

ISSUES

Collaboration between Industry and Academia Laying Groundwork for Data-Driven Materials Development

Advances in Materials Informatics and Potential for Advanced Measurement Technologies

Amid the quickening pace of development of new materials that will help to enhance infrastructure and to resolve environmental and energy problems in the move to a sustainable society, there are rising expectations for the use of materials informatics as a means of rapid and efficient development of high functional materials. Materials informatics is a field that combines materials and computational science, involving the recursive identification of rules and interrelationships between the structure and properties of materials from systematically acquired data and turning them into new material design principles. Achieving this requires a variety of different resources, including digital technologies (a platform on which to collect data about materials, data analysis, machine learning, artificial intelligence, and so on), solutions that can put these technologies to use in practical materials development, and the advanced measurement technologies needed for analysis. Hideki Yoshikawa, who is responsible for establishing a materials data platform for use in materials informatics at the National Institute for Materials Science, and Toru Hara, a recognized expert in measurement, talk to key people involved in the materials informatics at Hitachi about current trends in the field, the potential for advanced measurement technologies, and the future for the materials developments that rely on these technologies.

Approaching MI from Three Different Angles

Tanimoto: Materials research and development (R&D) has a long history of contributing to human progress through advances in materials that have gone hand in hand with the growth of industry. Materials informatics (MI) has attracted attention in recent years as a way of facilitating the R&D of materials and products by using the analysis of accumulated data to rapidly identify the interrelationships between the structure of materials and their properties and functions. Yoshikawa-san,

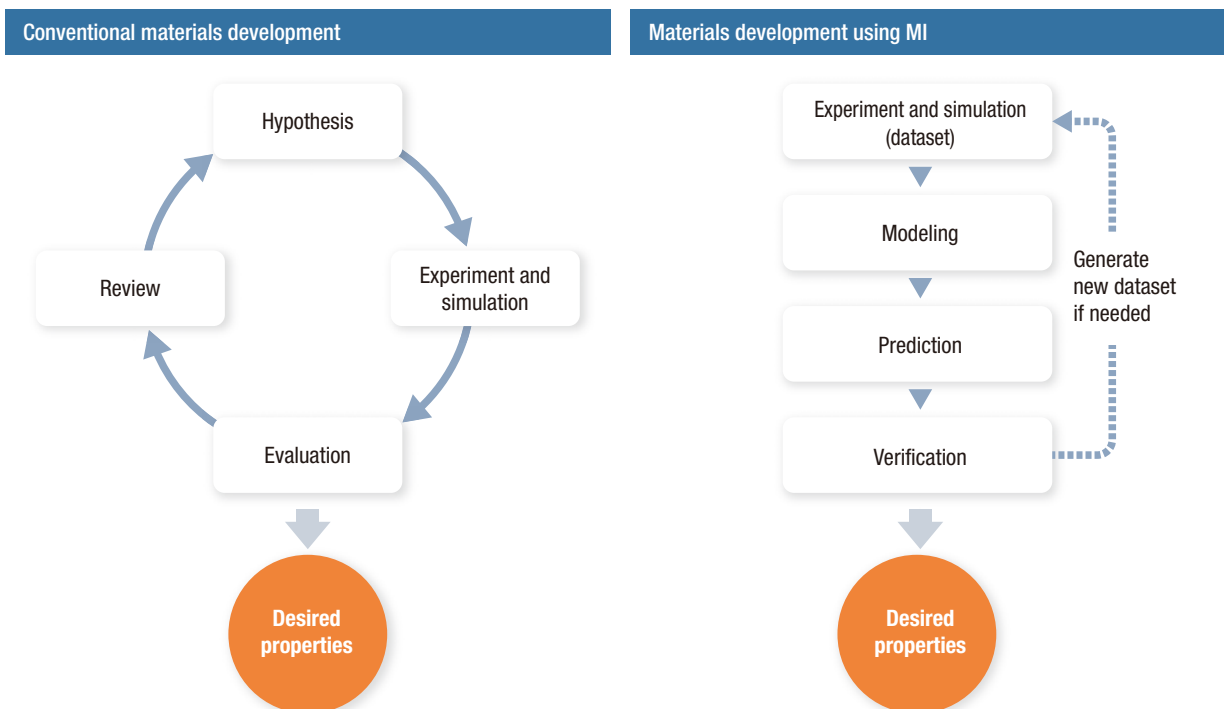


Hideki Yoshikawa
Deputy Managing Director,
Materials Data Platform Center,
Research and Services Division
of Materials Data and Integrated
System, National Institute for
Materials Science



Toru Hara
Platform Director, Structural
Materials Analysis Platform,
Research Center for Structural
Materials, National Institute for
Materials Science

Materials Developed Using Materials Informatics (MI)



you work for the National Institute for Materials Science (NIMS), which is leading work on MI in Japan. Can you please explain to us what is behind the emergence of this new trend?

Yoshikawa: MI has long been seen as a way to speed up materials R&D while reducing its cost. This has reached the point where MI is now starting to be used in earnest, with data-driven R&D coming to fruition thanks to rapid advances in information science over recent years that allow us to process and utilize large amounts of data. The launch by President Obama in 2011 of the Materials Genome Initiative (MGI) in the USA is another factor behind work on MI in Japan.

Tanimoto: What is NIMS doing with regard to MI?

Yoshikawa: Our approach is based around groups working on MI from three different angles: new materials development, performance improvement of practical materials, and data platform development. We established the Research and Services Division of Materials Data and Integrated System (MaDIS) to find new ways to research and develop properties and materials through data

integration, with the three groups forming part of this organization.

The new materials development group is working on the Materials Research by Information Integration Initiative (Mi²i), which seeks to use data science and the techniques of combinatorial chemistry for the high-throughput synthesis of materials, using this as a way to rapidly develop new materials with properties and features that have never previously existed. This work is primarily targeting battery materials, magnetic materials, and thermal management materials.

The SIP-MI Laboratory is working on structural materials such as steel, with the aim of boosting the efficiency and cutting the cost of development to improve the productivity and performance of practical materials that are already in widespread use in society. As part of the Structural Materials for Innovation Program, which is in turn part of the Cabinet Office's Strategic Innovation Promotion Program (SIP), the laboratory develops technologies for materials integration, focusing on structural materials.



Hidekazu Morita

Senior Engineer, HPC Solution Center for Governmental & Public Information Systems, Public Platform Solution Operation, Government & Public Corporation Information Systems Division, Social Infrastructure Systems Business Unit, Hitachi, Ltd.



Ryuichiro Tamochi

General Manager, Science Systems Sales & Marketing Division, Science & Medical Systems Group, Hitachi High-Technologies Corporation



Tomio Iwasaki

Chief Researcher, Biochemical Materials Research Department, Center for Technology Innovation – Materials, Research & Development Group, Hitachi, Ltd.



Sayaka Tanimoto

Chief Researcher, Center for Technology Innovation – Electronics, Research & Development Group, Hitachi, Ltd.

We have also established the Materials Data Platform Center (DPFC) to provide a platform for this data-driven R&D. As a national center for materials information, the DPFC operates both within NIMS and further afield with activities that range from collecting data for materials research and tools for its use to providing services.

Applying MI Discoveries to Device Materials

Tanimoto: Use of MI is also growing in industry. Iwasaki-san, as someone who plays a leading role in the use of MI at Hitachi, can you provide some specific examples?

Iwasaki: MI caught our attention at a comparatively early stage, primarily with regard to electronic materials. Our initial focus was on the problem of interfacing surfaces, something that plays an important role in the reliability of electronic devices. We were looking for materials to meet two criteria in particular: higher adhesive strength and reduced diffusion across grain boundaries.

In the case of adhesive strength, we were looking for materials that demonstrate strong adhesion when joining plastic to metal, metal to ceramic, and ceramic to plastic. Of all the parameters used in simulation, the prediction based on existing

knowledge was that surface tension represented the best way to achieve adhesion between surfaces. What the actual analysis found, however, was that the configuration of atoms has a large impact. It was MI that provided us with this new insight. From this we concluded that adhesion is improved if the atoms in the two different materials are arranged with similar periodicity (as indicated by their lattice constant), and this knowledge was utilized in actual device materials.

Diffusion across grain boundaries, meanwhile, is an important factor in wiring breaks and in fact has been a topic of study for some time, going back to the 1990s. When comprehensive analyses, including virtual material simulation, were undertaken to look for alloying elements that could reduce diffusion in the aluminum and copper used for wiring, thereby minimizing wiring breaks, it was found that atomic radius and cohesive energy are significant parameters. Materials developed based on this finding have been put to use in semiconductors and the finding itself has been presented at academic conferences. These two parameters are also important for choosing alloying elements to reduce the extent of elongation, a type of defect in lead-free solder, and this knowledge has led to the discovery of elements that work well for this purpose.

Rising Interest in MI by Industry

Tanimoto: I imagine numerous other manufacturers have taken an interest in MI as well as Hitachi. Morita-san, can you tell us about how MI is used in industry?

Morita: I work on materials development solutions that put MI to practical use in business to overcome the challenges faced by customers. Having spoken to more than 100 companies about MI, I am aware that they have a strong interest in the topic. While some are taking a wait-and-see attitude, those companies that are actively pursuing the technology are already putting it to use. We conducted a proof-of-concept for our first customer in October 2015.

In terms of solutions, we enter into confidentiality agreements with customers and carefully protect data from the fields that are important to them. All of the results of data analysis are supplied back to the customer, with most of the analysis techniques used being our own intellectual property. We are seeking to expand collaborative creation with customers by utilizing the data analysis techniques and know-how built up through this work across different applications, and by doing so to resolve a wide variety of the challenges facing customers and society as a whole.

Along with the development of new materials, applications for MI also include reducing the cost of existing materials and making experimental work more efficient. The last six months or so have also seen a rise in demand from customers wanting to be able to utilize artificial intelligence (AI) analytics by feeding the experimental data output by instruments directly into machine learning models.

Hara: What types of data are being collected?

Morita: As collecting all data is unrealistic, we have people focus on collecting those data that are easiest to use in AI and machine learning. As data on failures also play an important role in improving analysis accuracy, we urge customers who are able to do so to collect this data also.

Hara: How do you ensure the reliability of your data? I expect there will also be cases when the number of experiments will be too small because of the large number of factors to be considered.

Morita: This happens often. This is why we have adopted the verification and validation approach, conducting the actual experiments within a scheme that includes AI and machine learning and including data reliability in the verification process.

Data Interpretation a Challenge for Advanced Measurement Technologies

Tanimoto: You mentioned increasing demand from industry for feeding data from instruments into machine learning models. Tamochi-san, can you tell us about the challenges that arise in doing this?

Tamochi: A challenge recognized both by ourselves as an instrument vendor and by the user community is the difficulty of seeing the interrelationships between the material characteristics and the measurements acquired by instruments. Typically, different types of instruments produce data in different formats, and the current situation is that the measurement results only exist on the computer inside the instrument and the experimental results only in notes, such that the person doing the experiment is the only one aware of all of the relationships. There are significant obstacles to the integrated handling of different forms of measurements (experimental data) and the identification of interrelationships with

material properties, meaning that the challenge for us is to find ways to overcome these obstacles. **Yoshikawa:** Another challenge that presents itself to me as a user of measuring instruments is the differences in different people's interpretation of measurement results. My own field of surface analysis is one in which the interpretation of data is particularly difficult, with differences in experience and expertise showing up in how data is interpreted. Technicians have an unconscious tendency to selectively interpret results in ways that conform to their expectations, posing a risk that they will fail to see what is actually happening. Unfortunately, overcoming this is not just a simple matter of utilizing raw (uninterpreted) data in MI. That is, because raw data is only understood by the person who collects, it is difficult to use even in MI. It would be helpful, however, if instruments were equipped with functions for automatically imposing a degree of interpretation on data, if only in the nature of primary filtering.

Hara: It may be that electron microscope images are a particularly difficult type of experimental data. In the past, the images were distinctive enough that you could almost tell who had taken them just by looking. It is still true that the images produced by scanning electron microscopes are completely different depending on the detector used, with microscopes from different vendors producing different images. Observing the same specimen results in a variety of different images. Put another way, because a single image does not tell you everything about the microstructure, it is the purpose (what you are looking for) and the meaning of the data that are important. In future applications of MI, something will also need to be done about how data is acquired and collected. **Tamochi:** That's right. How to quantify the meaning of data acquired from an electron microscope poses a difficult challenge, and as such I believe

the suppliers of complex measuring instruments are duty bound to put a lot of effort into making their products as easy as possible to use regardless of who uses them.

Yoshikawa: By recording everything together in the same dataset, including raw data, data that has undergone a degree of interpretation, and the purpose for which the data was acquired, it should be possible to make data analysis easier by imposing a degree of universality on experimental data that in the past has been extremely personal.

Iwasaki: With regard to the interpretation of data, what people working on organic materials are hoping for from electron microscopes is a function that can determine the configuration of atoms from an image and present it as coordinates, even if it is only a prediction. If such a function were available, it would make it easier to create simulation models.

Hara: While I'm sure such a function would be helpful, wouldn't it be difficult to achieve with metals? Transmission electron microscopes typically only provide a projection, and the skill of sample preparation is a factor in the image interpretation problem. The obstacles are significant, including choosing what to quantify from an image of material structure that can serve as a feature value, and whether this can tell us the arrangement of atoms.

Yoshikawa: While getting a simulation model to converge on a single solution is difficult, from hundreds or thousands of sets of simulation results, it should be possible to isolate several dozen that explain an experimental result.

Hara: It may be that, by generating lots of structural models to put into the calculation, it is possible to home in on particular candidates.

Yoshikawa: It seems to me that one way to avoid the problem referred to earlier of selective interpretation by technicians is to offer them a

choice from a range of different interpretations. This makes it easier to check interpretations and improve data reliability. I imagine it is the same in other fields, but having the instruments perform a certain amount of interpretation automatically can be a means of providing advice to the technician and can raise their overall skill level. I see this as another area where MI shows promise.

Hara: That is certainly another aspect to consider. However, I am also apprehensive that too much automation might impede the progress of science by taking away the experience of trial and error whereby a hypothesis is proposed but a different result may be obtained, the opportunity to learn from this process, and the serendipitous discoveries that arise from failures.

Yoshikawa: As Hara-san touched on the sort of cutting-edge research that is not amenable to automation is vital for science. Nevertheless, when it comes to professional development for measurement technicians at institutions such as those that offer their facilities to others, we have yet to reach that level and are still at the stage of looking at how to collect routine measurement data. The systematic consolidation of data that in the past has been scattered across different locations represents a major step toward putting in place the conditions for professional development, something that should facilitate the elucidation of complex physical phenomena and the acquisition of new knowledge in cutting-edge research and elsewhere.

Progress on Establishing Platforms for Using Data

Tamochi: The New Energy and Industrial Technology Development Organization (NEDO) launched a project in April 2018 aimed at establishing ways of collating data in this manner. The project brings

vendors of measurement and analysis systems, Hitachi included, together with universities and research institutions to work on objectives that include the development of functions for integrating data across measurement or analysis systems of different types and from different vendors, and multi-purpose measurement and analysis systems that use AI to combine advanced analysis functions.

With regard to data integration, the project has started a process of converting the different data formats used by different measurement and analysis systems into common formats and viewing them on an integrated viewer to look for interrelationships, being currently at the stage of considering which types of data are needed. There are a lot of things to be decided, such as whether data on sample preparation and experimental method are also needed along with the raw data from the instruments, and how these different types of data are to be combined. The project is currently discussing the creation of standardized sample holders to ensure the uniformity of samples, and their labeling with identification numbers.

Yoshikawa: I look forward to seeing what you come up with. We are taking a different approach that seeks to control the terminology used in materials research. Terminology poses a problem if you want to handle data on materials in an integrated manner. In AI and machine learning, for example, the same things can end up being treated as if they are different simply because they are labeled differently. Because terminology grows as new discoveries are made and new technologies developed, it is at the heart of science. When science moves into a new area such as the use of AI, it needs to avoid having its progress blocked by a failure of the terminology to keep pace. As an institution specializing in materials research, I see our role as being to consider ways to reduce

differences in terminology to make it easier to use, and to establish these as an open resource available for widespread use.

Tamochi: Terminology is also a subject of discussion at the NEDO project. The problem of what to do in situations such as when different vendors give different names to the same function is ultimately a very important consideration. Moreover, given that it takes a lot of time for us to discuss the issues one by one, it would help if NIMS could deal with it all together.

Yoshikawa: Certainly, let's work together. With so much terminology to deal with, not just in Japanese, but also in foreign languages, we are currently looking to take advantage of the "wisdom of crowds" in how we address the problem by drawing on the assistance of people in academia and industry, both locally and overseas.

Morita: That is good news for industry. If NIMS is able to consolidate things centrally, we can share that information across the industry and do what we can to make active use of the resource.

Becoming World-leaders through Japan-wide Collaboration

Tanimoto: Thanks to what you have told me, I feel I now have a sense of where MI is at and where it is going, and of the relationship between MI and the sciences in which it is used.

Hara: There is a lot that remains unclear in my own field of metal microstructures and composition. Collecting data and knowledge systematically through a process that includes adding data acquired using new measurement technologies

to that which has been previously collected, and then making this available for use, may result in discoveries made by looking at things from a new direction. I sense that the way forward to achieving this is coming into view.

Yoshikawa: While industry is conscious of the need for materials data platforms, I also get feedback saying the cost is too high for it to be done by individual companies. I believe that a nationwide effort to establish platforms, including those of NIMS, would likely result in enhanced industrial competitiveness.

Morita: That's right. I believe there is potential to save costs in data platform development also through the wider application of model platforms, in the same way as those for data analysis. While data held by companies includes some that is not for public release, it should still be possible to provide Japan as a whole with better access to data if ways can be found to link together those parts that are open for joint use. There is no doubt that the sharing of data between industry and academia and the linking together of the things they are working on will facilitate progress on both sides.

Yoshikawa: I hope we can combine our respective strengths to achieve world leadership through a Japanese style of MI.

Tanimoto: By bringing together capabilities from a range of different fields, including materials research, the associated industries, the measuring instruments that are essential to R&D, and digital technologies such as data analysis and AI, I believe that the materials industries in which Japan has traditionally been strong can continue to prosper in the era of MI. Thank you for your time today.