

Before discussing accessibility to materials data, it is important to understand the various aspects of materials data and the databases that contain that data. The world of modern materials is large, diverse, and heterogeneous in a number of dimensions, and the data about materials reflect that diversity.

Consequently, materials data and databases can be viewed from a number of different perspectives, as shown in Table 1.

Database Perspective: Materials Properties

Structural (Crystallographic) Databases

The structure of a material is of fundamental interest in understanding and controlling properties. For materials with a regular periodic structure, the structure is characterized by crystallographic data. Computerized collections of crystallographic data are among the oldest scientific and technical databases. The first crystallography databases were built where programs for deconvolution of diffraction experiments led to building databases of crystal structures in the 1960s. The Cambridge Crystallographic Database was the first, and it collected full structural information on organic compounds [34]. This was followed by the Inorganic Crystal Structure Database [35]. In addition to being supported by the International Union of Crystallography (IUCr) with respect to standards for deposition and curation [36, 37], the data centers have traditionally charged fairly small fees for their use.

What is remarkable about the crystallographic databases is their completeness and coordination [38]. Because data have to be deposited in one of these databases (also considered repositories) before a research article is published, the incentive is high to make sure the data are deposited. While in recent years new online repositories have been created using Web technologies [39] [40], these standard databases remain fully engaged.

Phase Equilibria Databases

Metallic and ceramic materials usually change structure (phases) as a function of composition and temperature. The most definitive collections of binary and ternary alloy phase diagrams resulted from a decade-long (beginning in 1979) joint program by ASM International and the then National Bureau of Standards (NBS), supported in part by donations from industry [41]. These collections still provide materials scientists with fundamental phase data for these systems and

are available electronically. The primary set of ceramics phase diagrams is the result of a 70+-year collaboration between the American Ceramic Society and the National Institute of Standards and Technology (and its predecessor NBS). Based on a long-term publication series from the program, the Phase Diagrams for Ceramists Database is widely used by ceramists worldwide [42]. In more recent years, the continued progress of software-generated diagrams has supplemented experimentally determined diagrams, especially for higher-order systems [43–46].

Unlike for the case of crystallographic data, there are no central repositories into which new phase data are deposited, even though some journals are now requiring data deposition as a publication requirement, similar to that in the crystallographic community. Instead, the major collections have been built by extracting data from the open literature. The number of systems that are included in the data collections differ greatly—a few thousand binary and ternary alloy systems and a similar number of important ceramic systems versus the hundreds of thousands of crystallographic compounds that have been and are continually being generated.

Even though software-generated phase diagrams grow in number, the foundational knowledge base for phase diagrams is well established, and these phase diagram databases are not likely to grow substantially. This is in contrast to crystallographic databases, which continue to grow expansively as diffraction instruments become easier to use. This difference in size and growth rate between the two areas is also reflected in financial support requirements for the databases and the types of analysis tools being developed in conjunction with these databases. The crystallographic databases require greater support as they grow and expand. Further, the scientific opportunities for exploiting those crystallographic databases will naturally lead to new visualization, analytical, and predictive capabilities.

Thermal, Electrical, Optical, and Other Intrinsic Property Materials Databases

These important properties include thermophysical (coefficients of thermal expansion, thermal conductivity, etc.),

Table 1 Diverse perspectives for categorizing materials data and databases

| Perspective | Categories |
|------------------------|---|
| Materials properties | Structural (crystallographic), phase equilibria, thermal, electrical, optical, and other intrinsic properties, surface properties, failure (fatigue, tribology, corrosion), performance predictive (standardized tests) |
| Materials classes | Metals, alloys, ceramics, polymers, plastics, nanomaterials, composites, biomaterials, and others |
| Materials applications | Fundamental research, general characterization, design values, proprietary interests, failure analysis, EHS prediction, disposal |
| Interested parties | University, government laboratories, industry, defense, materials manufacturers, OEMs, testing laboratories, data collectors and providers |

electrical (conductivity, resistance, etc.), optical, elastic, magnetic, and other specialized properties. Many important sets of property data have been evaluated (for example [47]), but no coordinated effort has been undertaken to date to create comprehensive collections or databases of these properties. Many databases, however, include some of these data [48], but not systematically. For example, a review of ceramics databases showed that about 50% of the publicly available databases have some of these properties [13].

Surface Properties Databases

Surface properties databases fall into two major categories: surface analysis (characterization) and surface structure. Surface analysis databases include data on the composition and environment of the entities on a surface, which is critical for ascertaining the reactivity of surfaces. The NIST X-ray Photoelectron Spectroscopy Database was the first example [49]. Other surface analysis techniques have resulted in additional databases [50]. The structure of surfaces is important for designing catalysts and nanomaterials [51–53], and some of these data are available in databases. The growing interest and use of nanomaterials, for which surfaces are a major determinant of functionality, ensures that both surface composition and structure data will become increasingly important.

Performance Predictive Databases, with Standardized Tests, Including Failure Such as Fatigue, Tribology, and Corrosion

Many specialized collections of materials data are generated through standardized test methods, as shown, for example, by databases for metals [48], ceramics [13], and plastics [54]. Hundreds, if not thousands, of similar specialized materials databases can be found easily through a search of the Web. Today, however, few, if any, comprehensive databases or even comprehensive data directories for these property data exist for any material class, e.g., metals. It is useful to examine some of the reasons, as shown in Table 2, that historically have played a role in creating this rich, yet chaotic, situation.

Table 2 Factors challenging creation of comprehensive databases of materials performance prediction data

| Factor | Description |
|---------------------------------|---|
| Specialization | Numerous different tests for specific materials types (steels, aluminum, ceramics, etc.), performance characteristics (energy absorbing, hardness, fracture, etc.), or test conditions (temperature, environment, etc.) |
| Ownership of standardized tests | Hundreds of international, regional, national, and industry organizations |
| Proprietary issues | Test performed by organizations for proprietary reasons with reluctance to share results with competitors |
| Empirical nature of tests | Tests based on specified steps and not measuring intrinsic properties |

Specialization

Standardized testing of materials has been developed primarily to link easily obtained test results to accurate performance prediction, usually with some sort of safety factor included. Because materials in service are chosen for a large variety of performance characteristics—absorbing energy, deflecting force, preventing failure by wear, fatigue, or corrosion, to provide adequate strength, etc.—the development and prediction of materials performance has become very specialized. Specialization categories include materials type (metals, ceramics, polymers, composites of various types, etc.), applications (load bearing, energy absorption, electronic and magnetic performance, etc.), failure mechanisms, and performance criteria. This is especially true in critical applications, when the success of a product is determined by accurate prediction of material performance, and failure cannot be tolerated. Prior to the information age, these specialties were the subject of numerous hard copy handbooks and data tables, many of which have been directly converted into databases (*See* for example [9]). Very few efforts have been made to integrate these disparate databases into a comprehensive resource as has been done for crystallographic and phase data.

Ownership of Standardized Tests

The engineering materials community has done an outstanding job of developing needed tests on a non-proprietary basis, through national-, international-, and industry-specific formal and informal standards development bodies (SDOs). While this approach has in some sense maximized the use of knowledge spread across many companies and geographical areas, a side result is the plethora of actual and duplicative standard test methods. The vast majority of these methods have no specification for capturing test data and metadata in a standardized format. Even though most data are collected electronically through software on test equipment, collecting and homogenizing data from different test

methods is a time-consuming activity. In spite of numerous efforts, very little progress has been made to develop community-wide standards for materials performance data [55–57]. The SDOs that develop and maintain test method standards have little or no incentive to address data collection and exchange issues.

Proprietary Issues

The life cycle of materials data is complex and non-linear [58]. Many of the linked steps involve proprietary relationships that are well protected to ensure competitiveness and corporate well-being. This has two consequences. There are strong proprietary reasons for not making materials test data available, even though many companies have created internal databases containing test results for their own use. Companies also do not want others to know which materials they are interested in and what data they are using. They thereby limit their use of “publicly” available databases if not available to be installed for in-house use. This, in turn, has limited the market for more comprehensive, publicly-available databases of materials performance data. Many of the issues related to combining public and proprietary materials data have been discussed in a 2008 report from the National Research Council [45].

Empirical Nature of Tests

Most standardized materials performance tests have been based on a combination of empirical relationships and scientific principles, thereby inhibiting the growth of modeling as a source of data generation. There are a number of implications of these situations. The first is that small changes (e.g., compositional, processing, surface finishing) in a material may, in fact, lead to substantially different performance properties that are not easily predictable from existing models based on first scientific principles. The second implication is that it is difficult to develop predictive models for these tests such that the models span material types, test conditions, or performance environments, given the large number of independent variables that affect the measurement.

Given the difficulty in identifying all significant variables, the metadata requirements for careful documentation of a test can be quite large. For example, certain tests for composites have had several hundred metadata fields suggested for reporting [57]. Many standard test methods have specifications for specimen preparation and holding, loading rates, initial data analysis, and other parameters, including alternatives thereto, that require the reporting of many test parameters of different types [55, 56]. This makes comparisons of data from tests run by different investigators on different instruments at different times very difficult, again reducing the imperative for comprehensive databases.

Implications on Availability of Performance Test Data

As the result of the factors discussed above, the availability of comprehensive databases for performance test data is more limited than it might be otherwise. This is especially true with respect to the creation of comprehensive systems that could provide one-stop shopping for large amounts of these data. It is difficult to predict whether this will change significantly in the next few years, as it is not clear that users of these data are demanding greater access.

Database Perspective: Materials Classes

Most materials properties databases have focused on a specific materials class, especially for structural, phase equilibria, thermal/electronic properties, and standard test data. One obvious reason is that most databases are aimed at a specific user community rather than the general materials community. As most products can be classified in a single materials class—ceramic, metallic, plastic, and nanomaterials—the user in these cases is proficient with just that one type of material. This situation is common when designing to avoid or control materials failure in products, as different materials classes exhibit different failure mechanisms. Another reason for focus on a single materials class is that measurements are usually made by an expert in a single materials class. The standardized tests that generate most test data are produced by SDO committees that are almost always oriented to a single material class. Thus, most ceramic data are generated by ceramists; data on metals and alloys by metallurgists, and so on.

One major exception to the single materials class databases are comprehensive online materials data systems, which will be discussed later. The other exception is databases for multi-material classes to support materials selection [20, 59]. It should be noted that most materials selection software databases also usually focus on one materials class, such as plastics or metals.

Database Perspective: Materials Applications

A third perspective on materials databases is the purpose of the data collection, or what user interests are. Interests cover a broad range of applications that includes fundamental research, general characterization, design values, proprietary interests, failure analysis, and EHS prediction [60]. In the following paragraphs, we briefly look at how these different applications impact materials databases.

Fundamental Research

Most experiments done during the course of fundamental materials research are designed to gain understanding of some aspect of materials behavior [61]. Many lead to new

experiments that clarify or validate assumptions and build upon current understanding [62]. The data generated during these experiments are publishable in the archival literature and useful in documenting understanding, but rarely are of sufficient quality to be included in materials databases. If they are included, their associated uncertainties are difficult to ascertain. This is not to say that research data are not important, but that the major purpose is not to determine detailed properties, but rather to develop a better understanding of a phenomenon [63, 64].

General Characterization

Once a material is recognized as having potential for commercialization or other application, it is tested to generate a complete set of properties. These measurements are made by research institutes, companies, government labs, and testing houses, and the data generated are generally of high quality. Their availability is often limited, however, by patented interests, lack of circulation of published results (government and other kinds of reports, even though almost always electronic today, still are not widely noticed), and lack of appropriate data repositories (*See* for example Chap. 3 of [56]). As a result, even though much characterization is done, it is not always available.

Design Values

Several industries for which material failures cannot be tolerated, such as nuclear power, aerospace, and high-pressure vessels, have developed mechanisms to establish so-called design values for certain properties. The data are usually generated through specified testing protocols and analysis procedures. The resulting design data do not reflect an actual measurement result, but a recommended value based on analysis results and appropriate safety factors. Notable examples in the United States are the Military Handbook for aerospace metals [65] and composites [66] and the ASME boiler and pressure vessel code [67]. Most of the design value collections have been computerized and available on an ad hoc basis. There is no central directory for such resources, though users in the relevant industry are generally cognizant of their existence. Potential users of these high-quality data from other communities, however, are often unaware of their existence.

Proprietary Interests

Industry generates a great deal of materials data, and, with the exception of contributions to the calculation of design values, very little get released to the public. Many material producers maintain internal databases that they share with customers, though usually only those portions that directly affect a customer's purchasing decision. Producers also maintain product

description sheets that have “typical” values highlighting “attractive” features of an available material. Those data for plastics have sometimes been aggregated into public databases, but are not considered to be much more than marketing tools (*See* for example [68]).

Failure Analysis

Both materials producers and materials users maintain internal databases for failure analysis purposes. Few if any are publicly available. Various government agencies also have such databases, especially for advanced applications, including non-destructive testing results (*See* for example [69]).

Environmental, Health, and Safety Properties

The concern of possible environmental, health, and safety aspects of nanomaterials has given rise to efforts to develop standard tests and protocols for measuring these properties as well as accelerating development of the field of nanoinformatics. These include major European Union programs such as NanoReg [70], Future Nano Needs [71], and the Nanosafety cluster projects [72], United States efforts under the National Nanotechnology Initiative [73], including nanoinformatics programs funded by the National Institutes of Health [74], and other U.S. federal agencies; and standardization efforts by ISO Technical Committee 229 Nanotechnologies [75] and OECD Working Party on Manufactured Nanomaterials [76]. The focus is on developing standards for reporting data as well as demonstration databases. For traditional engineering materials, very few if any databases contain EHS-related properties.

Database Perspective: Interested Parties

Diverse communities are interested in materials data, including universities, government laboratories, industry, government agencies, materials manufacturers, testing laboratories, data collectors, and data providers. What should be apparent at this point is that few of these communities have a strong interest in publicly available comprehensive materials data systems. Proprietary interests are one major reason; specialized materials interests are another. One can say that most of these groups lack a strong business case for better materials data availability, though there are exceptions [20].

Comprehensive Online Materials Data Systems

In the 35 years since computerization of materials data has become a topic of major interest [7], a small number of efforts have tried to build comprehensive online systems with data on a wide variety of materials, properties, and sources. The most comprehensive effort was the National Materials Property

Data Network (NMPDN) in the late twentieth century. The prototype for the NMPDN was initially funded by NIST, the Department of Energy, and the Army, with the work being done at Lawrence Berkeley Laboratory and Stanford University [77]. It then was commercialized as the MPD Network by the Metals Properties Council [78] and later by Chemical Abstracts, but ceased operation in the late 1990s. During the same time, the European Demonstrator Project for Materials Data was put forward but never reached the commercialization stage [79]. While details about these systems can be found in the references cited, a few important conclusions can be put forward about these efforts and why they failed as well as the future of similar efforts.

Quite briefly, in the opinion of this author, they failed because of the effort required to put together a large enough collection of materials data to attract large numbers of users. The content and diversity of data content (at its largest, the MPD Network had a few tens of databases on a variety of materials) never reached the size necessary to generate enough user fees to sustain operation. One can ask why a comprehensive materials data system is needed in today's environment with powerful search engines and massive information archives that can quickly find millions of information resources on virtually any subject, including any material one can imagine. The present paradigm, however, does not work for materials data for the following reasons.

- Poor or non-existent data quality indicators
- Large volume of data with many duplicates, unknown sources, and poor documentation of test methods
- Lack of semantic content, limited and inconsistent metadata, inadequate display
- Difficulty in exchanging and merging data from different sources

The fragmented but very successful nature of today's Web and its search engines clearly demonstrates that a single integrated materials data system as described above is not only unnecessary but also impractical [80]. Easier and more comprehensive access to materials data, however, is still needed, and below we discuss critical issues, as shown in Fig. 2, involved in determining the success of such systems.

Comprehensiveness

The challenge of comprehensiveness is very difficult, given the multiplicity of potential data sources, which include peer-reviewed literature, manufacturers' data sheets, large and small scale testing programs that rarely get included in the archival resources, and the proprietary nature of much materials data. Yet that is what users want—the ability to find all available data for a specific material. The further the data type

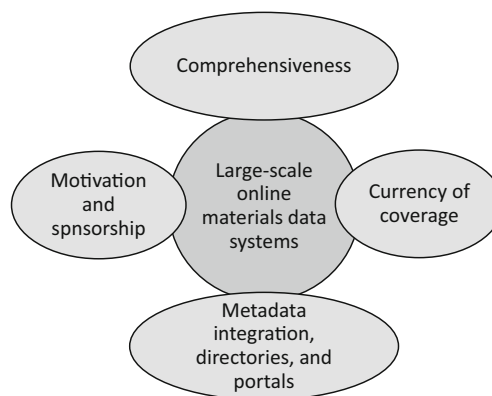


Fig. 2 Challenges for success of large-scale online materials data systems

is from fundamental physical data and the closer to complex test results, the more challenging comprehensiveness becomes, yet the more desirable the data.

One solution is to emphasize “reliable” data, which could be described as data that have been carefully selected for their pedigree and adherence to test quality standards [64]. This provides a more nuanced meaning of the term “comprehensive,” but one that is operationally slightly more reachable. One other aspect related to comprehensiveness that needs to be mentioned is the international nature of materials data. Given today's international marketplace, many materials have lost their geo-specificity, but through language and customary practices, data on those materials do not easily cross national borders.

Currency of Coverage

The task of creating a comprehensive online materials data system is compounded by the steady growth of more data on a growing number of materials. If a system is composed of a number of individual databases built and maintained by separate groups, then the effort to keep each of them up-to-date is remarkably difficult. Freiman's recent surveys of ceramics property data showed that the period of coverage of most available databases is extremely difficult to determine. Most of the databases identified have obvious coverage cut-off date years old [13].

A second aspect of the currency problem is related to the constant evolution of test methods themselves and the metadata connected therewith. Data generated under an older method may not be compatible with that generated under a new version of the “same” method, but the differences may be difficult, or impossible, to detect, especially as changes to test methods are not usually tracked by database providers. To date, automated data and metadata extraction have not been successfully applied to materials literature, though new approaches are being tried [81].

Metadata Integration, Database Directories, and Portals

When data within an online system have been put together by a single agent from multiple sources, the task of metadata integration comes into play. The task of integrating databases built and maintained by different groups is possible, either by choosing one metadata system as the “standard” and integrating others into it or else by developing a neutral metadata system that each individual database is translated into and from. The expectation is that after a sufficiently large number of databases have been integrated, the task becomes incrementally less taxing. Most terms and materials are already in the online system metadata dictionary. For a recent review of previous formal attempts at metadata integration, *See* Chap. 5 of [56].

In practice, given the large number of materials and especially the large number of properties and independent variables that need to be accounted for, the task does not seem to become easier. The lack of comprehensive materials database directories and portals (one-stop shopping) is a clear indication of the difficulties involved in indexing, harmonizing, and integrating individual databases into a system. While some effort is being put into using semantic Web technology to facilitate more detailed searching by modern search engines, it remains to be seen if material semantics are amenable to this approach [6, 82].

Motivation and Sponsorship

Online materials data systems have been developed for a number of reasons, including profitability, public service, support of national industry, and to advance the discovery of new materials. Each reason imposes different characteristics to the online system in terms of properties included, materials classes included, metadata used, analytical tools attached, and user interfaces developed. Also many companies have built internal materials data systems to support to their business; again these systems display features strongly dependent on the industry involved. It remains to be seen if any online system can approach the comprehensiveness and currency needed to perpetuate itself beyond a decade or so.

Different types of sponsorship for online data systems have been used, from government support to private investors. Government sponsorship sometimes is questioned when the primary goal is use by industry, with the feeling that industry itself should both invest and provide long term support for something that directly aids their bottom-line profitability. At the same time, private investors do not easily see that profitability will happen in a time period that is acceptable; though as shown for many of the databases discussed in this paper, private organizations are aggressively building individual data resources of many types. One contributing factor to the long-term support issue is the lack of glamor associated with an

online materials data system. “Why cannot you just use Google™?” is the question often asked, even though such systems do not provide any meaningful metadata integration nor useful data quality indicators.

Contemporary Efforts

The last few years have seen a global resurgence in interest in materials databases.

- The Materials Genome Initiative in the United States has focused on more rapid commercialization of new materials
- The European Standardization Organization is addressing materials data exchange approaches
- Open access policies are leading to new data repositories
- Nanomaterials informatics is critical in assessing EHS impacts on an international scale
- Big Data tools and new informatics approaches are coming to computational materials science

In this section, we briefly discuss these new materials data initiatives. The following section identifies some of the challenges they are facing and possible approaches to meeting those challenges.

Materials Genome Initiative

The Materials Genome Initiative (MGI) was launched in 2011 as a multi-federal agency effort of the U.S. Government to invest in research, tools, and prototypes for advancing next generation materials development and commercialization [2, 83, 84]. One of the major goals was to reduce the time for adoption for new materials from decades to less than a decade, especially through the development of advanced modeling (for example, *See* [85]). The generation and availability of materials data is a key component of this effort [61, 86].

In 2014, the MGI launched an open Materials Data Facility pilot as part of the National Data Service to boost data access and sharing, a consortium of research universities, national laboratories, and academic publishers [87]. This effort represents a major step forward in providing comprehensive access to materials data. At the same time, however, the issues outlined in this paper, including the proprietary nature of much materials data, the complexity of materials, materials properties and their associated metadata, and the commercial value of materials data themselves, must be addressed for this initiative to succeed.

Among the efforts included in the MGI is the Materials Data Curation System [88], which provides a mechanism for converting a wide variety of materials data into portable formats (e.g., XML, JSON) to improve data sharing and other uses.

European Workshops

The European Standardization Committee (CEN) has supported a series of projects—called Workshops in their parlance—to address issues related to the exchange of engineering materials data [55, 56, 89, 90]. The Workshops focus on the exchange of engineering materials data and feature close partnerships among materials scientists, information specialists, and industry materials experts to develop real-life technologies for sharing data. These are built on earlier standards work under ASTM and ISO [57].

Open Access Is Leading to Materials Data Repository Requirements by U.S. Funding Agencies

In 2010, the U.S. Federal Government began efforts to require the sharing of publicly funded research [91]. Federal agencies have established a variety of approaches. The National Institutes of Health have, for example, created an extensive array of data repositories for their different institutes and research areas [92]. Of particular interest to the field of materials data are the plans by the National Science Foundation to require data management plans for all new materials research proposals [93]. While data repositories for some types of S&T data are being created, the only mature examples in materials data are for crystallographic and thermochemical data, as discussed above.

The Emergence of Nanoinformatics

The scientific, technical, and commercial promise of nanomaterials has led to an explosive growth of research in this area. One area of great interest is the impact of nanomaterials on terms of environmental, health, and safety concerns. In support of the development of predictive techniques for EHS impact, the field of nanoinformatics has emerged, with considerable emphasis on building high-quality data repositories [29, 94, 95]. One interesting aspect of nanoinformatics is the collaboration between materials data and bioinformatics experts, which has resulted in the sharing of data tools from their different disciplines [96–98]. Though nanomaterials exhibit unique properties because of their size and reactive surfaces, they still are materials, and as such, the technologies important for traditional materials data are important in nanoinformatics.

Big Data and Modern Informatics

As discussed in “Why Digital Access to Materials Data is Becoming More Important,” Big Data and modern informatics open the possibility of discovering new knowledge and understand from existing data sets. While new analytical tools, including those for machine and deep learning, are being

aggressively developed both for general use and materials science specific applications, the need for complete and accurate evaluated data sets increases. Knowledge based on inaccurate data is not very reliable.

The FAIR Principles and Materials Data

FAIR Principles

In a recent seminal paper [99], a set of principles—the FAIR Guiding Principles for scientific data management and stewardship—have been enumerated. The four foundational principles are: Findability, Accessibility, Interoperability, and Reusability. It is instructive to draw upon the previous discussion and identify how these principles can be used in looking at some of the issues facing the materials data community in the coming years. We examine each of the principles in turn from the perspective of materials data.

Findability, also known as *discoverability*, is naturally the first key factor in using data and one that poses critical problems for materials data. We presented a vision at the beginning of this article of having “one-stop” access to large amounts of materials data for all users. This concept envisions having a single or small number of data portals, as found in other scientific disciplines, to a wide variety of data for a wide variety of user communities. The portal itself could access one or more *comprehensive centralized* systems, connect to *federated systems* with loosely linked, multiple data resources, or even simply be a *semantic-Web-based search system* with no special access to identified data resources. Another possibility is a portal that is a *register of databases*, similar to that developed by the Australian National Data Service [100] and the United States [11, 88, 101]. An issue with database registries is the difficulty in providing detailed and current lists of contents for the databases that have been registered for reasons such as described above. A third possibility, as suggested by the FAIR Guiding Principles, is a globally unique and persistent identifier for all metadata and data; though for materials, no meaningful steps have been taken.

Accessibility addresses the ability of users to retrieve data easily and using standard procedures. The present diversity of materials data and an equally large diversity of materials data resources present significant challenges to accessibility. With business cases for greater uniformity of access not well defined, given the commercial value of much materials data, there is little motivation for data providers to look beyond accessibility except in terms of their own data resource (for example, See [82]).

Interoperability of materials data is critical in today’s world of CAE. The broad range of data types and resources has provided strong challenges to making materials data interoperable. Numerous standards committees have worked in

different venues to put some degree of interoperability standards into place, especially in the context of materials testing and integration with CAE frameworks [55, 56, 94], but the lack of business cases has again hindered success. Some of the technical challenges that have to be overcome are discussed below.

Reusability refers both to the adequacy of metadata associated with materials data as well as appropriate data usage licensing. Metadata standards are still lacking for most materials data, though progress is slowly being made. More importantly, the commercial value of much materials data has led to quite restrictive data usage regimes.

Materials Data Challenges to FAIR

Below, we discuss seven key features of the materials data landscape, as shown in Fig. 3, that strongly affect the implementation of the FAIR Guiding Principles for materials data.

Diversity of Materials Data

Materials data are not homogeneous. They span the diversity of materials themselves, from nanomaterials of a few hundred atoms to bulk materials with Avogadro's number of atoms and more. They include metals and alloys, ceramics, polymers, and composites of all these. A similar diversity of properties means that each property has different metadata associated with it. The data themselves can range from raw measurements to published results to nominal values to design values. Because of this diversity of materials and property types, solutions for collecting, managing, disseminating, accessing, and using materials data require multiple approaches and methods. In turn, the expertise to build collections of diverse types of materials data that are accessible through a single portal is itself dispersed. Harmonizing and integrating

nomenclature, metadata, and test results remains a major challenge (for example, *See* [25, 30, 102, 103]).

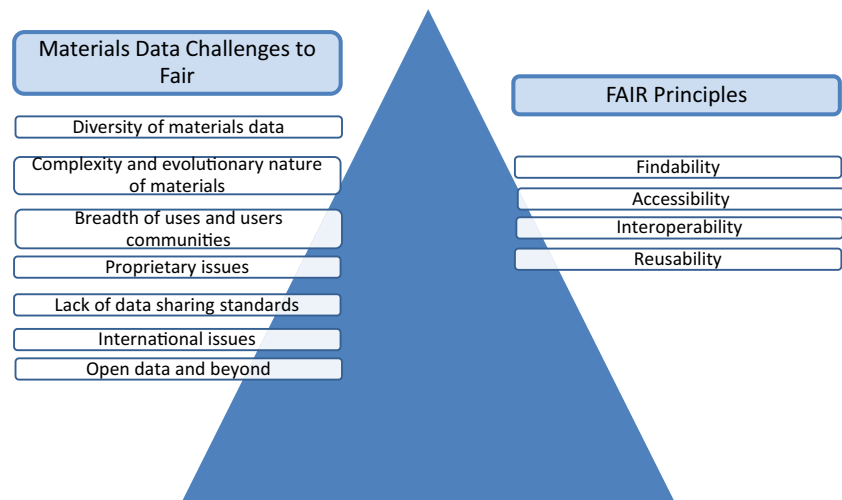
Complexity and Evolutionary Nature of Materials

Engineering materials are not static entities. Materials are used in products to provide specific product performance and small changes in a material can significantly affect that performance. Consequently, materials developers and producers are constantly looking for commercial advantages by altering and improving their materials. While attempts have been made to standardize the composition and structure of many materials, their producers still continuously seek to make improvements, such as through surface modification and slight compositional changes. What is an improved material today can easily become the standard material of tomorrow. In the case of more specialized materials, such as electronic materials or nanomaterials, the only materials standardization is through a commercial agreement between manufacturer and purchaser. Because the processing parameters and resulting compositions and performance are proprietary secrets, there is little incentive to share such information. The changing nature of materials means that materials data resources go out of date rapidly and having data on the newest materials becomes a major challenge.

Breadth of Uses and User Communities

The diversity of materials is matched by the diversity of uses. Every tangible object is made of a material. Use can involve highly controlled situations such as aircraft, high-pressure vessels, food packaging, and human implants. The materials data in these cases is carefully scrutinized and often subject to certification. Other uses have no such requirements, and the average ashtray producer does not spend much time on the quality of materials data. The range of uses between these

Fig. 3 FAIR principles and materials data challenges in meeting them



extremes is almost infinite, and this breadth of use is a major challenge. The types of materials data collections needed by different user communities impose different requirements for materials data systems, including data quality [64], presentation, documentation, uniformity, completeness, visualization, and standardization. Again, as a result, existing data resources are often incompatible in these features, thereby hindering their integration into a more comprehensive system. In many ways, the breadth of uses and user communities for materials data is more complex than the data themselves, resulting in additional challenges in building and disseminating materials databases [104, 105].

Proprietary Issues

Materials data have significant commercial value in many cases, and large amounts of materials data are generated in proprietary situations for that reason. Those data rarely get disseminated beyond corporate boundaries. As tools for predicting data (property prediction) and knowledge discovery evolve, their commercial potential obviously increases. Care must be taken to ensure a balance between public and proprietary interests [18].

Lack of Data Sharing Standards

Issues related to standards for materials data exchange and sharing have been reviewed recently [56]. The number of committees and other organizations involved in developing test method standards is quite large. As a result, for data format standards for materials data to evolve, a large number of groups have to be involved. To get metadata standards across material types, tests, and test committees is a significant challenge. A strong business case for materials data standards has yet to be made. For standards for data repositories, the situation may be better. These can be developed by the group(s) developing, controlling, and participating in the repository, which is a more coherent community (See for example [106, 107]). A greater issue here is to have coordination among the multiple repositories that are likely to arise.

International Issues

Materials have long been an international commodity and with the globalization of manufacturing, even more so today. Materials data are consequently equally an international commodity, though subject to significant constraints due to language, technical, and IPR issues. Perhaps, the technical issues are most difficult in that different countries have different specifications for materials that are essentially the same. One area in which international considerations is a major challenge is with materials test and data standards. The multiplicity of national and international standards development organizations

has made harmonization of test methods a lengthy and difficult process. While ISO and ASTM standards have been adopted in many situations, national test standards are still widely used. The same situation applies to materials data standards. Again, the existence of overlapping committees under different jurisdictions reduces the incentive to come up with harmonized data standards.

A final issue is related to the economic value of materials data themselves. Materials data resources are valuable to companies, and they are willing to pay significant fees for access to high-quality materials data. There is little incentive for countries to encourage materials data resources located in one country to reach out to similar organizations in other countries. This is especially true for data resources developed, built, and controlled by a national government [108–111].

Open Data and Beyond

Over the last 15 years, the movement towards open science, that is, the philosophy that publicly funded science is an economic resource that must be made available to everyone, has gained momentum and acceptance. As a corollary, the open data movement asserts that research data generated through public support should also be freely and openly available. As a result, government agencies throughout the world are demanding that researchers must share their research data [91]. One result is the growth of data management plans and data repositories as described previously. To date, this has had little impact on materials data, but that will change over the years. A challenge to a full commitment to open data is the cost of operating and maintaining data repositories over the long term, which is not a small number of years but a large number of decades. Repositories are expensive as data volumes increase, storage media changes, and dissemination technology advances. It remains to be seen how the cost issue will be resolved [112].

For materials data, the questions of proprietary and direct economic value also impact open data approaches. In areas of advanced materials development, such as for electronics and nanomaterials, even fundamental property data are enormously important and well protected, thus, challenging the spirit of open data.

Thoughts on the Future of Materials Data Access

In spite of the optimistic vision expressed at the beginning of this paper in terms of easy access to high-quality materials data, users of materials data still have significant difficulties in finding and using materials for the above-mentioned reasons. Much progress has been made, but much more is needed. We have reviewed many aspects of computerized materials data, especially those affecting accessibility. We have

tried to demonstrate that the diverse nature of materials, materials data, and users of materials data brings additional dimensions of complexity to data collection, management, and dissemination, all impacting accessibility. At the same time, the economic value of materials data is hard to overestimate. The first step to handling this complexity is recognizing its existence. Once that is done, solutions can be found to address its different dimensions.

We believe that new approaches to improving the quality and availability of materials data will continue to grow, including the ability to access and share materials easily and integrate them with other scientific and engineering software. The materials community expects progress, and the new initiatives and technologies, addressing the issues described above, should enable that progress.

Acknowledgements The author wishes to thank numerous colleagues for their thoughts, ideas, and discussion over the years, including Steve Freiman, Timothy Austin, Jack Westbrook, J. Gilbert Kaufman, and David Lide. In addition, the author thanks the reviewers for helpful suggestions.

Author's Contribution The author is the sole contributor to this paper.

Compliance with Ethical Standards

Disclaimer The mention of specific privately owned and operated data resources implies neither endorsement nor criticism. These data resources are mentioned as examples or for illustrative purposes.

Availability of Data and Materials Not applicable.

Competing Interests The author declares that there is no competing interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bureau of Economic Analysis USD of C (2016) Gross-domestic-product-(GDP)-by-industry-data, gross output 1947–2014, up to 71 industries, primary metals. In: Gross Domest. Prod. GDP Ind. Data. http://www.bea.gov/industry/gdpbyind_data.htm
- National Institute of Standards and Technology (2016) Materials Genome Initiative. In: Mater. Genome Initiat. <https://mgi.nist.gov/federalMgi>
- Agrawal A, Choudhary A (2016) Perspective: materials informatics and big data: realization of the “fourth paradigm” of science in materials science. *APL Mater* 4:053208
- Hill J, Mulholland G, Persson K et al (2016) Materials science with large-scale data and informatics: unlocking new opportunities. *MRS Bull* 41:399–409
- Lynch C (2008) Big data: how do your data grow? *Nature* 455:28–29
- Jaykumar N, Yallamelli P, Nguyen V, et al KnowledgeWiki: an open source tool for creating community-curated vocabulary, with a use case in materials science
- Westbrook JH, Rumble JR Jr. (1983) Computerized materials data systems. National Bureau of Standards
- Glazman JS (1989) Computerization and networking of materials data bases. ASTM International
- Chryssolouris G, Mavrikios D, Papakostas N et al (2009) Digital manufacturing: history, perspectives, and outlook. *Proc Inst Mech Eng Part B J Eng Manuf* 223:451–462
- Beckmann B, Giani A, Carbone J et al (2016) Developing the digital manufacturing commons: a national initiative for US manufacturing innovation. *Procedia Manuf* 5:182–194
- Bass J (2014) NIST Materials Resource Registry
- Dima A, Bhaskarla S, Becker C et al (2016) Informatics infrastructure for the Materials Genome Initiative. *JOM* 68:2053–2064
- Freiman S, Rumble J (2013) Current availability of ceramic property data and future opportunities. *Am Ceram Soc Bull* 92:34–39
- Kirklin S, Saal JE, Meredig B et al (2015) The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. *NPJ Comput Mater* 1:15010
- Odoh SO, Cramer CJ, Truhlar DG, Gagliardi L (2015) Quantum-chemical characterization of the properties and reactivities of metal–organic frameworks. *Chem Rev* 115:6051–6111
- Raghavachari K, Saha A (2015) Accurate composite and fragment-based quantum chemical models for large molecules. *Chem Rev* 115:5643–5677
- Jha R, Dulikravich GS, Colaco MJ, et al (2017) Magnetic alloys design using multi-objective optimization. In: *Prop. Charact. Mod. Mater.* Springer, pp 261–284
- Council NR, others (2008) Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security. National Academies Press
- Cheung KWS (2009) Developing materials informatics workbench for expediting the discovery of novel compound materials
- Ashby M (2011) Hybrid materials to expand the boundaries of material-property space. *J Am Ceram Soc* 94
- Mellody M (2014) Big Data in materials research and development: summary of a workshop. National Academies Press
- Rajan K (2008) Combinatorial materials sciences: experimental strategies for accelerated knowledge discovery. *Annu Rev Mater Res* 38:299–322
- Ghiringhelli LM, Vybiral J, Levchenko SV et al (2015) Big Data of materials science: critical role of the descriptor. *Phys Rev Lett* 114:105503
- Broderick SR, Santhanam GR, Rajan K (2016) Harnessing the Big Data Paradigm for ICME: shifting from materials selection to materials enabled design. *JOM* 68:2109–2115
- Kalidindi SR, Medford AJ, McDowell DL (2016) Vision for data and informatics in the future materials innovation ecosystem. *JOM* 68:2126–2137
- Stukowski A (2009) Visualization and analysis of atomistic simulation data with OVITO—the open visualization tool. *Model Simul Mater Sci Eng* 18:015012
- Obkhodsky A, Kuznetsov S, Popov A, et al (2017) Data visualization tools for materials properties research. In: *MATEC Web Conf. EDP Sciences*, p 00014
- Momma K, Izumi F (2011) VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J Appl Crystallogr* 44:1272–1276
- Hastings J, Jeliakova N, Owen G et al (2015) eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. *J Biomed Semant* 6:1

30. Cheung K, Drennan J, Hunter J (2008) Towards an ontology for data-driven discovery of new materials. In: AAAI Spring Symp. Semantic Sci. Knowl. Integr, pp 9–14
31. Ashino T (2010) Materials ontology: an infrastructure for exchanging materials information and knowledge. *Data Sci J* 9:54–61
32. Meredig B, Agrawal A, Kirklin S et al (2014) Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys Rev B* 89:094104
33. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
34. (2016) Cambridge Crystallographic Data Centre. In: Camb. Crystallogr. Data Cent. <http://www.ccdc.cam.ac.uk/>
35. FIZ Karlsruhe (2016) Inorganic Crystal Structure Database. In: *Inorg. Cryst. Struct. Database*. <https://www.fiz-karlsruhe.de/en/leistungen/kristallographie/icstd.html>
36. Hall SR, McMahon B (2005) International tables for crystallography, definition and exchange of crystallographic data. Springer Science & Business Media
37. International Union of Crystallography (2016) Crystallographic Information Framework (CIF). In: *CIF*. <http://www.iucr.org/resources/cif>
38. Mighell AD, Karen VK (1996) NIST Crystallographic Databases for Research and Analysis. *J Res-Natl Inst Stand Technol* 101:273–280
39. Gražulis S, Daškevič A, Merkys A et al (2012) Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res* 40:D420–D427
40. (2016) Crystallography Open Database. In: *Crystallogr. Open Database*. <http://www.crystallography.net/cod/>
41. (2016) ASM Alloy Phase Diagram Database. In: *ASM Alloy Phase Diagram Database*. <http://mio.asminternational.org/apd/index.aspx>
42. American Ceramic Society (2016) American Ceramic Society—NIST Phase Equilibria Diagrams Program. In: *Phase Equilibria Diagrams*. <http://ceramics.org/publications-and-resources/phase-equilibria-diagrams>
43. Spencer PJ (2008) A brief history of CALPHAD. *Calphad* 32:1–8
44. Bale CW, Bélisle E, Chartrand P et al (2016) FactSage thermochemical software and databases, 2010–2016. *Calphad* 54:35–53
45. Andersson J-O, Helander T, Höglund L et al (2002) Thermo-Calc & DICTRA, computational tools for materials science. *Calphad* 26:273–312
46. Kattner UR (1997) The thermodynamic modeling of multicomponent phase equilibria. *JOM* 49:14–19
47. Ho CY, Powell RW, Liley PE (1972) Thermal conductivity of the elements. *J Phys Chem Ref Data* 1:279–421
48. CINDAS LLC. <https://cindasdata.com/>. Accessed 16 Feb 2017
49. Lee AY, Blakeslee DM, Powell CJ, Rumble J (2002) Development of the web-based NIST X-ray Photoelectron Spectroscopy (XPS) Database. *Data Sci J* 1:1–12
50. Yoshikawa H, Yoshihara K, Watanabe D et al (2014) Proposal for common data transfer format for simulation softwares used in surface electron spectroscopies. *Surf Interface Anal* 46:931–935
51. Watson PR, Van Hove MA, Hermann K (1994) Atlas of surface crystallography based on the NIST Surface Structure Database (SSD). *J Phys Chem Ref Data Monogr*
52. Watson PR, Van Hove MA, Herman K (1995) Atlas of surface structure. [Volume IA, Monograph 5]. ACS Publications, Washington, DC
53. Van Hove MA, Hermann K, Watson PR (1997) The Surface Structure Database: SSD. *Surf Rev Lett* 4:1071–1075
54. CAMPUSplastics. <http://www.campusplastics.com/>. Accessed 17 Apr 2017
55. European Committee for Standardization (CEN) (2010) A Guide to the Development and Use of Standards Compliant Data Formats for Engineering Materials Test Data, CWA 16200:2010 (E). Technical Specifications. ftp://ftp.cen.eu/CEN/Sectors/List/ICT/CWAs/CWA16200_2010_ELSSI.pdf. Accessed 1 June 2017
56. European Committee for Standardization (CEN) (2016) ICT Standards in Support of an eReporting Framework for the Engineering Materials Sector, CWA 16762:2014 (E). ftp://ftp.cencenelec.eu/CEN/WhatWeDo/Fields/ICT/eBusiness/WS/SERES/CWA_16762_2014_SERES.pdf. Accessed 1 June 2017
57. Newton CH (1993) Introduction to the Building of Material Databases. *Man Build Mater Databases ASTM Man Ser MNL* 19:1–12
58. Mulholland GJ, Paradiso SP (2016) Perspective: materials informatics across the product lifecycle: selection, manufacturing, and certification. *APL Mater* 4:053207
59. Granta Design (2016) <https://www.grantadesign.com/>. Accessed 1 June 2017
60. O'Hare J (2013) Material selection: taking environmental business risks into account. 24th Adv. Aerosp Mater Process AeroMat Conf Expo
61. Ward CH, Warren JA, Hanisch RJ (2014) Making materials science and engineering data more valuable research products. *Integrating Mater Manuf Innov* 3:1
62. Ward L, Wolverton C (2016) Atomistic calculations and materials informatics: a review. *Curr Opin Solid State Mater Sci*
63. Michel K, Meredig B (2016) Beyond bulk single crystals: a data format for all materials structure–property–processing relationships. *MRS Bull* 41:617–623
64. Munro RG (2003) Data evaluation theory and practice for materials properties. Commerce Department
65. Metallic Materials Properties Development and Standardization Committee (2015) MMPDS-10, Metallic materials properties development and standardization (MMPDS) handbook. Battelle
66. Handbook M (2002) MIL-HDBK-17-2F: Composite materials handbook. *Polym Matrix Compos Mater Usage Des Anal* 17
67. (2016) ASME BPVC 2015 Boiler and Pressure Vessel Code. In: *Boil. Press. Vessel Code 2015 Version*. <https://www.asme.org/shop/standards/new-releases/boiler-pressure-vessel-code-2013>
68. Online materials information resource—MatWeb. <http://www.matweb.com/>. Accessed 17 Apr 2017
69. Francesco ED, Francesco RD, Leccese F, Cagnetti M (2016) A proposal to improve the system life cycle support of composites structures mapping zonal testing data on LSA Databases. In: 2016 I.E. Metrol. Aerosp. MetroAeroSpace. pp 151–155
70. Netherlands National Institute for Public Health and the Environment (RIVM) (2016) NANoReg Results Repository. <http://www.nanoreg.eu/>. Accessed 1 June 2017
71. Centre for BioNano Interactions (CBNI), School of Chemistry and Chemical Biology, University College Dublin (2016) EU FP7 FutureNanoNeeds. <http://www.futurenanoneeds.eu/>. Accessed 1 June 2017
72. EU Directorate General for Research & Innovation (2016) NanoSafety Cluster. <http://www.nanosafetycluster.eu/>. Accessed 1 June 2017
73. U.S. National Nanotechnology Coordination Office (NNCO) (2016) National Nanotechnology Initiative. *Nano.gov*. <http://www.nano.gov/>. Accessed 1 June 2017
74. National Cancer Institute NI of H (2016) caNanoLab. In: *caNanoLab*. <https://cananolab.nci.nih.gov/caNanoLab/#/>
75. International Organization for Standardization, ISO Technical Committee 229 Nanotechnologies. http://www.iso.org/iso/iso_technical_committee?commid=381983
76. OECD Working Party on Manufactured Nanomaterials (2016) OECD Working Party on Manufactured Nanomaterials. <http://www.oecd.org/science/nanosafety/>. Accessed 1 June 2017

77. Grattidge W, Westbrook J, McCarthy J, et al (1986) Materials Information for Science and Technology (MIST): project overview: phases I and II and general considerations. Sci-Tech Knowledge Systems, Scotia, NY (USA); Lawrence Berkeley Lab., CA (USA); Sandia National Labs., Albuquerque, NM (USA); National Bureau of Standards, Washington, DC (USA)
78. Kaufman JG (1989) The National Materials Property Data Network, Inc.—a cooperative national approach to reliable performance data. *Comput Netw Mater Data Bases*
79. Swindells N, Waterman N, Krockel H (1990) Materials information for the European communities. *Rep EUR* 13153
80. Kalidindi SR, De Graef M (2015) Materials data science: current status and future outlook. *Annu Rev Mater Res* 45:171–193
81. Seshadri R, Sparks TD (2016) Perspective: interactive material property databases through aggregation of literature data. *APL Mater* 4:053206
82. O'Mara J, Meredig B, Michel K (2016) Materials data infrastructure: a case study of the citrination platform to examine data import, storage, and access. *JOM* 68:2031–2034
83. Jain A, Ong SP, Hautier G et al (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. *Apl Mater* 1:011002
84. Jain A, Persson KA, Ceder G (2016) Research update: the Materials Genome Initiative: data sharing and the impact of collaborative ab initio databases. *APL Mater* 4:053102
85. Wong TT (2016) Building a materials data infrastructure. *JOM* 68:2029–2030
86. Jacobsen MD, Fourman JR, Porter KM et al (2016) Creating an integrated collaborative environment for materials research. *Integrating Mater Manuf Innov* 5:12
87. National Institute of Standards and Technology (2016) Materials Data Facility. In: *Mater. Data Facil.* <https://materialsdatafacility.org/>
88. National Institute of Standards and Technology (2016) Materials Data Curation System. In: *Mater. Curation Syst.* <https://mgi.nist.gov/materials-data-curation-system>
89. CEN Workshop on Standards Compliant Formats for Fatigue Test Data—FATEDA. <https://www.cen.eu/work/areas/ICT/eBusiness/Pages/WS-FATEDA.aspx>. Accessed 16 Feb 2017
90. CEN WS MeTeDa on mechanical test data. <https://www.cen.eu/news/workshops/Pages/WS-2016-011.aspx>. Accessed 16 Feb 2017
91. Office of Science and Technology Policy Increasing Access to the Results of Federally Funded Scientific Research. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
92. (2016) NIH Data Sharing Repositories. In: *NIH Data Shar. Repos.* https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
93. Directorate of Mathematical and Physical Sciences Division of Materials Research (DMR) Advice to PIs on Data Management Plans
94. Thomas DG, Gaheen S, Harper SL et al (2013) ISA-TAB-Nano: a specification for sharing nanomaterial research data in spreadsheet-based format. *BMC Biotechnol* 13:1
95. (2016) Alexander Tropsha (2016), "Nanomaterial Registry: present and future. <https://www.nanomaterialregistry.org/>. Accessed 1 June 2017
96. Hendren CO, Powers CM, Hoover MD, Harper SL (2015) The Nanomaterial Data Curation Initiative: a collaborative approach to assessing, evaluating, and advancing the state of the field. *Beilstein J Nanotechnol* 6:1752–1762
97. Richard LRM, Lynch I, Peijnenburg W et al (2016) How should the completeness and quality of curated nanomaterial data be evaluated? *Nano* 2016:25
98. Lowry GV, Hill RJ, Harper S et al (2016) Guidance to improve the scientific value of zeta-potential measurements in nanoEHS. *Environ Sci Nano* 3:953–965
99. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3
100. Australian National Data Service (2016) Australian National Data Service. <http://www.ands.org.au/>. Accessed 1 June 2017
101. National Data Service Consortium (2016) National data service. In: *Natl. Data Serv.* <http://www.nationaldataservice.org/>.
102. Austin TS, Over H-H (2012) MatDB Online—a standards-based system for preserving, managing, and exchanging engineering materials test data. *Data Sci J* 11:ASMD11-ASMD16
103. Rajan K (2008) Materials informatics part I: a diversity of issues. *JOM* 60:50–50
104. Austin T (2016) Towards a digital infrastructure for engineering materials data. *Mater Discov*
105. Lin L, Austin T, Ren W (2015) Interoperability of materials database systems in support of nuclear energy development and potential applications for fuel cell material selection. *Mater Perform Charact* 4:115–130
106. Curtarolo S, Setyawan W, Wang S et al (2012) AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput Mater Sci* 58:227–235
107. Nolan JW, Gkika DA, Vordos N et al (2015) On the archiving and visualisation of scientific data. *J Eng Sci Technol Rev* 8:40–43
108. NIMS (Japan) (2016) MATNavi NIMS Materials Database. In: *MATNavi NIMS Mater. Database.* http://mits.nims.go.jp/index_en.html
109. Gao Zhi-yu LG (2013) Recent progress of web-enable material database and a case study of NIMS and MatWeb. *J Mater Eng* 11: 89–96. doi:10.3969/j.issn.1001-4381.2013.11.015
110. YIN H, ZHANG R, LIU G et al (2014) Development of the material databases. *J Chin Ceram Soc* 1:007
111. Korea Materials Center (2016) Korea Materials Center. <http://www.matcenter.org/engMain.do?cmd=mainView>. Accessed 1 June 2017
112. Hodson Molloy current best practice for research data management policies. Zenodo. doi:10.5281/zenodo.27872