celonis

Object-Centric Process Mining

The next frontier in business performance

Prof. dr. ir. Wil van der Aalst
Chief Scientist, Celonis

Introduction

For years, organizations have embraced process mining technology to drive efficiency, improve operational excellence, and stay ahead of the competition. Today, the challenges of an unpredictable business landscape and macroeconomic headwinds require even greater speed and accuracy when it comes to getting to critical insights and optimizing business processes. Understanding how processes are interconnected — and how to improve them holistically — has become a mission-critical 'must-have'.

Enter Object-Centric Process Mining (OCPM). OCPM has emerged as a way to significantly increase the speed, value capture, accuracy, and use case potential of what is already a transformational discipline for the enterprise.

At Celosphere 2022, the annual Celonis user conference, Celonis revealed Process Sphere™, the first commercially available solution that fully embraces object-centric process mining.

OCPM is a revolutionary advancement for businesses as it allows them to:

- View all business activities from any perspective using a single source of truth
- Discover novel and valuable opportunities that live at the intersection points of processes and departments
- Move from two-dimensional views of processes to a three-dimensional and dynamic view of the entire business

Simply put, Object-Centric Process Mining acts as a force multiplier for operational visibility, alignment and efficiency, and enables organizations to reach new levels of business performance at the top, bottom and green lines.

This white paper will help you understand Object-Centric Process Mining: from the evolution of the discipline to its relevance for businesses, the challenges it overcomes and the powerful capabilities of Celonis Process Sphere.

1 A brief history of process mining

Before we look at Object-Centric Process Mining (OCPM) and Process Sphere, it is important to look back at traditional process mining. It is only by placing OCPM in the context of process mining as a whole that we can understand how innovative it is.

Process Mining was the first technology to connect process science and data science.

Practitioners of process science were more interested in processes than data and focused on process modeling, simulation, Business Process Management (BPM) and Workflow Automation. Meanwhile, data science practitioners who specialized in Machine Learning (ML) and Artificial Intelligence (AI) were very interested in data but less interested in processes.

Processes represent complex, dynamic and non-linear interdependencies of objects and events.[1] Therefore, supporting process improvement can be more challenging than, for example, creating an algorithm which classifies images of pets into 'cat pictures' and 'dog pictures'.

Process Mining was a much-needed bridge between process science and data science. The first of these bridges appeared in the late 1990s in the form of process discovery based on event data.

At the time, many workflow management projects were failing due to the gap between how people thought their processes ran and the reality of how those processes actually ran. Process discovery pulled back the curtain on the truth of how businesses were operating – there was no looking back.

This nascent form of process mining evolved beyond discovery to include other disciplines like conformance checking, prediction, organizational mining, decision mining and more. As the field matured over time, commercial products such as Celonis became available. (Fig. 1

[1] We will be using the terms 'events' and 'activities' interchangeably throughout this whitepaper, as is standard.

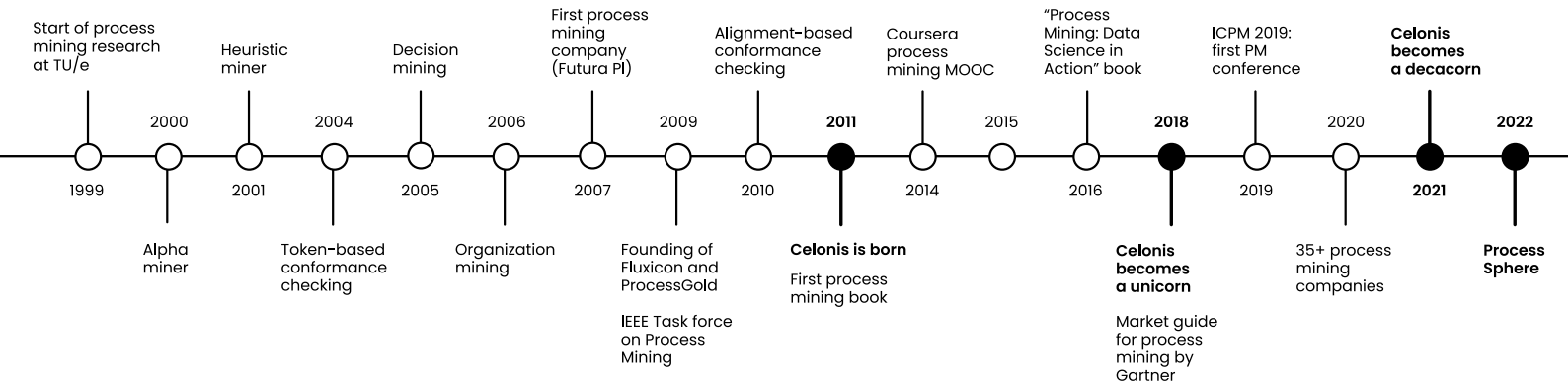


FIG. 1 Timeline of events in the history of process mining

The proliferation of new vendors and widespread adoption of process mining in recent years has shown that businesses are finding significant value in the adoption and implementation of this powerful discipline. This is also in line with the analyst forecasts about the market. Gartner predicts that the market for process mining will continue to grow and reach $2.3 billion by 2025, with a double-digit compound annual growth rate (CAGR) of 33%.[2] Process mining is the technology category that is growing the fastest compared to other categories, such as, Robotic Process Automation (RPA), Enterprise Low-code application platforms (LCAP), Integration Platform as a Service (IPaaS), and Business Intelligence (BI).

How process mining has worked up until now

There are two primary ways to think about process mining. Top-down, and bottom-up.

Taking the top-down view means you look at what process mining is through the prism of the tasks you're able to perform. You can see these four tasks in Figure 2 (below), but we will lay them out in more detail here.

[2] Gartner: Forecast Analysis: Process Mining, Worldwide (2021)

Process Mining begins with the extraction of event data from business information systems, such as ERP, CRM and SCM. Information systems are typically composed of multiple parts, sometimes by different vendors and even within the system of a single vendor, you may find dozens or thousands of different tables.

From these tables, event data is extracted and configured in order to discover how processes actually run. This first 'top-down task', process discovery, is often surprising because the process reality is often very different from what people believe is happening.

Once you understand how processes are truly running, you can move forward with the second 'top-down task': conformance checking. Conformance checking allows you to detect common problems in your processes by revealing deviations and variations from an ideal baseline. If you have enough data and your processes are stable, you can perform the third task: prediction. Process mining allows you to predict when and where bottlenecks will occur and where a process is likely to deviate from the desired path.

However, even if predictions are out of scope, through discovery and conformance checking, you can perform the fourth and final task: triggering action.

With Celonis, you can automatically set up and trigger an Action Flow to resolve compliance or conformance issues. An action flow is a small workflow composed of corrective actions triggered by process mining analytics. This enables real-time, corrective intervention into a process in order to improve it based on the diagnostics that are generated by process mining.

To put it another way, process mining best practice follows one of the two approaches below to close the gap between insights derived from data and action:

- Extract -> Discover -> Check -> Predict -> Act,
- Extract -> Discover -> Check -> Act.

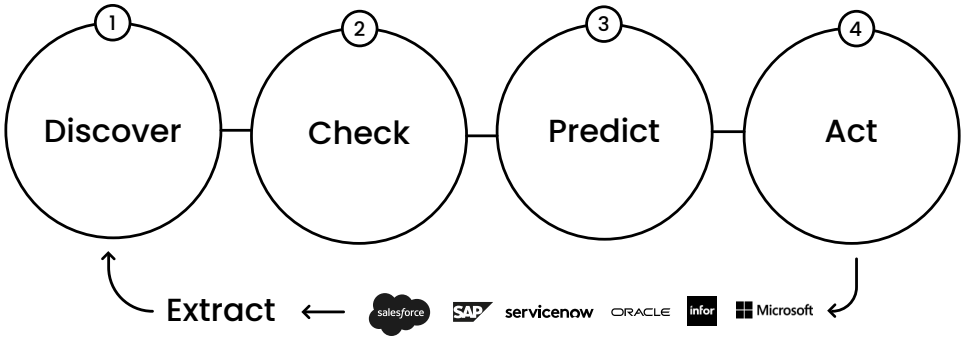


FIG. 2 Process Mining best practices approach.

Conversely, the bottom-up view allows you to understand what process mining is through the prism of the event data and process visualization.

As this is a method for understanding process mining, not a method of how to perform process mining, the first step is the same: You extract the event data from business information systems.

This is where things differ, but only because we are choosing to look at the process from a different angle, not because the process is different. By combining a Case ID, an activity and a timestamp, you get an event log which gives you an idea of what happened and when (Fig. 3). A "Case ID" in process mining is a unique identifier assigned to a specific instance of a process. It is based on the notion of a single object type involved in a business process (e.g. sales order item, delivery, etc.) and it groups all the activities and events associated with a particular process instance for analysis. For example, to create an event log, one needs to decide on which object type to use for the scope of analysis, like a "sales order item". Once the notion of "sales order item" is defined, every sales order item will present a Case ID in the event log that coincides with activities and time stamps.

There may be other helpful information attributes that can be added to the event log; for example, the resources being used, cost information, location and product information. This is something that we need to bear in mind when we start talking about object-centric process mining because a fundamentally different assumption about the fabric of event data is made.

CASE ID	ACTIVITY	RESOURCE	TIMESTAMP
6350	place order	Aiden	2018/02/13 11:34
6283	pay	Lily	2018/02/13 14:29
6253	send invoice	Sophia	2018/02/13 16:55
6283	deliver	Jack	2018/02/13 16:55

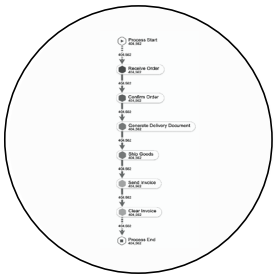
...

case +
activity +
timestamp +
= event

FIG. 3 The creation of an event log

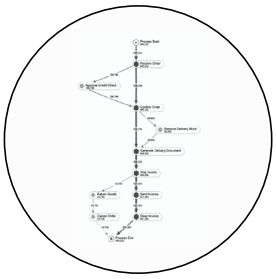
With the event log, we apply process mining to reveal the most common path of a given process. We can observe the most frequent sequences of activities happening for certain cases (Fig. 4a). If we start to incorporate paths that are less frequent, then we can see additional — often undesirable — activities, loops and decisions that have been made (Fig. 4b).

MOST COMMON PATH



4a

EXPECTED PATHS



4b

REALITY

4c

FIG. 4 The comparison between the most common path, a process with some expected variations and a true-to-life process map, sometimes referred to as a 'spaghetti diagram'.

Finally, when we look at all the data, we typically see a "spaghetti diagram" of reality that exposes all of the variations, connections and activities involved (Fig. 4c). The stark contrast between the spaghetti diagram and the most common path highlights opportunities for process improvement.

When looking at something as overwhelming as a spaghetti diagram, it's useful to apply the Pareto principle — also known as the 80/20 rule. Simply put, 20% of cases are responsible for 80% of the variations in the process. By only looking at the most common path, or including the closest set of deviations — which is often what happens in manual process discovery — you miss out on the major wins hiding at the fringes.

In summary, to perform process mining, we first want to discover what's really going on in the business process. Then we want to check where and when undesirable activities are happening and what the root causes are. If we have enough data, we can go as far as predicting what will happen to a case or a process. This forms a foundation to clearly see where opportunities exist in order to automatically take corrective action.

2 Challenges: What problems need to be solved?

Given its transformational potential, it's no wonder process mining is becoming increasingly popular. Large organizations across industries are using and benefiting from the value of process mining. So much so, that they want more.

However, we have reached a point where adopters of process mining have needs that traditional process mining can't meet.

For example, consider you're looking at a specific process but want to examine it from another point of view using a different notion for the Case ID (e.g., invoice instead of sales order item). For this, you must go back to the data and make another extraction to create a new event log. This can be a painful experience that impacts the quality and speed of delivering process transformation.

When analyzing a process for the first time it is not uncommon to spend 80% of your time on data extraction and configuration and only 20% on the actual analysis, which is clearly not ideal.

Furthermore, when the data is extracted and configured into an event log based on a Case ID, it is flattened in a 2-dimensional structure. When this flattening happens, the interactions between all the other objects involved in the business process are lost.

Traditional process mining only allows you to view a process from the perspective of a single object (e.g., sales order or invoice), rather than the relationship between all objects involved. This is why we describe the data as 'flattened' — because the focus is on a single object, and the model describes the lifecycle of one instance of that object, "a case", in isolation. Picking a Case ID corresponds to projecting reality onto a single object type for analysis (e.g., orders, customers, items, suppliers, deliveries, etc.).

For process mining to reach its full potential, it must empower us to explore the relationships between different types of objects in a process.

Introducing an object-centric approach

Realizing operational excellence requires tight alignment across departments working with the many objects involved in long-running end-to-end processes. This is what OCPM promises to enable with simplicity and speed.

Traditional process mining allows us to look at a single object, like a sales order. One can see how sales orders flow through the sales department. Within the sales department, we can apply process mining and find opportunities to improve that flow.

However, there are different departments that use different objects that are related to sales orders. For example, the production department uses production orders, finance uses invoices, the procurement department uses purchase orders and so on. These object types are often interrelated across the real-world execution of end-to-end business processes.

Let's look at a simple example of an order from an online retailer (Fig. 5). Suppose that a customer places an order that consists of four items. Of the four items, one is in stock and three need to be produced, which triggers the creation of three production orders in the company's production management system. The in-stock item can be immediately delivered to the customer, but the other three items must be shipped later. After all of the items are delivered, an invoice is sent. In this scenario, there is one sales order, four sales order items, three production orders, two shipments and one invoice.

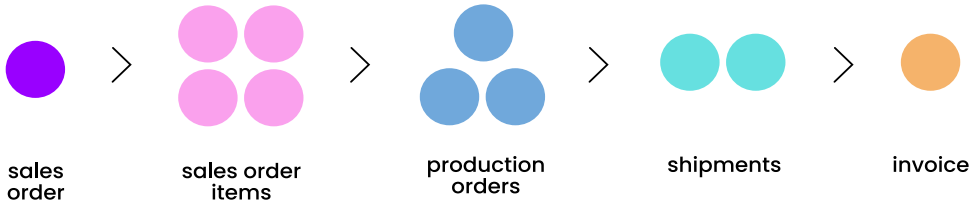


FIG. 5 The multiple objects involved in a single order

The main thing to take away from this example is that the processing of a single sales order can involve multiple objects scattered across different systems and departments. In addition, if we consider the reality of multiple sales orders, it is possible that a single shipment includes items from different orders.

This can happen when the same customer places multiple orders and the company batches the shipments together. This example reflects the complex real world of one-to-many and many-to-many relationships involved in end-to-end processes. Object-Centric Process Mining aims to capture these one-to-many and many-to-many relationships to more quickly and precisely enable optimal and coordinated performance across all operational processes. With this view and understanding, use cases like improving shipment efficiency through order bundling or reduction of end to end lead times from procurement through production become easier and faster to address.

Why is this important? As noted above, traditional process mining requires that we analyze each object separately. Shipments, orders, invoices and goods receipts require a separate event log for analysis in isolation. However, these objects and their corresponding activities are strongly intertwined. Problems that are related to one object type may have effects on many of the other object types and can create undesirable outcomes. For example: an issue procuring a battery could end up causing a late shipment, which causes a negative customer experience, which causes an order cancellation.

The reality in process mining today is that analyzing objects in isolation makes it more challenging to move the needle on higher-level company goals that are a result of several, interconnected processes working in harmony. Think about customer satisfaction as an example. High levels of customer satisfaction require coordination across many processes like; sales, procurement, production, quality, delivery, warranty, returns and customer service. Ensuring a positive experience for the customer at every touchpoint requires the ability to identify areas for improvement between these connected processes quickly and efficiently. Failures often happen at the intersection points between processes and can cause invisible breakdowns in the desired outcome. OCPM enables a more comprehensive and faster analysis of the business to better coordinate processes and move the customer satisfaction needle in a positive direction.

This is an example of 'butterfly effects' in business[3]. The solution is to approach processes in a more holistic way and look at multiple object types and their interrelatedness at the same time. This forms the basic idea of Object-Centric Process Mining. OCPM can be understood as a shift from 2-dimensional to 3-dimensional analysis that allows us to view a process from any angle without requiring multiple data configurations or data flattening.

[3]A butterfly effect in business is an event that occurs in one business process that can have significant effects in a related process down stream. They are very difficult to identify as they can appear benign, or even beneficial from a siloed point of view, but harmful from an end-to-end point of view.

3 A deeper technical dive: Dealing with convergence and divergence

As noted above, OCPM doesn't require data to be flattened like traditional process mining does. This is important, because when you flatten the data, you introduce the problems of convergence and divergence.

We can use the example of a race car pit stop to examine the challenges of convergence and divergence more closely. As with the example above, our race car scenario also illustrates how most real-life events involve multiple objects (often of different types) and that reality can complicate the analysis of a process using traditional methods.

A pit stop for a Formula One race car is an event that only takes two seconds. When we consider a pit stop as one high-level activity, we can see that many objects are involved in this activity:

- a car,
- a driver
- twenty pit crew members
- eight tires (four new and four old)
- and so on

To maintain context about the process — that we are in a race, with the goal of winning, and a pit stop is an activity to help realize the goal — it is important to capture the relationships between the multiple objects involved. Note that problems related to any of the objects will delay the whole pit stop.

If we look at traditional process mining, we are trying to follow a single object and we create a process model, for example, for a tire. If we did this, we would never see the multiple objects involved in the pit stop with full clarity. All of these other objects and events involved would be flattened in a disconnected manner behind the view of a tire, the single object.

Although there may be value in understanding the journey of a tire, this 2-dimensional view does not display the full reality. For example, one needs to cope with the problem of data convergence.

Convergence

Assume that for our single ‘pit stop’ event we select ‘tire’ as the notion for our Case ID. A traditional process mining analysis will tell us that the pit stop happens eight times.

This happens because we must replicate the data for each case (i.e., the ‘pit stop event’) once for each object of type ‘tire’. The unintentional replication of events caused by flattening the data is known as convergence.

Convergence can produce misleading diagnostics if one does not account for it. For example, in a sales order process, if events are unintentionally replicated (because we use sales order items as the notion for our Case ID), you could end up thinking more sales have occurred than actually took place. Such convergence problems can impact frequencies and time statistics without stakeholders realizing it.

Divergence

The other problem with flattening 3-dimensional data into 2-dimensional event logs is divergence.

Still using our pit stop example, say we want to look at low-level events, like:

- Stop
- Remove tire
- Mount tire
- Drive

If we consider these low-level events and use the object ‘car’ as our notion for the Case ID, we run into the divergence problem.

Our data shows four ‘remove tire’ events and four ‘mount tire’ events per pit stop. This is not inaccurate because these activities are happening multiple times per car. However, because the four ‘remove tire’ events and four ‘mount tire’ events are being analyzed through the same Case ID, the causalities between these events are lost.

Because we can’t tell which tire is being removed or mounted, we lose vital information about sequencing. We could see the sequence:

Therefore, you end up with a process model that suggests it is possible to mount a tire before removing it, which obviously isn’t true to the real-life process.

The example of a pit stop may seem far-fetched, but we see the same phenomena when considering sales orders and sales order items. Although events for sales order items follow a clear pattern (e.g., produce before ship), this information gets lost when using sales orders as a notion for the Case ID, because the model may suggest that an item may be shipped before it is produced. Next to concurrent process activities, this is one of the main reasons for the spaghetti-like process models shown in Fig. 4.

Put simply, convergence and divergence lead to misleading diagnostics (when not interpreted correctly) and overly complex process models that fail to capture clear patterns.

Object-centric process mining addresses these problems by using representations for event data and process models which mirror reality.

4 Solution: Object-Centric Process Mining

Object-Centric Process Mining allows you to look at business activity from any angle.

The following image (Fig. 6) shows an abstract visualization of objects and events. The black circles represent events like the placing of an order, the change of a ship-to, the change of order items, delivery, etc. The colored circles marked (ot1, ot2, ot3) represent object types (e.g., sales orders, customers, packages, etc.) and the start and end of objects of the corresponding type. The arrow lines correspond to the unique objects (e.g., sales order 11, sales order 12, etc.) and indicate how they flow through the network of events.

With every event, we can see the incoming objects, and with every event, one or more objects are involved. Fig. 6 shows the full reality of the process

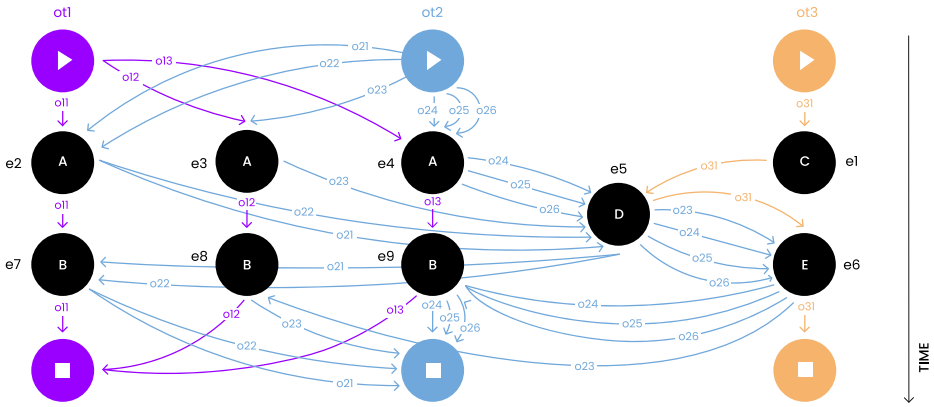


FIG. 6 An abstract visualization of objects and events that reflects the full reality of a business process.

If we see this view as reality but then try to visualize this with traditional process mining, we run into the challenges described above. In Fig. 7, we see the separation of the same objects shown in Fig. 6 grouped per object type. Because each object's lifecycle is captured in isolation, we need to replicate the events. For example, event "D" (a delivery) happens only once in Fig. 6, but is counted seven times in Fig. 7. This occurs because there were seven unique objects, across three object types involved with the same event. For every isolated object, this is considered to be a new event because we are flattening the object life cycles. This illustrates how easy it is to lose track of what's actually happening within a process using traditional process mining techniques.

Where Fig. 6 shows us the actual fabric of the process, Fig. 7 shows a flattened reality where objects no longer interact. Using a single case notion means focusing on all flattened object life cycles for a given object type, thus causing the convergence and divergence issues.

Object-Centric Process Mining and Celonis Process Sphere are designed to address these problems.

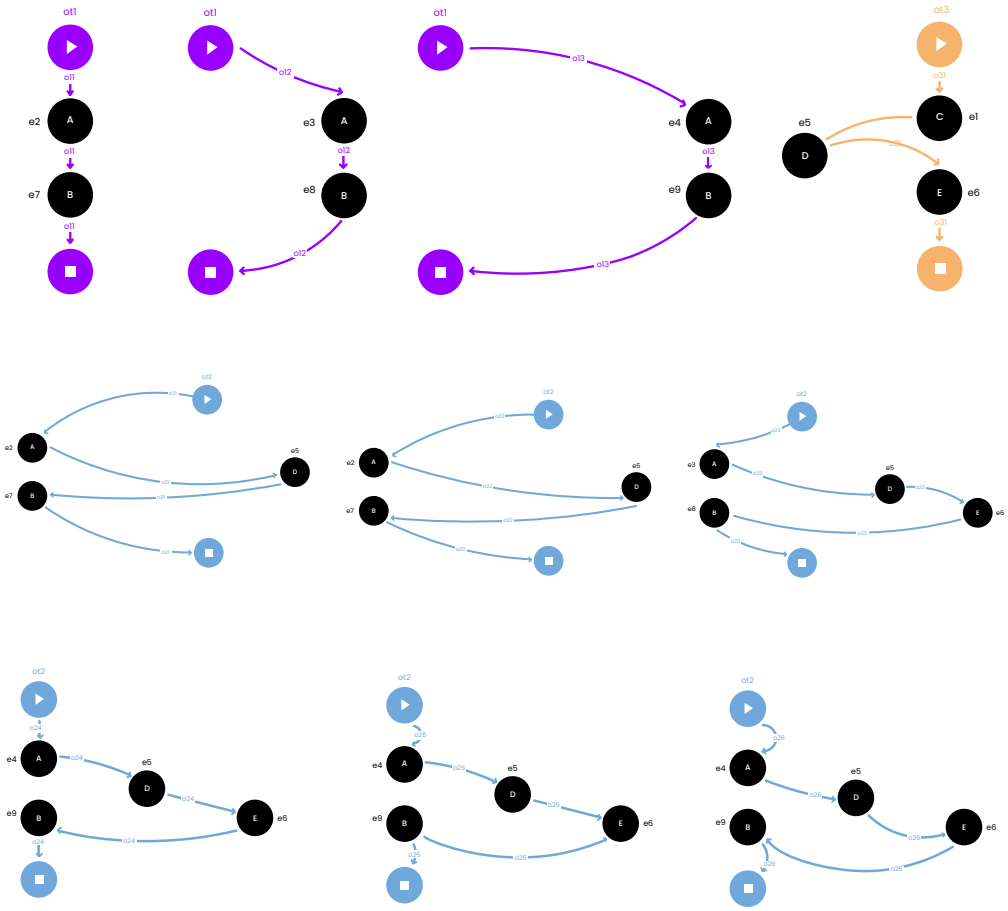


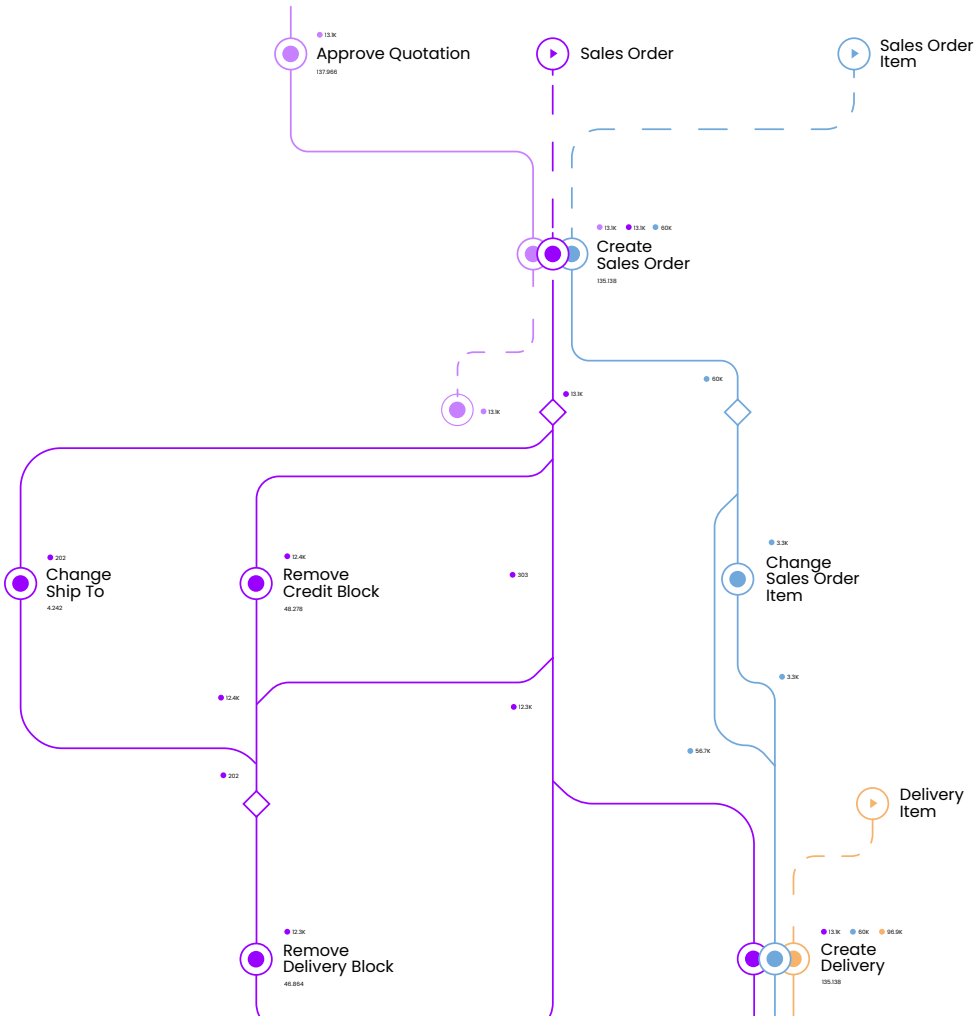
FIG. 7 A flattened representation of reality.

5 Meet Process Sphere: Object-Centric Process Mining in action

With Process Sphere™, Celonis is the first technology vendor to fully embrace Object-Centric Process Mining. Process Sphere overcomes the process mining limitations outlined above and therefore provides a more realistic picture of business operations. The following examples illustrate how Process Sphere visualizes real-world, end-to-end processes that involve multiple objects of different types connected across multiple processes.

Let's look at this in more detail. In Fig. 8, we can see multiple object types in a single diagram. We can select the objects and activities that we want to explore. With this concept, we can look at all objects and events for a comprehensive view of the business. However, what is most beneficial is to restrict the view to the relevant objects and events that are in the scope of analysis.

FIG. 8 Seeing multiple processes and object types simultaneously using Process Sphere.



The benefit here is that you do not need to go back to the data and configure a new event log when the line of questioning changes and requires the perspective of other objects or events. In addition, Process Sphere also shows the actual frequencies of the events without duplication, thus reflecting the most faithful picture of process reality while avoiding convergence and divergence problems. This enables powerful insights for the user. There is no longer a need to deal with the complex nuances of the data that we described in the challenges above because they no longer exist.

Another powerful capability of object-centric process mining and Process Sphere is the ability to select two event types (i.e., activities) in a process and measure the average throughput time between them (Fig. 9)

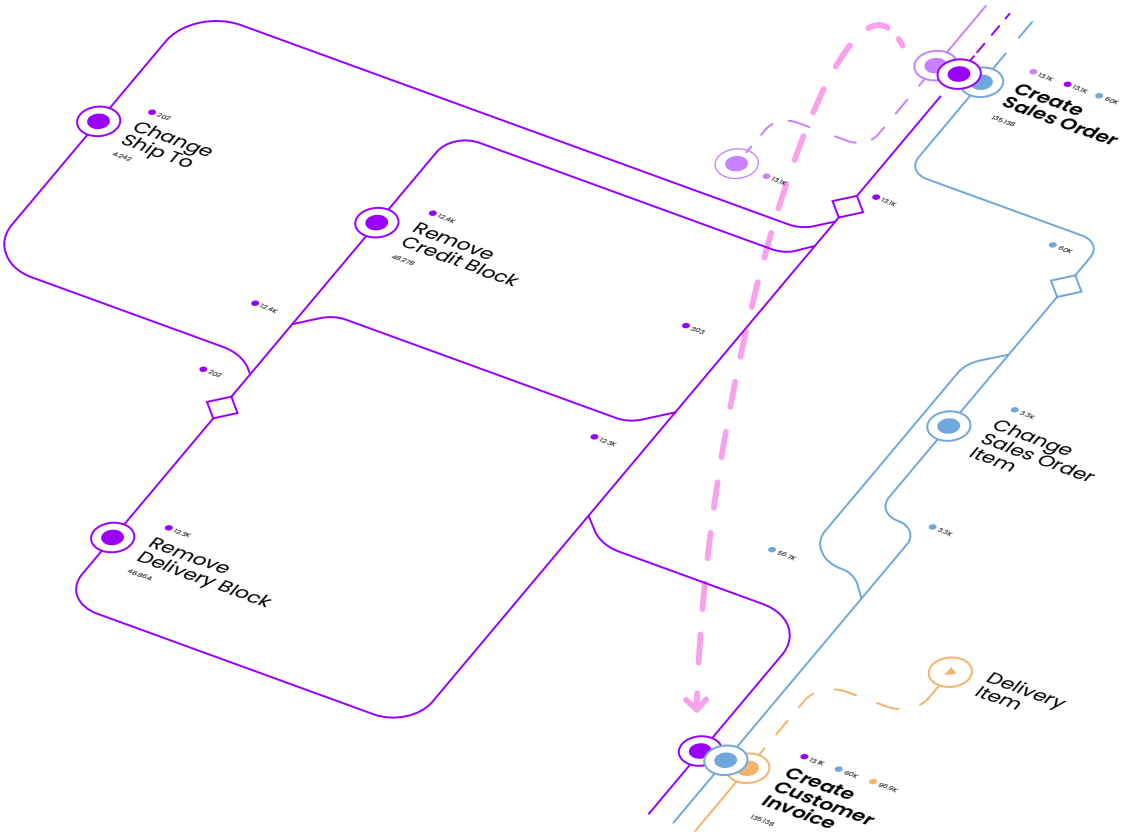


FIG. 9 Identifying the throughput time between multiple objects and events in Process Sphere.

Note that the time between two events is always with respect to an object or related objects. When we highlight the activity, we can also see which objects were involved in the execution of the section of the process that was selected. Each arc corresponds to an object type. The color makes it easy to identify the corresponding type and the frequency indicated reveals how often objects of the corresponding type moved from one activity to the next. This provides powerful insights to quickly determine the opportunities for improvement that are located at the intersection points of different business processes.

This type of ad-hoc process exploration promises to change the game for the process mining industry. The process analyst will be able to get to these insights much faster. This will also help the organization easily solve the up and down-stream 'butterfly effects' that hinder a company from achieving its top, bottom, and green line goals. Process Sphere will also serve as a powerful galvanizing canvas to unite stakeholders across lines of business for greater process centricity to deliver the most desirable outcomes for the company.

So not only will someone in Order Management be able to optimize their processes more quickly and simply, they will be able to see how their actions might affect those further downstream — their friends in Accounts Receivable, for instance — and how the actions of those upstream, like Procurement, affect them.

It means that a CEO will be able to see how pulling a lever to improve something in one department could have a knock-on effect in another department that causes a net-negative result for the business.

Going beyond that, it allows your sustainability lead to prove how initiatives like reducing emissions by optimizing the number of trucks on the road will positively affect the bottom line and finally get sign-off to go ahead.

The applications are endless, and the potential impact is vast.

6 Conclusion

We began by explaining what process mining is and its current limitations. With traditional process mining, every event needs to be linked to a single object type (a Case ID) and a new extraction and data configuration must be made each time a change of view is needed. This leads to the unintentional duplication of events (convergence) and the loss of causal information (divergence). Moreover, by focusing on a single object type, the interaction with other object types in the process is missed and the statistics that are produced may not reflect the 'truest' reality. The reason for this is that process realities have events that involve multiple objects and by flattening event data, we are squeezing the 3-dimensional reality into 2-dimensional event log process models.

Object-centric process mining addresses these challenges by using representations of data and processes that reflect reality. Process Sphere fully implements these ideas using powerful visualizations. As a result, we can realize the following advantages:

1 — **Data configuration is only done once.** After the data is loaded, the user chooses the view of objects and events that they want to look at. This is beneficial as users often describe processes in their unique ways based on their departmental understanding, or the problem they are trying to solve. For the first time, it is possible to do process mining queries in a truly flexible fashion without needing a significant data configuration project for every new question or process.

2 — **Interactions between objects are captured.** This changes the game in terms of the time to value of process mining projects. Tremendous efficiencies are gained with object-event capture without needing to create new models. This is immensely helpful in determining the 'butterfly effect' root causes in processes up or downstream very quickly, aligning departments to better work together, and it brings us closer to closing the loop between insights and action at the real speed of business.

3 — **3-dimensional process mining data models to better represent reality.** The models are 3-dimensional and easily extensible to add more objects and events to capture more use cases for fast process mining. Furthermore, object-centric process models represent the truest picture of what is happening in the business and overcome the challenges of divergence and convergence that were described before. Process Sphere offers powerful interactive visualizations that are easy to interpret.

For the first time, process mining is experiencing a renaissance that will amplify the transformational business value that traditional process mining has brought to practitioners. This is significant considering the hyper-complexity of business systems, processes and the economic climate of our time.

Process Sphere is just the starting point. Existing process mining capabilities, ranging from conformance checking to predictive analytics can benefit from object-centric process mining. This means that all preexisting types of process mining can be lifted from 2D to 3D, thus revolutionizing the entire industry.

7 References

W.M.P. van der Aalst. Process Mining: Data Science in Action. Springer-Verlag, 2016.

W.M.P. van der Aalst and A. Berti. Discovering Object-Centric Petri Nets. Fundamenta Informaticae, 175(1-4):1-40, 2020.

L. Reinkemeyer (eds). Process Mining in Action. Springer-Verlag 2020.

PADS. Object-Centric Event Log (OCEL) standard, http://www.ocel-standard.org/, 2020.

W.M.P. van der Aalst and J. Carmona (eds). Process Mining Handbook. Springer-Verlag, Berlin, 2022.

TITLE
Object-Centric Process Mining:
The next frontier in business performance

AUTHOR
Prof. dr. ir. Wil van der Aalst
Chief Scientist, Celonis

Celonis SE
Theresienstraße 680333 Munich

celonis