Big Semantic Data Processing in the Materials Design Domain

Patrick Lambrix, Rickard Armiento, Anna Delin, Huanyu Li

Definitions

To speed up the progress in the field of materials design, a number of challenges related to big data need to be addressed. This entry discusses these challenges and shows the semantic technologies that alleviate the problems related to Variety, Variability and Veracity.

Overview

Materials design and materials informatics is central for technological progress, not the least in the green engineering domain. Many traditional materials contain toxic or critical raw materials, whose use should be avoided or eliminated. Also, there is an urgent need to develop new environmentally friendly energy technology. Presently, relevant examples of materials design challenges include energy storage, solar cells, thermoelectrics, and magnetic transport (Ceder and Persson (2013); Jain et al (2013); Curtarolo et al (2013)).

The space of potentially useful materials yet to be discovered — the so-called *'chemical white space'* — is immense. The possible combinations of, say, up to six different elements, constitute many billions. The space is further extended by possibilities of different phases, low-dimensional systems, nanostructuring, and so forth, which adds several orders of magnitude. This space was traditionally explored by experimental techniques, *i.e.*, materials synthesis and subsequent experimental characterization. Parsing and searching the full space of possibilities this way is however hardly practical. Recent advances in condensed matter theory and materials modeling make it possible to generate reliable materials data by means of computer simulations based

on quantum mechanics (Lejaeghere et al (2016)). High-throughput simulations combined with machine learning can speed up progress significantly and also help to break out of local optima in composition space to reveal unexpected solutions and new chemistries (Gaultois et al (2016)).

This development has led to a global effort — known as the Materials Genome Initiative (`https://www.mgi.gov/`) — to assemble and curate databases that combine experimentally known and computationally predicted materials properties. A central idea is that materials design challenges can be addressed by searching these databases for entries with desired combinations of properties. Nevertheless, these data sources also open up for *materials informatics*, i.e., the use of big data methodology and data mining techniques to discover new physics from the data itself. A workflow for such a discovery process can be based on a typical data ming process, where key factors are identified, reduced and extracted from heterogeneous databases, similar materials are identified by modeling and relationship mining and properties are predicted through evaluation and understanding of the results from the data mining techniques (Agrawal and Alok (2016)). The use of the data in such a workflow requires addressing problems with data integration, provenance, and semantics, which remains an active field of research.

Even when a new material has been invented and synthesized in a lab, much work remains before it can be deployed. Production methods allowing manufacturing the material at large scale in a cost effective manner need to be developed, and integration of the material into the production must be realized. Furthermore, life-cycle aspects of the material need to be assessed. Today, this post-invention process takes typically about two decades (Mulholland and Paradiso (2016); Jain et al (2013)). Shortening this time is in itself an important strategic goal, which could be realized with the help of an integrated informatics approach (Jain et al (2013), Materials Genome Initiative `https://www.mgi.gov/`).

To summarize, it is clear that materials data, experimental as well as simulated, has the potential to speed up progress significantly in many steps in the chain starting with materials discovery, all the way to marketable product. However, the data needs to be suitably organized and easily accessible, which in practice is highly nontrivial to achieve. It will require a multidisciplinary effort and the various conventions and norms in use need to be integrated. Materials data is highly heterogeneous and much of it is currently hidden behind corporate walls (Mulholland and Paradiso (2016)).

Big Data Challenges

To implement the data-driven materials design workflow, we need to deal with several of the big data properties (e.g. Rajan (2015)).

Volume refers to the quantity of the generated and stored data. The size of the data determines the value and potential insight. Although the experimental materials science does not generate huge amounts of data, computer simulations with accuracy comparable to

experiments can. Moreover, going from state-of-the-art static simulations at temperature T=0K towards realistic descriptions of materials properties at temperatures of operation in devices and tools will raise these amounts as well.

Variety refers to the type and nature of the data. The materials databases are heterogeneous in different ways. They store different kinds of data and in different formats. Some databases contain information about materials crystal structure, some about their thermochemistry, others about mechanical properties. Moreover, different properties may have the same names, while the same information may be represented differently in different databases.

Velocity refers to the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. In computational materials science new data is generated continuously, by a large number of groups all other the world. In principle, one can store summary results and data streams from a specific run as long as one needs (days, weeks, years) and analyze it afterwards. However, to store all the data indefinitely may be a challenge. Some data needs to be removed as the storage capacity is limited.

Variability deals with the consistency of the data. Inconsistency of the data set can hamper processes to handle and manage it. This can occur for single databases as well as data that was integrated from different sources.

Veracity deals with the quality of the data. This can vary greatly, affecting accurate analysis. The data generated within materials science may contain errors, and it is often noisy. The quality of the data is different in different databases. It may be challenging to have provenance information from which one can derive the data quality. Not all the computed data is confirmed by lab experiments. Some data is generated by machine learning and data mining algorithms.

Sources of data and Semantic Technologies

Although the majority of materials data that has been produced by measurement or through predictive computation have not yet become organized in general easy-to-use databases, several sizable databases and repositories do exist. However, as they are heterogeneous in nature, semantic technologies are important for the selection and integration of the data to be used in the materials design workflow. This is particularly important to deal with Variety, Variability and Veracity.

Within this field the use of semantic technologies is in its infancy with the development of ontologies and standards. Ontologies aim to define the basic terms and relations of a domain of interest, as well as the rules for combining these terms and relations. They standardize terminology in a domain and are a basis for semantically enriching data, integration of data from different databases (Variety), and reasoning over the data (Variability and Veracity). According to Zhang et al (2015a) in the materials domain ontologies have been used to organize materials knowledge in a formal language, as a global conceptualization for materials information integration (e.g. Cheng et al (2014)), for linked materials data publishing, for inference support for discovering new ma-

terials and for semantic query support (e.g., Zhang et al (2015b, 2017)).

Further, standards for exporting data from databases and between tools are being developed. These standards provide a way to exchange data between databases and tools, even if the internal representations of the data in the databases and tools are different. They are a prerequisite for efficient materials data infrastructures that allow for the discovery of new materials (Austin (2016)). In several cases the standards formalize the description of materials knowledge (and thereby create ontological knowledge).

In the remainder of this section a brief overview of databases, ontologies and standards in the field is given.

Databases

The Inorganic Crystal Structure Database (ICSD, `https://icsd.fiz-karlsruhe.de/`) is a frequently utilized database for completely identified inorganic crystal structures, with nearly 200k entries (Belsky et al (2002); Bergerhoff et al (1983)). The data contained in ICSD serve as an important starting point in many electronic structure calculations. Several other crystallographic information resources are also available (Glasser (2016)). A popular open access resource is the Crystallography Open Database (COD, `http://www.crystallography.net/cod/`) with nearly 400k entries (Grazulis et al (2012)).

At the International Centre for Diffraction Data (ICDD, `http://www.icdd.com/`) a number of databases for phase identification are hosted. These databases have been in use by experimentalists for a long time.

Springer Materials (`http://materials.springer.com/`) contains among many other data sources the well-known Landolt Bornstein database, an extensive data collection from many areas of physical sciences and engineering. Similarly, The Japan National Institute of Material Science (NIMS) Materials Database MatNavi (`http://mits.nims.go.jp/index_en.html`) contains a wide collection of mostly experimental but also some computational electronic structure data.

Thermodynamical data, necessary for computing phase diagrams with the CALPHAD method, exist in many different databases (Campbell et al (2014)). Open access databases with relevant data can be found through OpenCalphad (`http://www.opencalphad.com/databases.html`).

Databases of results from electron structure calculations have existed in some form for several decades. In 1978, Moruzzi, Janak, and Williams published a book with computed electronic properties such as, e.g., density of states, bulk modulus and cohesive energy of all metals (Moruzzi et al (2013)). Only recently however, the use of such databases have become widespread, and some of these databases have grown to a substantial size.

Among the more recent efforts to collect materials properties obtained from electronic structure calculations publicly available a few prominent examples include the Electronic Structure Project (ESP) (`http://materialsgenome.se`)

with ca 60k electronic structure results, Aflow (Curtarolo et al (2012), http://aflowlib.org/) with data on over 1.7 million compounds, the Materials Project with data on nearly 70k inorganic compounds (Jain et al (2013), https://materialsproject.org/), the Open Quantum Materials Database (OQMD, http://oqmd.org/), with over 470k entries (Saal et al (2013)), and the NOMAD repository with 44 million electronic structure calculations (https://repository.nomad-coe.eu/). Also available is the Predicted Crystallography Open Database (PCOD, http://www.crystallography.net/pcod/) with over 1 million predicted crystal structures, which is a project closely related to COD.

As the amount of computed data grows, the need for informatics infrastructure also increases. Many of the databases discussed above have made their frameworks available, well-known examples include the ones by Materials Project and OQMD. Other publicly available frameworks used in publications for materials design and informatics include the Automated Interactive Infrastructure and Database for Computational Science (AiiDA, http://www.aiida.net/) (Pizzi et al (2016)), the Atomic Simulation Environment (ASE, https://wiki.fysik.dtu.dk/ase/) (Larsen et al (2017)), and the high-throughput toolkit (httk, http://www.httk.org) (Faber et al (2016)).

Ontologies

We introduce the features of current materials ontologies from a materials (Table 1) and a knowledge representation perspective (Table 2), respectively.

Most ontologies focus on specific sub-domains of the materials field (Domain in Table 1) and have been developed with a specific use in mind (Application Scenario in Table 1). The Materials Ontology in Ashino (2010) was designed for data exchange among thermal property databases. Other ontologies were built to enable knowledge-guided materials design or new materials discovery, such as PREMAP ontology (Bhat et al (2013)) for steel mill products, MatOnto ontology (Cheung et al (2008)) for oxygen ion conducting materials in the fuel cell domain, and SLACKS ontology (Premkumar et al (2014)) that integrates relevant product life cycle domains which consist of engineering analysis and design, materials selection and manufacturing. The FreeClassOWL ontology (Radinger et al (2013)) is designed for the construction and building materials domain and supports semantic search for construction materials. MMOY ontology (Zhang et al (2016)) captures metal materials knowledge from Yago. The ontology design pattern in Vardeman et al (2017) models and allows for reasoning about material transformations in the carbon dioxide and sodium acetate productions by combining baking soda and vinegar. Some ontologies are generated (Data Source in Table 1) by extracting knowledge from other data resources such as the Plinius ontology (van der Vet et al (1994)) which is extracted from 300 publication abstracts in the domain of ceramic

materials, and MatOWL (Zhang et al (2009)) which is extracted from MatML schema data to enable ontology-based data access. The ontologies may also use other ontologies as a basis such as for instance, MatOnto that uses DOLCE (Gangemi et al (2002)) and EXPO (Soldatova and King (2006)).

From the knowledge representation perspective (Table 2), the basic terms defined in materials ontologies involve materials, properties, performance, and processing in specific sub-domains. The number of concepts ranges from a few to several thousands. There are relatively few relationships and most ontologies have instances. Almost all ontologies use OWL as a representation language. In terms of organization of materials ontologies, Ashino's Materials Ontology, MatOnto, and PREMΛP ontology are developed as several ontology components that are integrated in one ontology. In Table 2 this is denoted in the modularity column.

Standards

There are currently not so many standards yet in this domain. Early efforts including ISO standards and MatML achieved limited adoption according to Austin (2016). The standard ISO 10303-45 includes an information model for materials properties. It provides schemas for material properties, chemical compositions and measure values (Swindells (2009)). ISO 10303-235 includes an information model for product design and verification. MatML (Kaufman and Begley (2003), `https://www.matml.org/`) is an XML-based markup language for materials property data which includes schemas for such things as materials properties, composition, heat, and production.

Some other standards that have received more attention are, e.g., ThermoML and CML. ThermoML (Frenkel et al (2006, 2011)) is an XML-based markup language for exchange of thermophysical and thermochemical property data. It covers over 120 properties regarding thermodynamic and transport property data for pure compounds, multicomponent mixtures, and chemical reactions. CML or Chemical Markup Language (Murray-Rust and Rzepa (2011); Murray-Rust et al (2011)) covers chemistry and especially molecules, reactions, solid-state, computation and spectroscopy. It is an extensible language that allows for the creation of sub-domains through the convention construct. Further, the dictionaries construct allows for connecting CML elements to dictionaries (or ontologies). This was inspired by the approach of the Crystallographic Information Framework or CIF (Bernstein et al (2016), `http://www.iucr.org/resources/cif`).

The European Committee for Standardization (CEN) organized workshops on standards for materials engineering data (Austin (2016)) of which the results are documented in CEN (2010). The work focuses specifically on ambient temperature tensile testing and developed schemas as well as an ontology (the ELSSI-EMD ontology from above).

Another recent approach is connected to the European Centre of Excellence NOMAD (Ghiringhelli et al (2016)). The NOMAD repository's (`https://repository.nomad-coe.eu/`) metadata structure is formatted to be

Table 1 Comparison of materials ontologies from a materials perspective

Materials ontology	Data Source	Domain	Application Scenario
Ashino's Materials Ontology Ashino (2010)	Thermal property databases	Thermal properties	Data exchange, search
Plinius ontology van der Vet et al (1994)	Publication abstracts	Ceramics	Knowledge extraction
MatOnto Cheung et al (2008)	DOLCE ontology[1], EXPO ontology[2]	Crystals	New materials discovery
PREMΛP ontology Bhat et al (2013)	PREMΛP platform	Materials	Knowledge-guided design
FreeClassOWL Radinger et al (2013)	Eurobau data[3], GoodRelations ontology[4]	Construction and building materials	Semantic query support
MatOWL Zhang et al (2009)	MatML schema data	Materials	Semantic query support
MMOY Zhang et al (2016)	Yago data	Metals	Knowledge extraction
ELSSI-EMD ontology CEN (2010)	Materials testing data from ISO standards	Materials testing, Ambient temperature tensile testing	Data interoperability
SLACKS ontology Premkumar et al (2014)	Ashino's Materials Ontology, MatOnto	Laminated composites	Knowledge-guided design

[1] DOLCE stands for Descriptive Ontology for Linguistic and Cognitive Engineering.
[2] EXPO ontology is used to describe scientific experiments.
[3] Eurobau.com compiles construction materials data from ten European countries.
[4] GoodRelations ontology (Hepp (2008)) is used for e-commerce with concepts such as business entities and prices.

Table 2 Comparison of materials ontologies from a knowledge representation perspective

Materials ontology	Ontology Metrics	Language	Modularity
Ashino's Materials Ontology Ashino (2010)	606 concepts, 31 relationships, 488 instances	OWL	✓
Plinius ontology van der Vet et al (1994)	17 concepts, 4 relationships, 119 instances[1]	Ontolingua code	
MatOnto Cheung et al (2008)	78 concepts, 10 relationships, 24 instances	OWL	✓
PREMAP ontology Bhat et al (2013)	62 concepts	UML	✓
FreeClassOWL Radinger et al (2013)	5714 concepts, 225 relationships 1469 instances	OWL	
MatOWL Zhang et al (2009)	(not available)	OWL	
MMOY Zhang et al (2016)	544 metal concepts, 1781 related concepts, 9 relationships, 318 metal instances 1420 related instances	OWL	
ELSSI-EMD ontology CEN (2010)	35 concepts, 37 relationships, 33 instances	OWL	✓
SLACKS ontology Premkumar et al (2014)	34 concepts and 10 relationships at least[2]	OWL	

[1] 103 instances out of 119 are elements in the periodic system.
[2] The numbers are based on the high-level class diagram and an illustration of instances' integration in SLACKS shown in (Premkumar et al (2014)).

independent of the electronic-structure theory or molecular-simulation code that was used to generate the data and can thus be used as an exchange format.

Conclusion

The use of the materials data in a materials design workflow requires addressing several big data problems including Variety, Variability and Veracity. Semantic technologies are a key factor in tackling some of these problems. Currently, efforts have started in creating materials databases, ontologies and standards. However, much work remains to be done. To make full use of these resources there is a need for integration of different kinds of resources and reasoning capabilities should be used, as in the bioinformatics field in the 1990s (Lambrix et al (2009)). Databases could use ontologies to define their schemas and enable ontology-based querying. Integration of databases is enabled by the use of ontologies. However, when databases have used different ontologies, alignments between different ontologies are needed as well (Euzenat and Shvaiko (2007)). Further, more effort should be put on connecting ontologies and standards (as started in the CML, CEN and NOMAD approaches), which may also lead to connections between different standards. Reasoning can be used in different ways. When developing resources reasoning can help in debugging and completing the resources leading to higher quality resources (Ivanova and Lambrix (2013)). Reasoning can also

be used during querying of databases as well as in the process of connecting different resources.

References

Agrawal A, Alok C (2016) Perspective: materials informatics and big data: realization of the Fourth paradigm of science in materials science. APL Mater 4:053,208:1–10, DOI 10.1063/1.4946894

Ashino T (2010) Materials Ontology: An Infrastructure for Exchanging Materials Information and Knowledge. Data Science Journal 9:54–61, DOI 10.2481/dsj.008-041

Austin T (2016) Towards a digital infrastructure for engineering materials data. Materials Discovery 3:1–12, DOI 10.1016/j.md.2015.12.003

Belsky A, Hellenbrandt M, Karen VL, Luksch P (2002) New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. Acta Crystallographica Section B: Structural Science 58(3):364–369, DOI 10.1107/S0108768102006948

Bergerhoff G, Hundt R, Sievers R, Brown ID (1983) The inorganic crystal structure data base. Journal of Chemical Information and Computer Sciences 23(2):66–69, DOI 10.1021/ci00038a003

Bernstein HJ, Bollinger JC, Brown ID, Grazulis S, Hester JR, McMahon B, Spadaccini N, Westbrook JD, Westrip SP (2016) Specification of the crystallographic information file format, version 2.0. J Appl Cryst 49:277–284, DOI 10.1107/S1600576715021871

Bhat M, Shah S, Das P, Reddy S (2013) Premλp: knowledge driven design of materials and engineering process. In: ICoRD'13, Springer, pp 1315–1329, DOI 10.1007/978-81-322-1050-4˙105

Campbell CE, Kattner UR, Liu ZK (2014) File and data repositories for Next Generation CALPHAD. Scripta Materialia 70(Supplement C):7–11, DOI 10.1016/j.scriptamat.2013.06.013

Ceder G, Persson KA (2013) How Supercomputers Will Yield a Golden Age of Materials Science. Scientific American 309

CEN (2010) A guide to the development and use of standards compliant data formats for engineering materials test data European Committee for standardization

Cheng X, Hu C, Li Y (2014) A semantic-driven knowledge representation model for the materials engineering application. Data Science Journal 13:26–44, DOI 10.2481/dsj.13-061/

Cheung K, Drennan J, Hunter J (2008) Towards an Ontology for Data-driven Discovery of New Materials. In: McGuinness D, Fox P, Brodaric B (eds) Semantic Scientific Knowledge Integration AAAI/SSS Workshop, pp 9–14

Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor R, Nelson L, Hart G, Sanvito S, Buongiorno-Nardelli M, Mingo N, Levy O (2012) AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. Computational Materials Science 58(Supplement C):227–235, DOI 10.1016/j.commatsci.2012.02.002

Curtarolo S, Hart G, Buongiorno-Nardelli M, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nature Materials 12(3):191, DOI 10.1038/nmat3568

Euzenat J, Shvaiko P (2007) Ontology Matching. Springer

Faber F, Lindmaa A, von Lilienfeld A, Armiento R (2016) Machine Learning Energies of 2 Million Elpasolite $(AB–C˝˙–2˝–D˝˙–6˝)$ Crystals. Physical Review Letters 117(13):135,502, DOI 10.1103/PhysRevLett.117.135502

Frenkel M, Chiroco RD, Diky V, Dong Q, Marsh KN, Dymond JH, Wakeham WA, Stein SE, Knigsberger E, Goodwin ARH (2006) XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML) (IUPAC Recommendations 2006). Pure Appl Chem 78:541612, DOI 10.1351/pac200678030541

Frenkel M, Chirico RD, Diky V, Brown PL, Dymond JH, Goldberg RN, Goodwin ARH, Heerklotz H, Knigsberger E, Ladbury JE, Marsh KN, Remeta DP, Stein SE, Wakeham WA, Williams PA (2011) Extension of ThermoML: The IUPAC standard for thermodynamic data communi-

cations (IUPAC Recommendations 2011). Pure Appl Chem 83:19371969, DOI 10.1351/PAC-REC-11-05-01

Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L (2002) Sweetening ontologies with dolce. Knowledge engineering and knowledge management: Ontologies and the semantic Web pp 223–233, DOI 10.1007/3-540-45810-7˙18

Gaultois MW, Oliynyk AO, Mar A, Sparks TD, Mulholland GJ, Meredig B (2016) Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. APL Materials 4(5):053,213, DOI 10.1063/1.4952607

Ghiringhelli LM, Carbogno C, Levchenko S, Mohamed F, Huhs G, Lueders M, Oliveira M, Scheffler M (2016) Towards a Common Format for Computational Materials Science Data. PSI-K Scientific Highlights July

Glasser L (2016) Crystallographic Information Resources. Journal of Chemical Education 93(3):542–549, DOI 10.1021/acs.jchemed.5b00253

Grazulis S, Dazkevic A, Merkys A, Chateigner D, Lutterotti L, Quiros M, Serebryanaya NR, Moeck P, Downs RT, Le Bail A (2012) Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Research 40(Database issue):D420–D427, DOI 10.1093/nar/gkr900

Hepp M (2008) Goodrelations: An ontology for describing products and services offers on the web. Knowledge Engineering: Practice and Patterns pp 329–346, DOI 10.1007/978-3-540-87696-0˙29

Ivanova V, Lambrix P (2013) A unified approach for debugging is-a structure and mappings in networked taxonomies. J Biomed Semant 4:10:1–10:19, DOI 10.1186/2041-1480-4-10

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials 1(1):011,002, DOI 10.1063/1.4812323

Kaufman JG, Begley EF (2003) MatML: A Data Interchange Markup Language. Advanced Materials And Processes 161:35–36

Lambrix P, Stromback L, Tan H (2009) Information Integration in Bioinformatics with Ontologies and Standards. In: Bry F, Maluszynski J (eds) Semantic Techniques for the Web, pp 343–376, DOI 10.1007/978-3-642-04581-3˙8

Larsen AH, Mortensen JJ, Blomqvist J, Castelli IE, Christensen R, Marcin Duak, Friis J, Groves MN, Hammer B, Hargus C, Hermes ED, Jennings PC, Jensen PB, Kermode J, Kitchin JR, Kolsbjerg EL, Kubal J, Kristen Kaasbjerg, Lysgaard S, Maronsson JB, Maxson T, Olsen T, Pastewka L, Andrew Peterson, Rostgaard C, Schitz J, Schtt O, Strange M, Thygesen KS, Tejs Vegge, Vilhelmsen L, Walter M, Zeng Z, Jacobsen KW (2017) The atomic simulation environment - a Python library for working with atoms. Journal of Physics: Condensed Matter 29(27):273,002, DOI 10.1088/1361-648X/aa680e

Lejaeghere K, Bihlmayer G, Bjrkman T, Blaha P, Blgel S, Blum V, Caliste D, Castelli IE, Clark SJ, Corso AD, Gironcoli Sd, Deutsch T, Dewhurst JK, Marco ID, Draxl C, Duak M, Eriksson O, Flores-Livas JA, Garrity KF, Genovese L, Giannozzi P, Giantomassi M, Goedecker S, Gonze X, Grns O, Gross EKU, Gulans A, Gygi F, Hamann DR, Hasnip PJ, Holzwarth NaW, Iuan D, Jochym DB, Jollet F, Jones D, Kresse G, Koepernik K, Kkbenli E, Kvashnin YO, Locht ILM, Lubeck S, Marsman M, Marzari N, Nitzsche U, Nordstrm L, Ozaki T, Paulatto L, Pickard CJ, Poelmans W, Probert MIJ, Refson K, Richter M, Rignanese GM, Saha S, Scheffler M, Schlipf M, Schwarz K, Sharma S, Tavazza F, Thunstrm P, Tkatchenko A, Torrent M, Vanderbilt D, van Setten MJ, Speybroeck VV, Wills JM, Yates JR, Zhang GX, Cottenier S (2016) Reproducibility in density functional theory calculations of solids. Science 351(6280):aad3000, DOI 10.1126/science.aad3000

Moruzzi VL, Janak JF, Williams ARAR (2013) Calculated electronic properties of metals. Pergamon Press, New York

Mulholland GJ, Paradiso SP (2016) Perspective: Materials informatics across the product lifecycle: Selection, manufacturing, and certification. APL Materials 4(5):053,207, DOI 10.1063/1.4945422

Murray-Rust P, Rzepa HS (2011) CML: Evolution and design. Journal of Chem-

informatics 3:44:1–44:15, DOI 10.1186/1758-2946-3-44

Murray-Rust P, Townsend JA, Adams SE, Phadungsukanan W, Thomas J (2011) The semantics of Chemical Markup Language (CML): dictionaries and conventions. Journal of Cheminformatics 3:43, DOI 10.1186/1758-2946-3-43

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) AiiDA: automated interactive infrastructure and database for computational science. Computational Materials Science 111(Supplement C):218–230, DOI 10.1016/j.commatsci.2015.09.013

Premkumar V, Krishnamurty S, Wileden JC, Grosse IR (2014) A semantic knowledge management system for laminated composites. Advanced engineering informatics 28(1):91–101, DOI 10.1016/j.aei.2013.12.004

Radinger A, Rodriguez-Castro B, Stolz A, Hepp M (2013) Baudataweb: the austrian building and construction materials market as linked data. In: Proceedings of the 9th International Conference on Semantic Systems, ACM, pp 25–32, DOI 10.1145/2506182.2506186

Rajan K (2015) Materials Informatics: The Materials Gene and Big Data. Annu Rev Mater Res 45:153–169, DOI 10.1146/annurev-matsci-070214-021132

Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C (2013) Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). JOM 65(11):1501–1509, DOI 10.1007/s11837-013-0755-4

Soldatova LN, King RD (2006) An ontology of scientific experiments. J R Soc Interface 3(11):795–803, DOI 10.1098/rsif.2006.0134

Swindells N (2009) The representation and exchange of material and other engineering properties. Data Science Journal 8:190–200, DOI 10.2481/dsj.008-007

van der Vet P, Speel PH, Mars N (1994) The Plinius ontology of ceramic materials. In: Mars N (ed) Workshop Notes ECAI'94 Workshop Comparison of Implemented Ontologies, pp 187–205

Vardeman C, Krisnadhi A, Cheatham M, Janowicz K, Ferguson H, Hitzler P, Buccellato A (2017) An Ontology Design Pattern and Its Use Case for Modeling Material Transformation. Semantic Web 8:719–731, DOI 10.3233/SW-160231

Zhang X, Hu C, Li H (2009) Semantic query on materials data based on mapping matml to an owl ontology. Data Science Journal 8:1–17, DOI 10.2481/dsj.8.1

Zhang X, Zhao C, Wang X (2015a) A survey on knowledge representation in materials science and engineering: An ontological perspective. Computers in Industry 73:8–22, DOI 10.1016/j.compind.2015.07.005

Zhang X, Pan D, Zhao C, Li K (2016) MMOY: Towards deriving a metallic materials ontology from Yago. Advanced Engineering Informatics 30:687–702, DOI 10.1016/j.aei.2016.09.002

Zhang X, Chen H, Ruan Y, Pan D, Zhao C (2017) MATVIZ: a semantic query and visualization approach for metallic materials data. International Journal of Web Information Systems 13:260–280, DOI 10.1108/IJWIS-11-2016-0065

Zhang Y, Luo X, Zhao Y, chao Zhang H (2015b) An ontology-based knowledge framework for engineering material selection. Advanced Engineering Informatics 29:9851000, DOI 10.1016/j.aei.2015.09.002