

The logo for Sansan DSOC. The word "sansan" is in a lowercase, italicized, sans-serif font. Below it, "DSOC" is in a larger, bold, uppercase, sans-serif font. The entire logo is centered on a black rectangular background.

# **sansan** **DSOC**

Data Strategy & Operation Center

## **Presentation Material for DSOC, the Data Integration Division**

Sansan, Inc.  
April 14, 2021

Thank you for the opportunity today to introduce the Data Strategy & Operation Center (DSOC).

Until now, we didn't have much of an opportunity to spend time explaining to many investors or analysis about the Data Center & Operation Center, DSOC. Today, I would like to explain about the source and schemes of our competitive advantages with a focus on the details about the technology.

It would be a great pleasure if you would use this opportunity to deepen your understanding of DSOC.

## Introduction

---



### Satoru Joraku

---

**Sansan, Inc.**  
**Director, CISO <sup>(1)</sup>, DPO <sup>(2)</sup>, Managing Executive (DSOC <sup>(3)</sup>)**

Satoru cofounded Sansan Inc. in 2007 and he initially led product development of the Sansan, a B2B cloud-based contact management service. He now serves as the Director of Sansan's research hub DSOC, which conducts analysis and application of collected data, and refines the company's digitization technology. He also drives and implements internal security measures as CISO.

(1) Chief Information Security Officer  
(2) Data Protection Officer  
(3) Data Strategy & Operation Center

First, please let me introduce myself. Please see page two.

Currently, I service as Director, CISO, and DPO while being the Managing Executive of DSOC.

Sansan was founded in 2007 by five colleagues, including CEO Terada, and I was one of the co-founders.

I will deliver today's presentation as the Managing Executive of the Data Integration Division DSOC, but I would also like to briefly explain my roles as CISO and DPO.

At Sansan, we position information security and data protection as top priority management items. We are taking all conceivable measures for these purposes. The role of CISO and DPO is to promote these measures. Specifically, I am in charge of overseeing the penetration tests, where we use a white hat hacker to attack the internal system, and promoting the acquisition of personal information protection qualifications by employees.

Next, I would also like to give the history up until I joined Sansan.

My background is in engineering. Perhaps because of that, I had no interest in business cards. I had a strange uneasiness about business cards in paper, given that so many things were being digitized around the time of our founding, such as emails and PDFs. That's why I thought digitizing business cards might be a promising business.

Back then, just like how it is today, the US was the global leader of innovation in IT. But, because we started out with business cards, we thought there was potential for us to create some kind of new innovation that could be transmitted from Japan to the rest of the world and change how people encounter others. That was the motivation behind founding the business.

## Table of Contents

---

- 1** DSOC Overview
- 2** DSOC's Roles
- 3** DSOC in the Future

Next, I would like to go into the main topic.

## Table of Contents

---

### **1** DSOC Overview

- Mission and Roles
- History
- Organizational Structure
- Outstanding Engineers and Researchers

### **2** DSOC's Role

### **3** DSOC in the Future

I would like to give the presentation divided into three parts.

## Mission and Roles

---

Create a competitive advantage in the market as a data management division

### Mission

# Activating Business Data

Through publicly available business data, such as business cards, documents, financial results, and stock information, we build "Encounter Database" that creates new possibilities leading to the future of business and society.

### 3 Roles of DSOC

## Generating, Organizing, and Utilizing data

First, I will give an overview of DSOC. Please see page five.

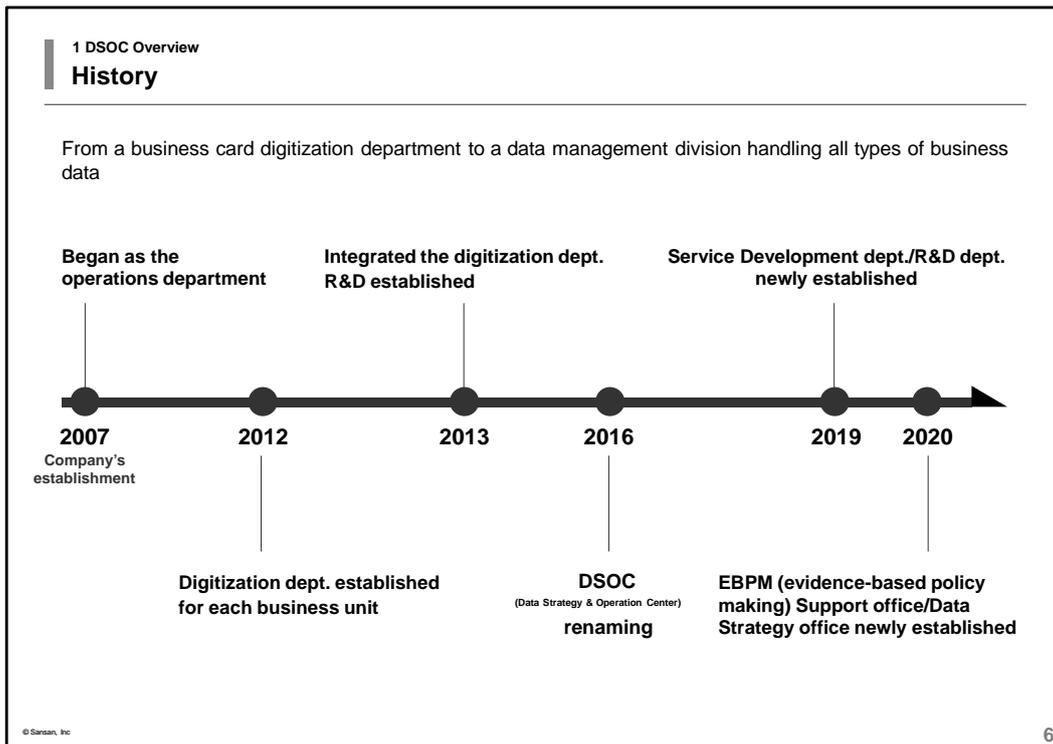
The greatest role of DSOC is to boost the speed of business growth in the Data Integration Division and to establish a competitive advantage in the market. So far, we have been working on digitizing business cards with speed and accuracy and at a low cost. We have supported the Company's robust growth by generating technologies that underpin our business activities.

Since our founding, we have been working on the opportunities created through business encounters involving the exchange of business cards. Currently, the business data that we handle include personnel transfer information, news, and other business data. It also includes invoices, contracts, and other business documents, which can be said to be proofs of business encounters between companies.

We uphold the mission, Activating Business Data, at DSOC to unlock the opportunities in this expanding range of handled data.

Under the mission, we have written the three roles of DSOC as the Data Integration Division. The first role is to generate data by aggregating diverse sets of business information, including business cards. The second role is to organize data by standardizing it and turning it into rich data. And the third role is to utilize that data.

I will explain these three roles in more detail in the second chapter.



Next, on page six, I would like to explain the history of DSOC.

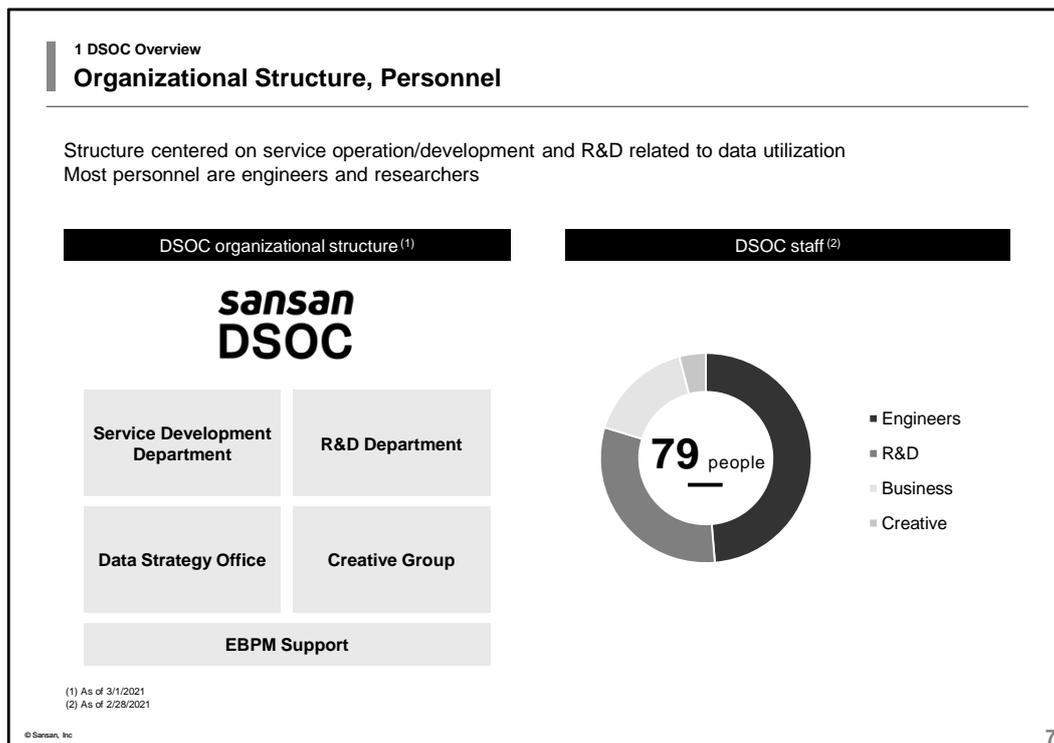
Back when the Company was established in 2007, we launched an operation department as the business card digitization division of the Sansan Business. After we launched the Eight Business in 2012, the business card digitization division was established separately in the Sansan and Eight Businesses due to differences in focal points such as speed and cost depending on each service.

However, operations fell completely behind the demand for business card digitization, which had increased in tandem with the growth of each service, and the Business Card Digitization Business had become a hindrance to business growth.

In 2013, we created an R&D organization that deals with the efficiency gains and automation of business card digitization by integrating the Business Card Digitization Division separately established in the two Business Divisions.

In 2016, we renamed the organization DSOC and newly added a role to it whereby it would not only digitize data but seeks ways to utilize it.

Ever since, the types of data handled have increased and DSOC started handling data other than business card data. That's when we set forth DSOC's mission as "Activating Business Data," and established teams such as the Service Development Department and R&D Department.



Next, I will introduce the structure of DSOC.  
 Please see page seven.

Currently, DSOC consists of three groups in addition to the Service Development Department and R&D Department.

The Service Development Department is made up of engineers and business professionals who serve to organize and generate data, starting with operations to digitize business cards. The department is responsible for developing and operating services provided to our Business Divisions. The R&D Department has researchers who are specialists in diverse fields enrolled in it and conducts R&D for the utilization of data.

In addition, we have established a Creative Group, where there are creators enrolled who are responsible for DSOC's branding, and a Data Strategy Office that formulates our data utilization strategy and conducts new business planning.

Recently, we have newly established an EBPM Support office. EBPM is an

abbreviation for Evidence-Based Policy Making and it literally means that policies are made based on evidence. This is an initiative that is strongly promoted by Japanese government agencies. EBPM Support offers evidence of evaluations and verifications based on the research results accumulated at DSOC aimed at supporting administrative activities.

## Outstanding Engineers and Researchers

In addition to engineers involved in digitization technology, DSOC has specialists with diverse backgrounds, such as social science and economics researchers

Image Processing

Natural Language Processing

Deep Learning

Machine Learning

Complex Networks

Economics

Business Administration

Sociology

kaggle



Grandmaster

PhDs

Science (2), Economics (1), Engineering (2)

Please see page eight. I would also like to introduce DSOC's personnel. At DSOC, we believe that innovation is born from completely new combinations, and we conduct our hiring activities with an emphasis on the diversity of members. As a result, we have personnel with various backgrounds and expertise playing active roles regardless of new grads or mid-career hires.

For example, ever since the launch of DSOC, we have had data scientists who are active members of Kaggle join the organization. Kaggle is a competition where data scientists from around the world compete to create an optimal model for a problem, and about 10 Japanese people possess the highest-ranking title of Kaggle Grandmaster. In addition, we have researchers who possess PhDs.

As for their areas of expertise, we have appointed not only engineers specializing in data science, such as image processing related to digitization technology or machine learning, but also researchers in the humanities, social sciences, and economics.

## Table of Contents

---

### 1 DSOC Overview

## 2 DSOC's Role

- Data Generation
- Data Organization
- Data Utilization

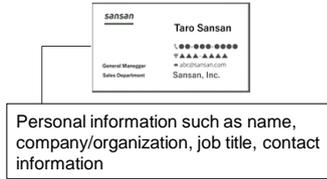
### 3 DSOC's in the Future

Next, I would like to explain the role played by DSOC.

## Data Generation: Unique Issues in Business Card Digitization

Business card information must be accurately digitized for effective data utilization  
Practically difficult to correctly convert business card info into data using only OCR because of unique issues in business card digitization

### Accuracy needed for business card digitization



If even one character in the data is incorrect, the value as information is damaged.

Digitization of business cards requires as much as 100% accuracy.

### Issues in business card digitization

- **Atypical**  
Varying designs, shapes, fonts
- **Image Quality**  
Depends on shooting environment
- **Item Identification**  
Judgment of names, titles, etc.  
Unified notation rules

Accurate digitization is difficult if only using OCR.

Please see page 10.

I will explain the three roles of DSOC in order. First, I will explain about the role of data generation.

As a premise of data generation, I will explain the unique issues in business card data conversion.

This is quite obvious, but if you make a mistake in the phone number or email address by just one number or letter, you can't make the call or send the email correctly. In other words, inputting invalid information leads to damaging the value of the information. To utilize data, it is essential to convert the business card information into data with an accuracy that is as close to 100% as possible.

On the other hand, there are various unique issues in converting business card information into data.

For example, the design, shape, and font size are different and unique.

Also, depending on the environment in which the user takes the photo, there might be problems in image quality, such as an unstable photo, a photo with a shadow appears in the business card image, or characters that are difficult to recognize due to lighting.

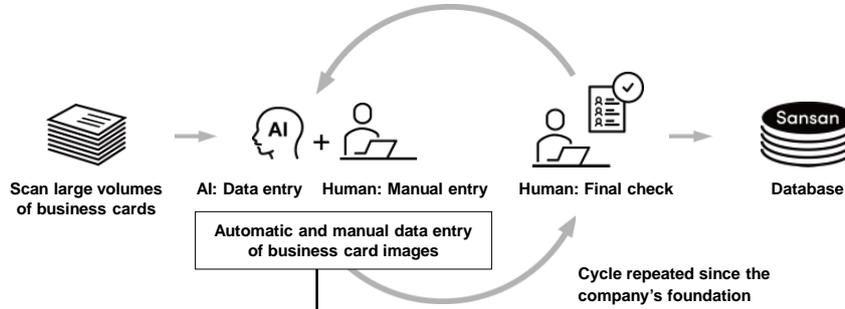
Another issue is to identify items such as the name and title of the person. For example, it is necessary to distinguish whether the line that says Analyst is your title or a qualification. Even though a human person might be able to judge such things instantaneously, there is a very high hurdle to be cleared for machines to be able to identify such information.

Due to these issues, it is currently difficult to correctly automate and digitize business cards using only general OCR.

## Data Generation: Business Card Digitization Operations

Mechanism & tech to achieve 99%+ accuracy in digitization – source of competitive advantage

Prompt and accurate digitization of a large volume of business card information



### Combination of Technologies

Automatic detection and microtasking of business card information

Automatic detection of business card language

Automatic sorting to operators

Diverse network of data entry operators

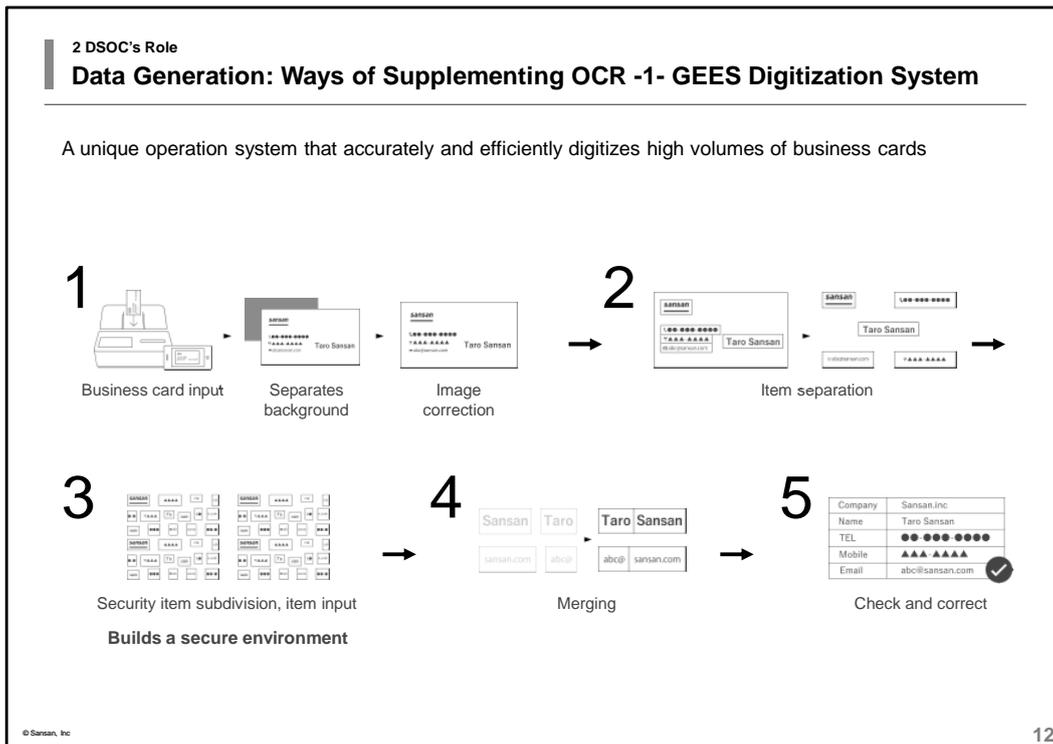
Image processing technology

lease see page 11.

As a solution to the problem stated on the previous page, we are complementing the machine data conversion with AI OCR and human power. Our strength is that we can utilize not only technology but also human power.

By combining this human-powered operation and technology, “Sansan” has achieved a business card data conversion accuracy of 99.9%, and we have acquired an overwhelming market share of 83% in the cloud business card management service market.

From the next page, I would like to introduce specific examples of complementing AI OCR.



Please look at page 12.

At “Sansan”, we have built our own business card data conversion operation system, which we call GEES. GEES is an operation system originally developed by DSOC that converts a large number of business cards into data accurately and efficiently. It divides the business card image captured by the machine and makes the unit of work smaller. By doing so, it creates a scheme where input work can be performed anytime and anywhere while ensuring accuracy and security.

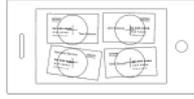
The business card data conversion using GEES is conducted based on the following process. The user first scans or takes a photo of the business card. Then, the imported image of the business card is sent to us. Then, the characters are processed so that they are easy to read.

Next, the group of characters on the business card is divided by a machine. After that, the information is classified by item, such as company name, first name, and last name. Then, in consideration of security, it is segmented into small pieces of information until it is no longer legible.

After that, the data is inputted by machines and humans. In the human input process, we prevent input mistakes by having two or more operators input the same image, and the input process is performed until the results match. As a result, the accuracy of data conversion is improved.

## Data Generation: Ways of Supplementing OCR -2- AI/Image Recognition Tech

High-speed and high-precision digitization of business cards using unique image recognition technology with AI



### Smart Capture

Technology that enables users to receive results



### Item Segmentation

Distinguishes items from the business card design without reading letters



### Language determination

Judges language without reading letters



### Mistake Detector

Learns errors trends and predicts potential mistakes

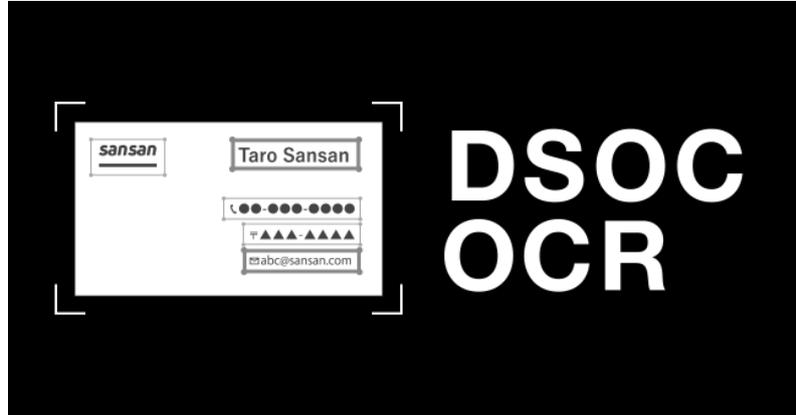
Please turn to page 13.

DSOC has independently developed various image recognition technologies and AI to realize high-speed and high-precision business card data conversion.

We are developing technologies that contribute to automatic data conversion, such as smart capture that recognizes business cards and delivers the result of data conversion to the user in a few seconds, and item segmentation that estimates the item areas such as the first name, last name, and company name from the business card image. In addition, there is a certain degree of regularity to errors made by humans when inputting data. So, we use AI to analyze those errors to complement the data conversion accuracy.

### Data Generation: Ways of Supplementing OCR -3- DSOC OCR

We independently developed an OCR engine specializing in business cards, and that can read email addresses with 99.7%+ accuracy. We will continue researching and developing digitization for all business card items.



Please see page 14.

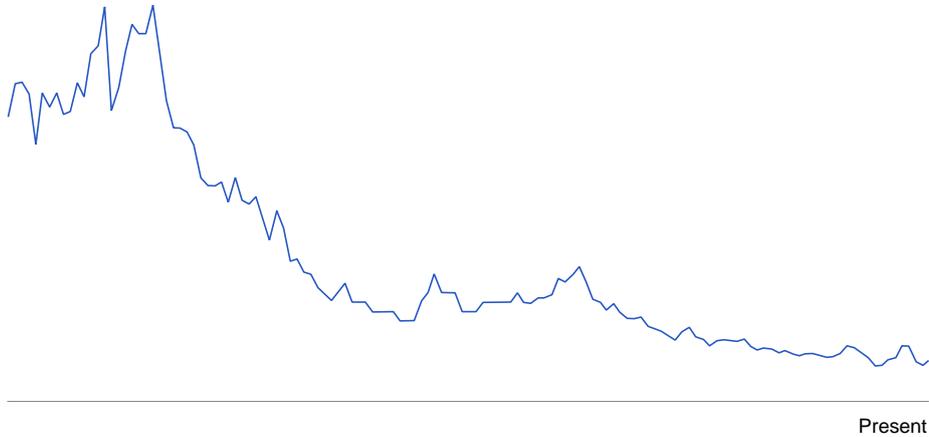
The DSOC R&D department is working on developing an original OCR engine specializing in importing business cards and offers this service as DSOC OCR.

I mentioned earlier that achieving high accuracy with the current general OCR is difficult in terms of converting business card data. However, we have taken on the challenge of developing our own OCR engine. The DSOC OCR makes it possible to automatically convert e-mail addresses into data with an accuracy of 99.7% or higher.

DSOC OCR can be used not only for email addresses but also names, and we are proceeding with development so that it can also be applied to all items for business cards in the future.

**Data Generation: Changes in Digitization Cost per Business Card**

With the evolution of business card digitization, the cost per card will be less than 1/20 what it originally was.



Please see page 15. This is a graph showing the trend for data conversion cost per business card until now.

Looking back, it is unthinkable how we conducted our operations when the Company was first established. We used to do everything by hand, and during the busy season, it was quite common for us to come to the office and input business card data. We conducted data input only by manual labor, but it was clear that there was a limit to continuing these labor-intensive operations, as the demand for data conversion increased as our services grew.

That was how we developed GEES, an operation system combining machines and humans, which I explained earlier. GEES divides operators into three levels to maximize the efficiency of human resources.

First, it consists of about 300 directly hired operators who are responsible for highly confidential and complex data conversions. Second, it consists of over 70 outsourced operators. Lastly, it consists of hundreds of thousands of managed crowd workers. With such a system, we have built a system

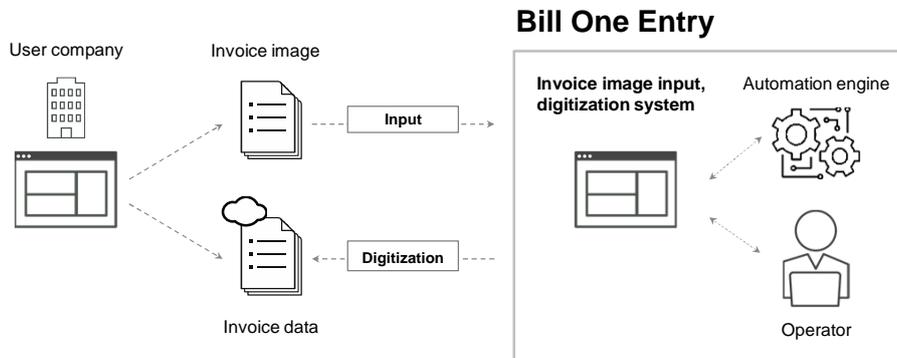
that can flexibly respond to the waves of busy and off-season demand by changing the human resource costs related to data conversion into variable costs.

As a result of such efforts, we have lowered the data conversion costs per business card to less than one-tenth of what it used to cost at the time of our founding. At the same time, we have significantly improved the speed of data conversion. At the time of our founding, it took us up to three months until delivery, but we are now able to handle the conversion within several hours. We have created a structure where the number of sheets that can be converted to data in a month has increased by several tens of thousands of times compared to when we were first founded.

## Data Generation: Using Digitization in Operations for Other Fields

Expanding accurate business card digitization operations to other fields, such as invoice digitization

**Bill One**  
powered by Sansan



Please see page 16.

Since our founding, we have grappled with the data conversion of business cards. We have also started activities to utilize our business card data conversion operations in fields other than business cards such as invoices and contracts.

For example, for “Bill One”, a cloud-based invoice receiving service, we receive invoices in various methods or formats, such as paper invoices delivered by mail or PDF invoices attached to emails, and we accurately convert them to data.

Invoices have different industry rules from business cards such as itemized lists of amounts or rows that indicate tax charges. But we horizontally apply the business card data conversion operation and image recognition technology that we have accumulated, up to now, to make it possible to conduct accurate data conversion within several hours of receiving the invoice.

In addition, we are conducting similar activities in terms of Contract One by offering a data conversion solution for contracts.

## Data Organization: Challenges in Effective Data Use

Some accumulated data cannot be fully utilized for sales and marketing because of data problems such as outdated information and empty data items.

Company name	SAN SAN, Inc.	Old business name
HP		Empty data field
Tel.	+81-3-6758-0033	
Address	5-52-2 Jingumae, Shibuya-ku, Tokyo, 150-0001, Japan	
Contact name	Taro Yamada	
Department	Digital Marketing	
Job title	Director	Old job title
E-mail	Yamada.taro@33i.co.jp	

Next, I will explain the second role: data organization.  
Please see page 17.

It goes without saying that the importance of utilizing all kinds of data in business is growing. There are some issues in data stored by companies in the past that prevent it from being effectively utilized due to imperfect data such as the company name being old, missing information, or the address being incorrect.

Companies have adopted various types of SaaS tools, spurred in part by the recent trend of digital transformation. However, because customer information is distributed separately across each service, even the same customer information is difficult to manage or operate as a whole company.

Therefore, it is crucial for the data to be organized, and DSOC is conducting various R&D related to organizing data aimed at effectively utilizing data.

## Data Organization: Building Data Infrastructure

Use accurate and up-to-date business card data, enrich with other items, establish data infrastructure for effective business use

### Normalizing & updating

Company name	Sansan, Inc.	Correct name
URL	https://www.corp-sansan.com/	
Tel.	+81-3-6758-0033	Fill in missing info
Address	5-52-2 Jingumae, Shibuya-ku, Tokyo, 150-0001, Japan	
Contact name	Taro Yamada	
Department	Digital Marketing	
Job title	Director	Updated job title
Email	Yamada.taro@33i.co.jp	

### Enrichment with added info

Reg. name	Sansan, Inc.	Est.	June 2007
Corp #	4010001120965	Fiscal end	May
TDB Code	989671019	Position	Representative Director & CEO
Main business	Information processing	Name	Chika Terada
Industry	Package software	IPO	Yes
Capital (yen)	1-10 billion	+	
Employees	500-1,000	Performance information, affiliated companies, company news, etc.	
Revenue (yen)	10-30 billion		

Please see page 18. As an example of organizing data, I will explain the process after converting business card data with “Sansan”.

“Sansan” makes it possible to update old company names with new ones after data conversion of business cards in the case the company name changes and also standardize URL domains. Furthermore, we create a database that is based on the business card information but also supplemented with corporate data from Teikoku Databank and earnings information, so that we can deliver relevant news.

In this way, we can standardize and optimize the business card information converted to data. And we add all kinds of other business information to it to turn it into rich data. We organize the data so that it can serve as a data platform that can be effectively utilized in business.

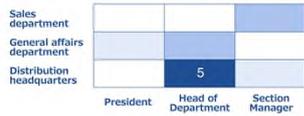
Moreover, although this is a paid option, it is also possible to utilize data organized on “Sansan”'s database to cleanse the customer information in other companies' tools and integrate the data.

**Data Utilization: "Sansan Labs" changing the future of business**

"Sansan" provides experimental functions to solve business problems via use of cutting-edge data

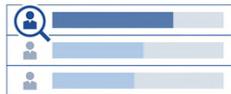
Simply scanning a business card lets the user experience DSOC's experimental analysis functions that support efficient sales activities and in-house staff's knowledge and network utilization.

**Sales Tech: supporting sales activities**



**An Overview of Points of Contact for Each Customer**

Account-based marketing (ABM) dashboard (β)



**Find a Key Person in a Business Area**

Hot lead recommendations (β)

**HR Tech: talent management**



**Search Your Colleagues Based on Their Expertise**

Search internal knowledge (β)



**Understanding Strengths through Trends in Connections**

Businessperson type analysis (β)

Finally, I would like to explain the third role: data utilization. Please see page 19.

We offer "Sansan Labs," a function to support the streamlining of sales activities and utilization of the knowledge and connections of in-house personnel just by scanning the business card. This group of functions on "Sansan" allows users to experience the results of research in various business realms conducted by social science researchers.

For example, the colleague knowledge search function allows users to search the knowledge of colleagues about specific industries or fields by entering a keyword. Similarly, by searching the name of a colleague, it is possible to know what kind of knowledge your colleague possesses. It is possible to use this function for the selection of project members of the collection of internal information.

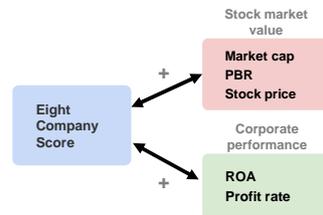
## Data Utilization: Work by the EBPM Support -1- "Eight Company Score"

This unique index quantifies corporate assessments using the business network of the business card management app "Eight." The score is closely related to the social aspect of Environment, Social, Governance (ESG), and we have started research on developing an investment trust index.

# Eight Company Score



Can quantify and quantitatively grasp the strength of relationships between B2B companies



Research results showing relationship between stock market value and corporate performance

Please see page 20.

We established an EBPM support office to bolster administrative policy-making based on evidence. In this support office, we have developed proprietary indicators for quantitatively and qualitatively evaluating the reputation of companies among external stakeholders. We call this survey "Eight Company Score."

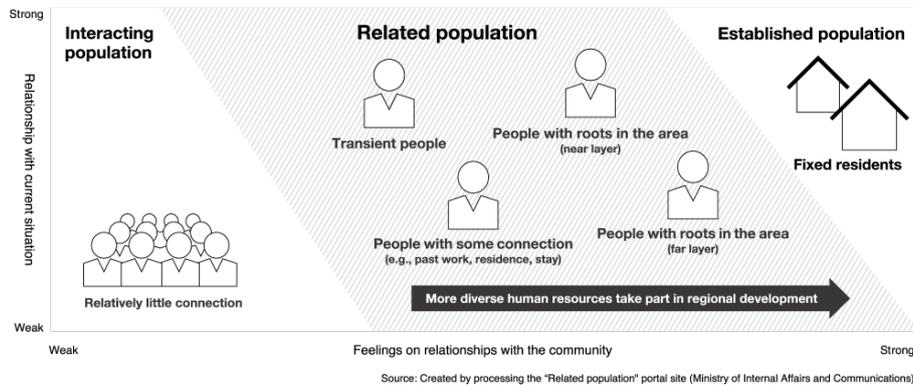
Until now, there has never been data that directly measure corporate evaluation by stakeholders who have direct contact points in business with the Company. By using "Eight Company Score," it is possible to quantify the survey results targeting the users of our business card app Eight. In this way, we have been able to develop a unique indicator.

Currently, we are conducting survey on roughly 1,400 companies. When verifying the correlation of this indicator with stock market valuation and corporate earnings to study the effectiveness of the indicator, we found that there was a correlation with P/B, market capitalization, and profit margins, demonstrating that there is a strong connection between a company's social reputation and profit generation.

We believe Eight Company Score will be a beneficial indicator not only for investors but also people within companies as a way of measuring a company's future potential and sustainable growth.

## Data Utilization: Work by the EBPM Support -2- “Business-related population”

This unique index shows the degree of the relationship between the region and business. It supports administrative efforts, looking at use in regional revitalization and disaster-recovery policies.



Please see page 21.

I would like to introduce another example of the activities carried out by the EBPM support office. This is the “Business-related population.”

“Business-related population” is an original indicator showing the degree of relationship between a businessperson in one region and a businessperson in another region. This data is calculated based on the number of users imported from business cards of people who are based in the city, ward, town, or village of the indicator by using Eight’s business network. We expect this indicator to be used for regional revitalization and disaster recovery policies. We intend it to support administrative decision-making.

## Table of Contents

---

1 DSOC Overview

2 DSOC's Role

**3 DSOC in the Future**

– DSOC's Aims as We Move Forward

Lastly, this is our vision of DSOC in the future.

**Become business infrastructure**

**Optimize business encounters to solve  
business problems faster and more efficiently**

Please see page 23.

I would like to give two broad explanations of what DSOC aims for in the future.

The first one is to become a business infrastructure.

Starting from this year, the Company has set forth the vision to become a business infrastructure. We hope to become a company that feels like common sense to be implemented by offering services that blend into customer businesses. That is the type of world we hope to create.

Based on this vision, DSOC aims to be an infrastructure for business data as a Data Integration Division. We hope to create a world where the most accurate and quickest way of looking up a company or businessperson is search about them using our product.

The second vision is to optimize business encounters to solve business problems faster and more efficiently.

By encounters we do not mean just the encounters between people. We

also mean the encounters between companies and individuals and companies and other companies. Oftentimes, when talking about the keyword encounters, people tend to focus on who they are encountering.

But our pursuit doesn't end there. We also aim to provide services that make encounters even more important than just who we meet.

For example, when you want to sell a product, I think you would select the target customers based on criteria like industry or number of employees.

But that's insufficient. There are other factors that need to be considered, such as the timing or phase during which these customers would want the product. We aim to be able to know those changes and timings based on data.

By coming face-to-face with these various factors and business data associated with such encounters and continuing to conduct research, we aim to create schemes and technologies for solving business issues.

Ultimately, we would like to link that to the Company's mission of turning encounters into innovation.

This concludes my presentation. Thank you.

**sansan**

