Development of Materials Using Materials Informatics

Hitachi's Chemicals Informatics Solution for Environmental Management

In recent years, chemical and materials manufacturers have been tasked with the early development of highly functional materials that will help realize a circular economy. On the other hand, since conventional materials development requires trial and error over a long period of time, there has been a focus on materials informatics, which combines materials science with information science. A chemicals informatics solution developed by the Hitachi Group supports the acceleration of strategic research and development by utilizing public data such as patents to narrow down materials with good characteristics that have not been patented yet. This will not only reduce the loss of waste materials by improving the efficiency of trial production and measurement evaluations, but will also contribute to environmental management by promoting the development of materials such as biodegradable plastics that help reduce environmental impacts, lithium-ion battery materials that are essential for decarbonization, catalysts that produce gases such as hydrogen, methane, and ammonia, and natural and bio-derived materials that do not increase CO_2 emissions even when burned.

Rei Kurokawa
Tomio Iwasaki, Ph.D.
Takashi Isobe, Ph.D.
Yoshihiro Okada

1. Introduction

As environmental issues become ever more serious on a planetary scale, manufacturing companies that make the products supporting people's everyday lives and economic activities are expected to conduct sustainable and responsible production activities across the entire supply chain, from the procurement of raw materials to manufacturing and disposal. In particular, there is growing demand for the early development of new, highly functional materials that address environmental issues by the chemical and materials industries that make the materials used in manufacturing. In this context, the current trend is to move away from conventional methods of materials development that relied on manual work by humans toward trying new methods that utilize information science.

2. Evolution of Materials Development Brought about by Materials Informatics

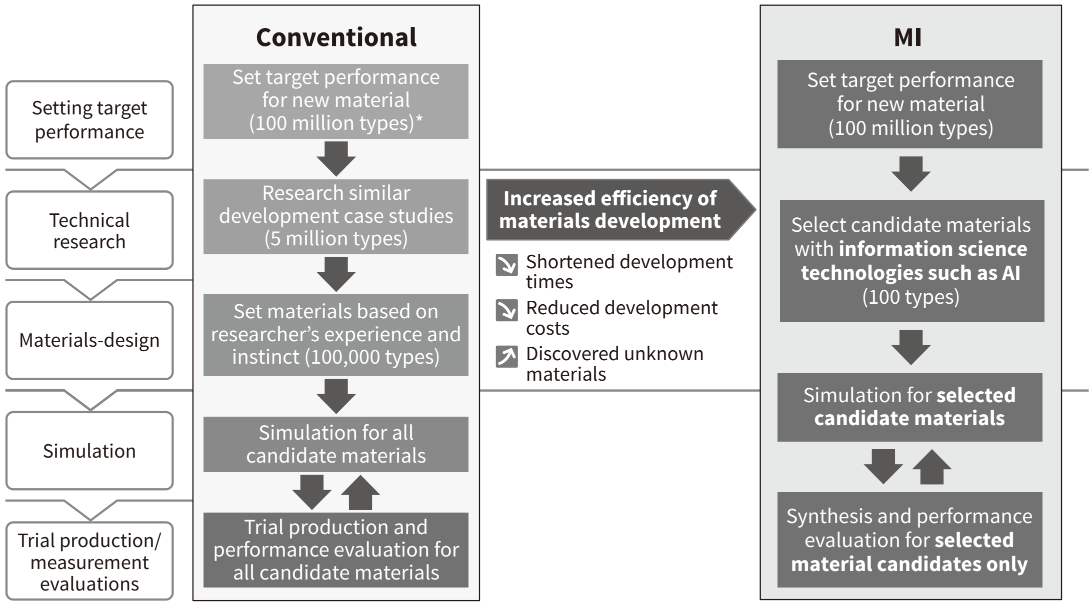
Materials informatics (MI) is an approach to highly efficient materials development based on information science techniques that use methods such as statistical analysis. Compared to conventional methods that are dependent on the experience and instinct of researchers, this method is expected to deliver dramatically shorter development times and also enable the discovery of previously unknown materials[1].

2. 1 MI Transforms the Materials Development Process

Figure 1 shows the transformation of the materials development process brought about by MI. In the conventional materials development process, the target performance of

Figure 1 — Comparison of Conventional Materials Development and Materials Informatics

Improved materials development efficiency is expected by utilizing information science techniques. Created by Hitachi with reference to "The Use of Materials Informatics for Materials Development" published by Sumitomo Mitsui Banking Corporation in November 2019.



MI: materials informatics AI: artificial intelligence

* The values in parentheses indicate examples of the number of material candidates.

the new material is set first, then the material researchers investigate case studies of similar developments, and perform simulations, trial production, and measurement evaluations for all the materials that are designed based on the experience and knowledge of the researchers. When utilizing MI, after setting the target performance, technologies such as artificial intelligence (AI) are used in the technical research and materials-design stages to narrow down the material candidates, and the simulations, trial production, and measurement evaluations are performed only for the narrowed down materials. This shortens development times and reduces costs, while also leading to the potential discovery of unknown materials.

2. 2 Role of MI in Environmental Management

Although the focus on MI has been mainly due to its role in discovering new materials and improving development efficiency, it also has great potential for contributing to environmental management at corporations as they need to respond ever more urgently to environmental and social issues. In recent years, manufacturing companies are expected to use materials that are sustainable throughout the entire lifecycle of a product.

In this context, the chemical and materials industries are utilizing MI to attempt the early development of highly functional materials that support technological innovation. These include the development of new materials to substitute for those that are problematic in terms such as environmental compatibility, biocompatibility, procurement or recycling; battery materials that stimulate energy transformation to realize a sustainable society; and semiconductor materials that stimulate the transformation of information and communication technology, which is a key part of social infrastructure[(2), (3), (4)].

The role of MI in environmental management is not limited to such transformation of materials for development, but has grown dramatically in recent years as activities shift to sustainable and responsible production that eliminates as much waste as possible from the materials development process.

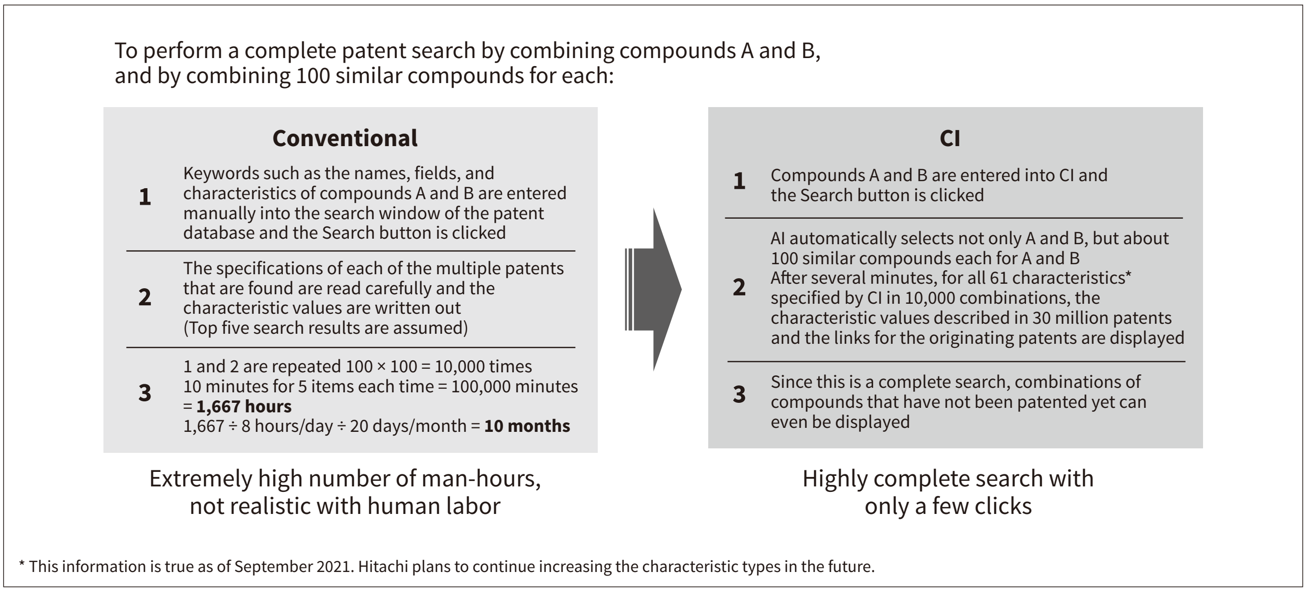
3. Chemicals Informatics Solution for Realizing Strategic Materials Development

Hitachi High-Tech Solutions Corporation held meetings with more than 20 chemical and materials manufacturers, and then developed a chemicals informatics solution (hereinafter "CI") that is a proprietary MI technique for supporting upstream technical research and materials-design in the materials development process.

CI is a cloud service comprising a database constructed after analyzing public technical documents, such as patents and technical papers, using natural language processing technology. Based on the characteristics of this information, such as the material structure and elemental composition, CI can search for and propose materials that were ignored in the past, as well as compounds with new structures formed by AI. The predicted characteristic values and related patents for these materials can also be displayed.

Figure 2 — Example of CI Utilization

The chemicals informatics solution (CI) enables a complete search of patent documents with only a few clicks by displaying candidate compounds and materials expected to have good characteristics from a huge amount of reference data such as patents, together with their predicted characteristic values and related patent links.



3. 1 Utilizing Huge Amounts of Public Data to Transform the Materials Development Process from Upstream

Figure 2 shows an example of CI utilization. A scenario is assumed where a patent search is performed for a combination of two compounds, A and B, which will form the basis of a certain material, and 100 similar compounds for each A and B to improve completeness. The patent database is searched by entering the names of compounds A and B, then the specifications of multiple displayed patents are read, and characteristic values and other data are written out. This work is repeated 100 × 100 = 10,000 times, and if the reading is assumed to take 10 minutes each time, this will require more than 1,600 hours. Since this is unrealistic, in conventional methods, the candidate compounds were preselected based on the instinct and experience of the materials developer before performing the trial production and measurement evaluations. In other words, the selection of compound candidates with good characteristics can be expected if the materials researcher is highly experienced and knowledgeable. If this is not the case, meanwhile, there is a risk of overlooking materials, or needing to wastefully repeat trial production and measurement evaluations due to the selection of materials with characteristics that are not relevant to the purpose or materials that conflict with existing patents.

When CI is used and compounds A and B are entered, CI automatically selects about 100 compounds that are similar to compounds A and B. Then, it performs a complete search for 100 × 100 = 10,000 combinations, and after several minutes, displays the candidate compounds and materials that are expected to have good characteristics together with characteristic value predictions and related patent links. Since it performs a complete search of public patents, CI can even display compounds and materials that have not been patented yet even though they can be expected to have good characteristic values. This makes a complete search of patent documents, which was unrealistic using human labor, possible with only a few clicks.

CI contains a rich data set, and by utilizing it upstream of the materials development process, filtering can be performed after considering a wide range of compound and material candidates, and simulations, trial production, and measurement evaluations are only required for the promising filtered candidates. This results in strategic research and development that increases the overall efficiency of the materials development process, increases the probability of development success, and produces the intended effects at an early stage.

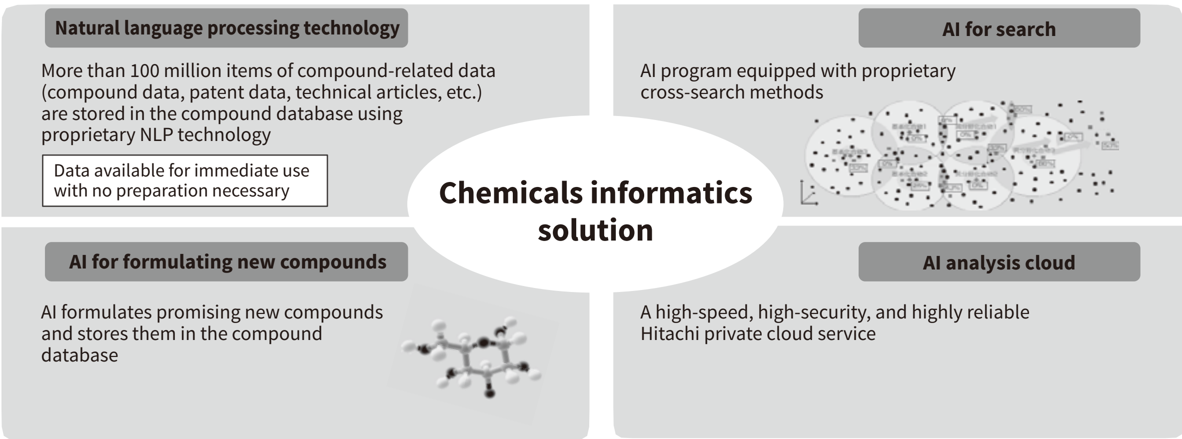
3. 2 CI's Technical Advantages

CI has three technical elements that support innovation (see **Figure 3**). The first is natural language processing technology. CI is comprised of a proprietary database created by using natural language processing technology to extract the characteristic values, fields, and applications of compounds and materials from sources such as 30 million patents and 33 million technical papers from 105 countries and regions that are publicly available on the Internet. The database already contains 117 million recorded compounds, 119 field and application types, 61 characteristic types, and more than 401 million characteristic values.

The second technical element is AI for formulating new compounds. This randomly changes a part of the structure

Figure 3 — CI's Technical Elements for Supporting Innovation

From compound-related data collected by proprietary natural language processing technology and new compounds formulated by AI, an efficient and complete search for candidate compounds and materials that are expected to have good characteristics is performed using a proprietary search AI. This is stabilized and processed on a Hitachi private cloud service.



NLP: natural language processing

of known compounds, and has stored 11 million compounds whose novelty has been confirmed.

The third element is AI for search. CI provides not just a simple keyword search, but a search based on features such as structure and elements, which is more likely to return materials with the target characteristics. Multiple compounds can be specified to enable discovery of compounds that combine the characteristics of multiple compounds, or composite materials that combine compounds.

CI integrates these proprietary technologies as a cloud service, providing an environment available for immediate use by simply connecting to it via a browser.

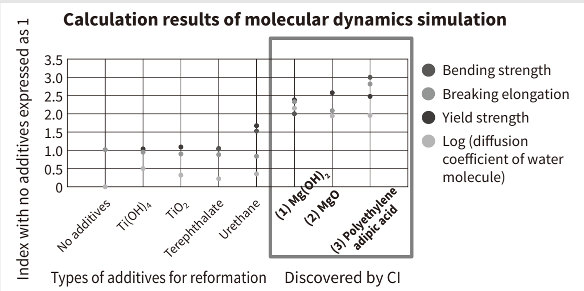
4. Case Study of CI Utilization Contributing to Environmental Management

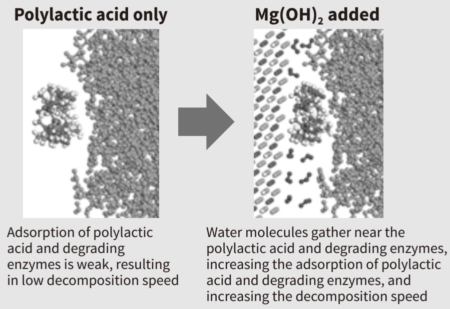
Utilizing CI in processes upstream of the materials development process makes all of the processes more efficient, which not only reduces the loss of waste materials used

Figure 4 — Case Study of Additive Search for Biodegradable Plastics for Reducing Environmental Impact

CI was utilized for an efficient and complete search of new additives that increased the strength of biodegradable plastics and also made plastics easier to decompose in natural environments. Their validity was then confirmed in a molecular dynamics simulation.

Purpose	Search for additives that increase the strength and decomposition speed of biodegradable plastics
Search conditions	Search from the starting points of polylactic acid that is often used as biodegradable plastic, and metal hydroxides, metal oxides, and organic substances that have potential to be used as additives
Results	Three additive candidates were found with high yield strength, high breaking elongation, and high moisture vapor transmission rate, and their validity was confirmed in a molecular dynamics simulation





Effects	• Selection and simulation cycle for candidate compounds shortened to one-sixtieth the time (normally approx. 120 days → 2 days) • Number of trial production runs/measurement evaluations and amount of test material waste reduced to one-fifth the level (normally 10 cases → 2 cases) • Completeness of candidate compound study increased more than 800 times (600 instances → approx. 500,000 instances)

$Ti(OH)_4$: titanium hydroxide TiO_2: titanium dioxide $Mg(OH)_2$: magnesium hydroxide MgO: magnesium oxide in trial production and measurement evaluations, but also accelerates the development of highly functional materials that reduce environmental impact (see **Figure 4**).

4. 1 Improvement of Decomposition Performance of Biodegradable Plastics, and Increased Efficiency of Trial Production and Measurement Evaluations

The disposal of plastic products such as household goods, clothing fibers, and fishing nets causes environmental pollution. In response to this issue, there is increased interest in materials that reduce environmental impact, such as biodegradable plastics and marine-degradable plastics[(5), (6)]. This section describes a case study of CI utilization that resulted in the rapid discovery of new additives that increased the strength of biodegradable plastics and also made plastics easier to decompose in natural environments.

The compounds specified in CI as search conditions were polylactic acid that is often used as biodegradable plastic, 15 types of metal hydroxides, 15 types of metal oxides, and 20 types of organic substances that were already known and had potential to be used as additives. When the CI search was performed, in addition to the specified compounds, about 100 types of similar compounds were selected automatically and included in the search. This enabled simultaneous search for (1 polylactic acid × 100 similar compounds) × [(15 metal hydroxides + 15 metal oxides + 20 organic substances) × 100 similar compounds] = 500,000 combinations.

The results identified three types of additives that could be expected to provide higher strength and decomposition performance than conventional additives; one metal hydroxide type, one metal oxide type, and one organic substance type. Furthermore, a molecular dynamics simulation demonstrated the high strength and decomposition performance of biodegradable plastics that use these additives, thereby confirming their validity. This work only required two days. Then, taking into account a precondition that an organic material must be included due to manufacturing restrictions, trial production and measurement evaluations using the actual materials became necessary for only two cases, which combined one metal hydroxide type + one

Table 1 — Record of CI Results

CI provides comprehensive support for organic, inorganic, single compounds, composite materials, and high and low molecular weights, and has a proven track record for search in a wide range of fields, including electronic components, optical materials, energy materials, and natural and bio-derived materials.

No.	Category	Search record	Target characteristics
1	Inorganic materials	Search for new combinations of oxide LIB solvents	Li conductivity (S/cm)
2	Inorganic materials	Search for new combinations of sulfide LIB solvents	Li conductivity (S/cm)
3	Inorganic materials	Search for new combinations of phosphors/quantum dots with specific wavelengths	PL wavelength (nm), PL quantum yield (%)
4	Inorganic materials	Search for structural factors influencing gas reactivity of semiconductors	Gas reactivity (Å/cycle)
5	Inorganic materials	Search for new combinations of structural factors influencing hardness of alloy coating films	Vickers hardness (HV)
6	Organic materials	Search for new combinations of LIB electrode materials with high discharge capacity	Discharge capacity (mAh/g)
7	Organic materials	Search for synthetic methods starting from catalyst/compound structures	Reaction yield (%), activation energy (eV)
8	Organic materials	Search for new combinations and new structures for low dielectric resin	Relative dielectric constant, dielectric loss tangent, coefficient of thermal expansion (ppm/K)
9	Organic materials	Search for new combinations and structural influence factors for smartphone resin lenses	Refractive index, light transmittance (%)
10	Organic materials	Search for new combinations of photoreactive resin with short wavelengths and low exposure	Relative dielectric constant, dielectric loss tangent, exposure (mJ/cm^2), exposure wavelength (nm)
11	Organic materials	Search for new combinations of water-repellent materials with high contact angles	Contact angle (deg)
12	Organic materials	Search for new combinations of hydrophilic films	Contact angle (deg)
13	Organic materials	Search for new combinations of liquid black inks	Optical density OD value, absorption wavelength (nm)
14	Biochemical	New structure search demonstration for 4-HPPD inhibitor agricultural chemical	IC_{50} (nM, μM, mM), agricultural chemical used amount (kg/ha)
15	Biochemical	Search for new structures of kinase inhibitors	IC_{50} (nM, μM, mM), targets (120 enzyme types and biosynthetic circuits)
16	Biochemical	Search for candidate treatment drugs for novel coronavirus infection from existing drugs	IC_{50} (nM, μM, mM), targets (120 enzyme types and biosynthetic circuits), filter application coronavirus
17	Natural materials	Search for combinations of moisture-resistant and strongly adhesive peptides	Adhesive strength (MPa), bending strength (MPa), moisture vapor transmission rate (g/m^2/day)
18	Natural materials	Search for additives for marine-degradable plastic (PHB)	Bending strength (MPa), yield strength (MPa), water absorption rate (g/g), tensile elongation at break (%), activation energy (eV)
19	Natural materials	Search for additives for biodegradable plastic (polylactic acid)	Bending strength (MPa), yield strength (MPa), moisture vapor transmission rate (g/m^2/day), tensile elongation at break (%)

LIB: lithium-ion battery Li: lithium PL: photoluminescence OD: optical density 4-HPPD: 4-hydroxyphenylpyruvate dioxygenase IC_{50}: 50% inhibition concentration
PHB: polyhydroxybutyrate

organic substance type, and one metal oxide type + one organic substance type.

Under conventional methods using manual work by humans, selecting compounds with an equivalent level of completeness is almost impossible. Therefore, candidate compounds are preselected based on the knowledge of the developer and only about (15 metal hydroxides + 15 metal oxides) × (20 organic substances) = 600 simulations can be expected to be performed. Even if a high-memory-capacity, one-teraflop-level computer can be dedicated to the task, only about five cases can be simulated in one day, requiring 120 days just for 600 simulations. Furthermore, the simulation results can be expected to be narrowed down to about 10 cases, and trial production and measurement evaluations will be required for each of these 10 cases.

Therefore, CI shortens the time required for materials selection and simulations to one-sixtieth the time of conventional methods, and reduces the number of times trial production and measurement evaluations need to be performed and hence the amount of test material waste to one-fifth the level. Even so, the coverage in studies of candidate compounds is more than 800 times higher than in conventional methods, making this a good example of the tangible effects of utilizing CI in upstream processes.

4. 2 Utilization of CI in Wide Range of Materials Development

In addition to the above, CI has a proven track record for search in a wide range of fields, including electronic components, optical materials, energy materials, and natural and bio-derived materials (see **Table 1**).

CI provides comprehensive support for organic and inorganic, single compound and composite materials, and both high and low molecular weights, enabling its utilization in research and development at a wide range of chemical and materials companies.

5. Conclusions

This article focused on MI utilization in materials development, but in actual operations, there is a need after this process to start up the production line, transition to the mass production phase, adjust manufacturing conditions in each phase, and dispose of any resulting material waste.

Hitachi High-Tech Solutions is optimizing this entire series of processes from development to mass production by developing measurement control technologies for obtaining manufacturing data, and techniques for optimizing manufacturing conditions that reduce the need for rework by utilizing this manufacturing data. In this way, the company supports the growth of the chemical and materials industries, and so the realization of a sustainable and prosperous society.

References

1) Sumitomo Mitsui Banking Corporation, "The Use of Materials Informatics for Materials Development" (Nov. 2019) in Japanese, https://www.smbc.co.jp/hojin/report/investigationlecture/resources/pdf/3_00_CRSDReport090.pdf
2) Mizuho Research & Technologies, Ltd., "New Trend of Material Development" in Japanese, https://www.mizuho-ir.co.jp/publication/report/2018/mhir15_mi_01.html
3) Mitsubishi Research Institute, Inc., MRI Monthly Review (Nov. 2017) in Japanese, https://www.mri.co.jp/knowledge/mreview/201711/i6sdu6000000r947-att/mr201711.pdf
4) Ministry of Economy, Trade and Industry, "The 6th Strategic Energy Plan" (Oct. 2021), https://www.enecho.meti.go.jp/en/category/others/basic_plan/pdf/6th_outline.pdf
5) European Commission, "Circular Economy Action Plan: For a Cleaner and More Competitive Europe" (Mar. 2020), https://ec.europa.eu/environment/pdf/circular-economy/new_circular_economy_action_plan.pdf
6) Mitsubishi Research Institute, Inc., "Challenges and Future Prospects for Biodegradable Plastics" (Apr. 2019) in Japanese, https://www.mri.co.jp/knowledge/column/20190408.html

Authors

Rei Kurokawa
ICT Business Planning Department, ICT Business Division, Hitachi High-Tech Solutions Corporation. *Current work and research:* Business development and marketing.

Tomio Iwasaki, Ph.D.
Materials Process Department, Center for Technology Innovation – Production Engineering and MONOZUKURI, Research & Development Group, Hitachi, Ltd. *Current work and research:* Materials design using molecular simulations. *Society memberships*: A Fellow of the Japan Society of Mechanical Engineers (JSME), an editorial committee member of the Society of Polymer Science, Japan (SPSJ), and a member of the Society of Materials Science, Japan (JSMS), the Institute of Electrical Engineers of Japan (IEEJ), and the Japan Institute of Electronics Packaging (JIEP).

Takashi Isobe, Ph.D.
Hitachi High-Tech America, Inc. *Current work and research:* Research and development of AI solutions.

Yoshihiro Okada
ICT Business Division, Hitachi High-Tech Solutions Corporation. *Current work and research:* Direction of AI and cloud solutions business. *Society memberships*: Japan Social Data Science Society (JSDSS).