



AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN

Informe final publicable de proyecto

Minería de procesos y datos para la mejora de procesos colaborativos aplicada a e-Government

Código de proyecto ANII: FMV_1_2021_1_167483

Fecha de cierre de proyecto: 01/11/2024

DELGADO CAVALIERE, Andrea (Responsable Técnico - Científico)

CALEGARI GARCÍA, Daniel (Investigador)

GONZÁLEZ LÓPEZ, María Laura (Investigador)

MAROTTA LLAMBÍ, Adriana Alejandra (Investigador)

TANSINI MERCADER, Libertad (Investigador)

ANDRADE ACOSTA, Daniela Andrea (Investigador)

UNIVERSIDAD DE LA REPÚBLICA. FACULTAD DE INGENIERÍA (Institución Proponente) \\ UNIVERSIDAD DE LA REPÚBLICA. FACULTAD DE INGENIERÍA

Resumen del proyecto

Los procesos de negocio son un conjunto de actividades realizadas en coordinación en un entorno organizacional y técnico para alcanzar un objetivo de negocio. Estos procesos raramente están explicitados en modelos que provean una guía para su ejecución y mayoritariamente se encuentran implícitos en sistemas de información organizacionales tradicionales. Si bien en sistemas basados en procesos estos modelos están presentes, la ejecución real de los procesos puede incluir desviaciones, excepciones y/o la necesidad de nuevos caminos o condiciones. La aplicación de técnicas de minería de procesos y datos para el descubrimiento de modelos de procesos, predicción y análisis de su ejecución, provee a las organizaciones con información valiosa sobre eficiencia, calidad o cumplimiento de normativas, para descubrir oportunidades de mejora basada en evidencia.

Un proceso colaborativo involucra dos o más organizaciones que actúan coordinadamente para llevar adelante distintas sub-partes del proceso que hacen al proceso general. Los procesos colaborativos, a

diferencia de los de tipo orquestación que se realizan en una única organización, agregan diversos desafíos tanto para su implementación como para el análisis de su ejecución. En particular, las trazas de ejecución de los procesos colaborativos se encuentran diseminadas en varias organizaciones, con modelos de datos y tecnologías heterogéneas, lo que complejiza la recolección de datos asociados para la aplicación de minería de procesos.

El objetivo principal de este proyecto es definir y evaluar técnicas, algoritmos y estrategias de minería de procesos y datos para el descubrimiento de modelos de procesos colaborativos, predicción y análisis de su ejecución, a los efectos de identificar oportunidades de mejora. En particular, se propone el estudio de procesos colaborativos de Gobierno Electrónico (e-Government), vinculando consistentemente las definiciones con resultados previos relacionados al entorno integrado para la aplicación de ciencia de datos organizacional denominado PRICED (Process and Data sCience for oRganizational improvEment).

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Gestión de Procesos de Negocio, Minería de Procesos y datos

Palabras clave: Minería de procesos y datos / Procesos colaborativos / Gobierno Electrónico /

Antecedentes, problema de investigación, objetivos y justificación.

Antecedentes

Un proceso de negocio define el conjunto de actividades que se realizan en coordinación en un entorno organizacional y técnico, para alcanzar un objetivo del negocio [WES19]. La Gestión de Procesos de Negocio (Business Process Management, BPM) soporta el ciclo de vida de los procesos de negocio desde su modelado, diseño, configuración y ejecución hasta su evaluación y mejora [WES19, DUM18, vAHW03]. Un problema que enfrentan las organizaciones cuando intentan basar su operativa en los procesos que realizan, es el poco conocimiento real que se tiene sobre estos. Estos procesos raramente están explicitados en modelos que provean una guía para su ejecución, por ejemplo utilizando el lenguaje estándar BPMN [OMG11] y mayoritariamente se encuentran implícitos en sistemas de información tradicionales. Si bien en el caso de Sistemas de Información basados en Procesos (Process Aware Information Systems, PAIS) [DUMA05] y plataformas BPMS (Business Process Management Systems) [BPMS06] estos modelos están presentes, la ejecución real de los procesos puede incluir desviaciones, excepciones y/o la necesidad de nuevos caminos o condiciones. Para sustentar la mejora de procesos basada en evidencia, se debe poder acceder a los datos de su ejecución real de forma de analizarlos y contrastarlos con los objetivos y funcionamiento que tiene definido la organización. A partir de estos datos mediante el uso de técnicas de minería de procesos (process mining) [vaHW16] se pueden realizar análisis complejos de la ejecución real de los procesos, proveyendo a las organizaciones con información valiosa sobre eficiencia, calidad o cumplimiento de normativas, para descubrir oportunidades de mejora basada en evidencia.

La minería de procesos [vaHW16] se ha desarrollado en las últimas décadas, y provee técnicas y algoritmos para analizar la ejecución de procesos de negocio. En primer lugar, el descubrimiento de modelos de procesos desde los datos de ejecución como base para discusión con los involucrados y mejora. Ejemplos de algoritmos de descubrimiento son: alpha, inductive, heuristic, fuzzy, BPMN miner [ACDLM19]. En segundo lugar el chequeo de conformidad de modelos permite comparar los logs de eventos con modelos descubiertos o descriptivos, para contrastar la operativa real con su definición. Se utilizan medidas como el fitness (similar recall), precisión, overfitting y underfitting para evaluar la calidad de los modelos con respecto al log [CvDSW18]. Finalmente, es posible extender los modelos obteniendo medidas clave de la ejecución como: cuellos de botella (bottlenecks), duración de instancias particulares, participantes, roles y relaciones, sub utilización de recursos, etc.

Este tipo de análisis post mortem es muy útil para las organizaciones sentando la base para la mejora de los procesos basada en evidencia, aportando elementos para la toma de decisiones con base en la operativa diaria. Por otro lado, la posibilidad de explotar los datos de ejecución de instancias pasadas para predecir la ejecución de las instancias actuales, está siendo cada vez más requerida por las organizaciones. Las técnicas de minería de datos permiten encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto [SUMA06]. Los objetivos principales son la predicción de tendencias o comportamientos y el descubrimiento de patrones y modelos previamente desconocidos [HERN4]. Las propuestas para el análisis predictivo de la ejecución de procesos de negocio se han incrementado en la última década [FGMM18], siendo la predicción del tiempo restante de ejecución de una instancia o el próximo evento en su ejecución de los más atacados.

Tanto la aplicación de minería de procesos (descubrimiento, conformidad, extensión) como de minería de datos (descripción, predicción) sobre los datos de ejecución de procesos de negocio se han enfocado principalmente en procesos de tipo orquestación [vaHW11] que se realizan en una única organización (intra-organizacionales), y recientemente a procesos colaborativos [CHQTC20] pero no en BPMN. Esto determina que las trazas de ejecución del proceso se encuentran en un entorno controlado y accesible, aún siendo distribuidos. Por el contrario, un proceso colaborativo [vaHW11] involucra dos o más organizaciones que actúan coordinadamente para llevar adelante distintas sub-partes del proceso del proceso general (inter-organizacionales). Los procesos colaborativos, a diferencia de los de

tipo orquestación, agregan diversos desafíos tanto para su implementación como para su descubrimiento, predicción y análisis de su ejecución. En particular, las trazas de ejecución de los procesos colaborativos se encuentran diseminadas en varias organizaciones, con modelos de datos y tecnologías heterogéneas, lo que complejiza la recolección de datos para la aplicación de minería de procesos, y su integración. Adicionalmente, los modelos descubiertos por las implementaciones existentes, no incluyen en su visualización los participantes o roles asociados (incluso en BPMN), los que se tratan aparte en la extensión del modelo. Esto limita la comprensión y utilidad de los modelos descubiertos.

El e-Government es un campo de aplicación interesante ya que los procesos de e-Government suelen ser colaborativos. La plataforma de Gobierno Electrónico (PGE) de Uruguay brinda una plataforma de interoperabilidad que permite a las organizaciones gubernamentales llevar a cabo procesos colaborativos comunicándose entre sí y con los ciudadanos involucrados [DGC17]. La plataforma ofrece soporte para exponer e invocar servicios para interacción de organizaciones, registrando la coreografía de dichos procesos. También se proporciona un sistema de trazabilidad para registrar las actividades de cada organización dentro de los procesos colaborativos. Esta información se utiliza, por ejemplo, para brindar a ciudadanos y organizaciones información en tiempo real sobre la ejecución de procesos. Sin embargo, estos procesos están implícitos y la responsabilidad de rastrearlos se delega en cada organismo. Por tanto, no existe una evaluación adecuada de su ejecución.

Gran parte del equipo de trabajo (Delgado, Calegari, González, Marotta, Tansini, Rubio y Peña) participaron activamente del proyecto I+D "Minería de procesos y datos para la mejora de procesos en las organizaciones" financiado por la Comisión Sectorial de Investigación Científica (CSIC), que sienta las bases para este proyecto. También participaron integrantes de la Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento (AGESIC) proveyendo datos, sin aporte monetario. Como resultado principal del proyecto CSIC previo se definió el framework PRICED (Process and Data sCience for oRganizational improvEment) [DMGTC20] y su metodología [DCMGT21]. El proyecto tuvo diversidad de resultados con publicaciones en congresos latinoamericanos, iberoamericanos e internacionales de primer nivel [RDT21] [DMGTC20] [DCMGT21] [DC20] [ABCD21] [BPMD21] [GD21] con participación de estudiantes de grado y tesista de posgrado. Se desarrollaron herramientas de soporte: (a) automatización de la integración de datos de procesos y organizacionales, carga del metamodelo y matching asociado y generación automática de logs de eventos extendidos, y (b) plug-ins para el framework ProM para evaluación de calidad, de cumplimiento en coreografías, y aplicación integrada de minería de procesos y datos. Los investigadores Reijers y Weber participaron en proyectos previos y en actividades del grupo COAL desde el año 2012 (BPMuy), con los investigadores M. Sepúlveda y J. Muñoz-Gama integramos el simposio SLPNASO (<https://clei2021.cr/simplatpnaso>) de la Conferencia Latinoamericana en Informática (CLEI).

Problema de investigación y justificación

El crecimiento exponencial de la cantidad de datos disponibles en internet y en las organizaciones o el fenómeno de la "explosión de datos" que se viene experimentando en los últimos años, ha puesto en primera plana la aplicación de ciencia de datos para el análisis de datos, permeando de forma universal en diversas áreas de conocimiento. La ciencia de datos [IEEEEDS, vaHW16] surge en los últimos años como una disciplina en sí misma, interdisciplinaria, para brindar respuesta al problema de la gestión, análisis y descubrimiento de información en grandes volúmenes de datos que son generados a gran velocidad y con gran variedad (las tres V) [FURH16], considerando también la veracidad de los datos [ONG16], almacenados en forma estructurada o no estructurada.

Estos datos incluyen los asociados a la ejecución de procesos de negocio y organizacionales, generados por la operativa diaria en las organizaciones tanto desde sistemas de información tradicionales como de Sistemas de Información basados en Procesos (Process Aware Information Systems, PAIS) [DUMA05] y plataformas BPMS (Business Process Management Systems) [BPMS06]. Estos datos se encuentran cada vez más almacenados con modelos de datos y tecnologías heterogéneas, distribuidos en diversos sistemas e infraestructuras, más en el caso de los procesos colaborativos, cuyas trazas se encuentran diseminadas entre varias organizaciones. Para poder analizar la ejecución de los procesos, estos datos deben ser recolectados desde sus fuentes de datos, y transformados a formatos manipulables por las técnicas, algoritmos y herramientas de minería de procesos y datos.

Las técnicas de minería de procesos (process mining) [vaHW16] permiten realizar análisis complejos de la ejecución real de los procesos, proveyendo a las organizaciones con información valiosa sobre eficiencia, calidad o cumplimiento de normativas, para descubrir oportunidades de mejora basada en evidencia. Este tipo de análisis post mortem es muy útil para las organizaciones sentando la base para la mejora de los procesos basada en evidencia, aportando elementos para la toma de decisiones con base en la operativa diaria. Por otro lado, la posibilidad de explotar los datos de ejecución de instancias pasadas para predecir la ejecución de las instancias actuales, está siendo cada vez más requerida por las organizaciones. La minería de datos [SUMA06, [HERN4] permite realizar predicciones sobre la ejecución de los procesos considerando las trazas en el log de eventos como insumo de entrada, para anticipar posibles desviaciones, violaciones y demoras en los procesos, de forma de tomar medidas preventivas (por ej. re-asignación de recursos) [FGMM18]. Además la minería de datos permite analizar los datos integrados de procesos y organizacionales, por ej. caracterizando variantes según tipos de datos organizacionales que manejan.

Tanto la aplicación de minería de procesos (descubrimiento, conformidad, extensión) como de minería de datos (descripción, predicción) sobre los datos de ejecución de procesos de negocio se han enfocado principalmente en procesos de tipo orquestación [vaHW11] que se realizan en una única organización (intra-organizacionales). Los procesos colaborativos (inter-organizacionales), a diferencia de los de tipo orquestación, agregan diversos desafíos tanto para su implementación como para su descubrimiento, predicción y análisis de su ejecución. El foco del proyecto está en proveer soporte al análisis de la ejecución de procesos colaborativos en entornos del tipo de e-Government, donde varias organizaciones interactúan entre sí para la realización de los procesos. El proyecto aportará en la definición y evaluación de técnicas, algoritmos y estrategias de minería de procesos y datos para el descubrimiento de modelos de procesos colaborativos, predicción y análisis de su ejecución, así como herramientas de soporte.

Objetivos

El objetivo general de este proyecto es definir y evaluar técnicas, algoritmos y estrategias de minería de procesos y datos para el descubrimiento de modelos de procesos colaborativos, predicción y análisis de su ejecución, a los efectos de identificar oportunidades de mejora basada en evidencia. En particular, se propone la aplicación a procesos colaborativos de Gobierno Electrónico (e-Government), vinculando consistentemente las definiciones con resultados previos relacionados al entorno integrado para la aplicación de ciencia de datos organizacional PRICED (Process and Data sCIence for oRganIzational improvEment). Estos resultados buscan reducir el esfuerzo de identificación y aplicación integrada de técnicas, algoritmos y herramientas de minería de procesos y datos, ampliando las posibilidades de análisis, evaluación y mejora de los procesos colaborativos en las organizaciones.

(O1). Extender el relevamiento del estado del arte existente en la temática de técnicas, algoritmos y estrategias enfocadas en la minería de procesos colaborativos y datos, incluyendo descubrimiento, predicción y análisis de su ejecución.

(O2). Evaluar, seleccionar técnicas y algoritmos de minería de procesos colaborativos y datos, o definir nuevos adecuados para el descubrimiento de procesos colaborativos y predicción de su ejecución.

(O3). Evaluar y definir estrategias para minería de procesos colaborativos y datos incluyendo recolección, integración, limpieza, análisis y obtención de información de valor asociada a la ejecución de procesos colaborativos

(O4). Ensayar las técnicas, algoritmos y estrategias para la recolección, limpieza, análisis y obtención de información de valor asociada a la ejecución de procesos colaborativos, experimentando sobre conjuntos de datos (logs de eventos) relevantes.

(O5). Integrar/desarrollar herramientas de soporte a las técnicas, algoritmos y estrategias definidas para minería de procesos colaborativos y datos.

(O6). Evaluar la aplicabilidad de la propuesta en casos de estudio de e-Government con conjuntos de datos (logs de eventos) de ejecución real de procesos colaborativos, para validar la propuesta e identificar oportunidades de mejora.

Metodología/Diseño del estudio

La hipótesis de investigación se define como: "es posible aplicar y/o definir nuevas técnicas, algoritmos y estrategias de minería de procesos y datos para el descubrimiento de modelos de procesos colaborativos, predicción y análisis de su ejecución, a los efectos de identificar oportunidades de mejora basada en evidencia."

A partir de esta hipótesis, se derivan el objetivo general del proyecto y preguntas que se relacionan con sus objetivos específicos.

(P1) ¿Qué propuestas existen de técnicas, algoritmos y estrategias de minería de procesos y datos para el descubrimiento de modelos de procesos colaborativos, predicción y análisis de su ejecución? (O1)

(P2) ¿Las propuestas existentes de técnicas, algoritmos y estrategias de minería de procesos y datos para el descubrimiento de modelos de procesos colaborativos, predicción y análisis de su ejecución son adecuadas, o es necesario mejorarlas/ extenderlas/unificarlas en algún sentido? (O2, O3, O4)

(P3) ¿Existe algún vínculo entre las propuestas de técnicas, algoritmos y estrategias de minería de procesos y datos para procesos de tipo orquestación y para procesos colaborativos? (O1, O2, O3)

(P4) ¿Qué soporte tecnológico es necesario integrar/desarrollar para implementar las técnicas, algoritmos y estrategias seleccionadas/definidas en la propuesta? (O5)

(P5) ¿Qué tan aplicables son las técnicas, algoritmos y estrategias seleccionadas/ definidas en la propuesta a procesos colaborativos de e-Government? (O6)

Para llevar adelante el trabajo de investigación en las distintas líneas de trabajo definidas en el grupo COAL, trabajamos con un enfoque de investigación integrado, para el cual identificamos cuatro etapas genéricas a las que asociar diversas estrategias de investigación, las que definen las actividades específicas a llevar adelante. Las etapas son:

1. Identificar hipótesis y preguntas de investigación que guiarán el trabajo a realizar
2. Realizar estado del arte al inicio de la investigación sobre la temática seleccionada
3. Plantear y realizar la investigación sobre la temática seleccionada, definiendo actividades específicas y resultados, generando los avances de conocimiento asociados
4. Validar los resultados obtenidos tanto con publicaciones como mediante estrategias de validación adecuadas a la temática

Cada una de las etapas identificadas son análogas a las etapas de cualquier proyecto en que hay una primera etapa de identificación de necesidades, una etapa siguiente de planificación de tareas, una tercera de ejecución del proyecto y una final de evaluación. Para cada sub-temática asociada a las líneas de investigación definidas, el trabajo de investigación será guiado por las preguntas de investigación definidas asociadas a los objetivos planteados, comenzando con el estado del arte en la temática como elemento inicial y base de la investigación. Luego definimos las actividades específicas a realizar (que están asociadas directamente al contenido de la investigación) y los resultados esperados, ejecutando las tareas definidas. Finalmente, validamos dichos resultados tanto con publicaciones en revistas y congresos de nivel como mediante estrategias de validación asociadas.

En particular, planteamos utilizar las siguientes estrategias de investigación en cada etapa:

a. Para el estado del arte seguiremos los lineamientos de Revisiones sistemáticas de la literatura y mapping studies planteados en [Kit04, KB07]. Es un enfoque formal para realizar relevamientos del estado del arte en la temática definida, que permite asegurar que el resultado obtenido es exhaustivo y repetible. En este proyecto la revisión se realizará sobre propuestas de técnicas, algoritmos y estrategias para descubrimiento y predicción de procesos colaborativos. Se consideran tres etapas principales: planificación de la revisión, desarrollo de la revisión y publicación de los resultados. En la etapa de Planificación se define la pregunta de investigación para el estado del arte, las palabras claves y las cadenas de búsqueda a utilizar para recuperar los trabajos relevantes de las fuentes seleccionadas, así como los criterios de inclusión y exclusión para seleccionar trabajos relevantes y estudios primarios. En el Desarrollo las cadenas de búsqueda definidas son ejecutadas en las fuentes seleccionadas, aplicando criterios de inclusión/exclusión para seleccionar artículos relevantes. Luego de descartar estudios repetidos el resto son leídos en profundidad y si cumplen el criterio de inclusión son seleccionados como estudios primarios. De estos estudios se realiza la extracción de datos para el análisis de resultados. Finalmente la Publicación de los resultados permite la divulgación de los mismos.

b. Para guiar la investigación en cada sub-temática de las líneas de investigación planteadas seguiremos los lineamientos del método de Design Science para Sistemas de Información

[Wer14,HMPR04, Ven06], eventualmente en combinación con Action-Research [Woo85] según [LV09]. En Design-Science se pueden identificar dos procesos principales: construir y evaluar, y la producción de cuatro artefactos clave: construcciones, modelos, métodos e instancias. Se construyen artefactos de utilidad para resolver problemas no resueltos aún, y se evalúan con respecto a su utilidad en la solución de dichos problemas [HMPR04]. Es importante distinguir entre diseño rutinario o construcción de sistemas e investigación Design-Science. En el primer caso, se aplica conocimiento existente a problemas organizacionales utilizando artefactos "buenas prácticas" conocidos (construcciones, modelos, métodos e instancias); en el segundo caso se atacan problemas no resueltos aún, utilizando enfoques únicos e innovadores, o problemas resueltos pero con enfoques más efectivos o eficientes, contribuyendo al conocimiento en fundamentos (teorías, frameworks, construcciones, modelos, métodos, instancias) y metodologías (técnicas de análisis de datos, formalismos, medidas, criterios de validación) [Wer14, HMPR04]. En conjunción con Action-Research el principal elemento es la existencia y trabajo conjunto con un "cliente" organizacional quién participa de la definición del problema y del ciclo de generación de soluciones, pruebas en la organización, validación y mejora.

c. Para realizar la validación de resultados aplicaremos evaluación tanto de tipo "artificial" como "naturalista"[Ven06, LV09] mediante estudios de caso y/o experimentos controlados [Yin17, Woh12]. La validación de resultados requiere el análisis de la utilidad de los artefactos generados (construcciones, modelos, métodos, instancias) en el marco del problema para el cual fueron propuestos. Se podrán realizar experimentos controlados para validar la adecuación de las propuestas generadas en el contexto de usuarios (ej. sobre los modelos de procesos colaborativos descubiertos), estudios de caso para validar la adecuación de la propuesta en el contexto de procesos colaborativos existentes. Adicionalmente todos los resultados serán validados en la comunidad científica correspondiente según la temática, con publicaciones en congresos de nivel regional e internacional, y revistas de primer nivel regional e internacional.

Resultados, análisis y discusión

Los resultados obtenidos en el proyecto son de varios tipos:

En primer lugar en cuanto al contenido de la investigación propuesta, se extendieron técnicas y estrategias definidas en el proyecto previo, para integración de datos de procesos y organizacionales desde fuentes de datos heterogéneas (bases de datos relacionales y no SQL), se aportaron nuevas técnicas y estrategias para el descubrimiento de procesos colaborativos inter-organizacionales y de coreografías de procesos colaborativos, con base en técnicas existentes de descubrimiento para orquestaciones (en particular el algoritmo BPMNminer que devuelve procesos en BPMN), utilizando los logs de eventos extendidos en formato XES que definimos previamente (proyecto anterior), y utilizando también estos logs como entrada se aportaron nuevas técnicas y estrategias para la predicción de procesos colaborativos, con base en técnicas existentes de predicción para orquestaciones. Adicionalmente se comenzó a trabajar también en la línea de sostenibilidad de procesos de negocio, con foco en el consumo energético de los algoritmos de minería de procesos, extendiendo el framework de comparación de algoritmos existente [ACDLM19] con la dimensión de sostenibilidad, y se trabajó en una plataforma de microservicios que integre algoritmos de minería de procesos en un único lugar de acceso. El valor de estos aportes fue confirmado por las publicaciones alcanzadas en congresos latinoamericanos, iberoamericanos e internacionales de primer nivel [DC22], [DGMCGP23], [PADC24a], [PADC24b], [DC24], y extensiones para artículos de revista en preparación [DGMCGP25], [DC25], así como artículos en preparación [RD25], [D25]. Se continuará trabajando en líneas iniciadas en este proyecto sobre sostenibilidad e integración de datos organizacionales y procesos con el nuevo estándar OCED (doctorado), y predicción de compliance para procesos colaborativos (maestría).

En segundo lugar, se desarrollaron diversas herramientas de tipo open-source de soporte a las propuestas definidas, en particular herramientas de soporte al descubrimiento de procesos colaborativos [PDC23] y coreografías de procesos colaborativos [ADC23], predicción de procesos colaborativos [ERDC24], plataforma de microservicios para minería de procesos [PLXD24], que suman a las desarrolladas previamente para el framework PRICED y serán integradas al mismo.

En tercer lugar se aportó a la formación de estudiantes de las carreras de Computación (Ingeniería, Licenciatura), y se continuará aportando en posgrado en curso (maestría, doctorado), así como al ingreso de becarios y técnicos que trabajaron en el proyecto al grupo COAL como docentes contratados del Instituto de Computación (InCo) grado 1 (Daniela Andrade) y grado 2 (Martín Rubio). Se aportó también a la formación de profesionales y estudiantes de grado y posgrado con la realización de cursos de posgrado y educación permanente y módulo taller, dictados por los profesores visitantes participantes del proyecto en la 1era Escuela Latinoamericana de Minería de Procesos (LASPM) <https://www.fing.edu.uy/inco/eventos/bpmuy/2022/index.html>, así como el dictado de cursos de minería de procesos como optativa para la carrera de Ingeniería en Computación y para posgrados académico y profesional y educación permanente en dicho marco.

Finalmente, se continúan consolidando las colaboraciones con las universidades y grupos de investigación de los profesores visitantes, tanto de Latinoamérica como de Europa, que dieron lugar a nuevos trabajos incluso en otras áreas dentro de la disciplina de procesos de negocio, en las que seguimos trabajando en conjunto. Adicionalmente, con la realización de la 1era Escuela Latinoamericana de Minería de Procesos (LASPM) y trabajos en curso en procesos colaborativos, se abrieron varias puertas con la comunidad internacional, participando como co-chair en el 2nd Workshop on Collaboration Mining for Distributed Systems (CoMinDS 2023) <https://pros.unicam.it/cominds/2023/index.html> en la 5th International Conference on Process Mining (ICPM), 23-27 October, 2023 (<https://icpmconference.org/2023/>), y como co-chair del workshops track (<https://icpmconference.org/2024/organizing-committee/>) en la 6th International Conference on Process Mining (ICPM), 14-18 October, 2024 (<https://icpmconference.org/2024/>). Finalmente un resultado de alto impacto para el país y para toda Latinoamérica es el otorgamiento de la sede de la 7th International Conference on Process Mining (ICPM) que tendrá lugar del 20 al 24 de Octubre de 2025 (<https://icpmconference.org/2025/>) en la Facultad de Ingeniería (FING), Universidad de la República (UdelaR), Montevideo, Uruguay, organizada por el grupo COAL (<https://icpmconference.org/2025/organizing-committee/>).

Esto se suma al fortalecimiento de la colaboración con AGESIC que ha participado activamente del proyecto, lo que permite abrir nuevas puertas colaboración con organismos públicos y privados y la aplicación de los resultados obtenidos en problemas de relevancia para nuestro país, así como esperamos para ICPM 2025 que se sumen organismos públicos y de la industria de software, tanto como apoyo a la conferencia como con participación de estudiantes, investigadores y profesionales del medio en las sesiones de la conferencia.

Conclusiones y recomendaciones

Como conclusiones más importantes podemos mencionar que en estos últimos años la disciplina de minería de procesos ha evolucionado considerablemente tanto en técnicas y algoritmos que continúan aportando en las distintas perspectivas del análisis de procesos, como en integración de técnicas de inteligencia artificial en particular aprendizaje automático e IA generativa y en herramientas de soporte cada vez más maduras que ofrecen estas técnicas a los usuarios en distinto formato (stand alone, web, librerías java, python, R, etc). Sin embargo, el foco principal de estos esfuerzos continúan siendo los procesos de tipo orquestación que se visualizan sin participantes ni la interacción entre éstos. En este sentido, si bien en la revisión sistemática de la literatura sobre descubrimiento de procesos encontramos unas pocas propuestas (6) y solo dos de ellas en BPMN 2.0, creemos que el aporte que hacemos con nuestro enfoque de descubrimiento de procesos colaborativos y coreografías es significativo para esta línea de trabajo, que viene creciendo en los últimos tres años. Nuestra propuesta toma como entrada el log de eventos extendido definido en un proyecto previo CSIC, que incluye además de los datos de ejecución del proceso colaborativo, datos sobre los participantes y los mensajes intercambiados, y utilizando algoritmos existentes de descubrimiento para cada participante, devuelve un modelo de procesos colaborativo en BPMN 2.0 y la coreografía asociada en BPMN 2.0 también. El código fuente es abierto y se encuentra disponible en el repositorio open-coal en gitlab de FING, UdelaR así como la evaluación comparativa realizada en base a logs colaborativos disponibles de una propuesta relacionada.

En cuanto a la predicción de procesos colaborativos, es una línea en la que no encontramos aún ninguna propuesta, ya que el foco del monitoreo predictivo de procesos está en los procesos de tipo orquestación. Creemos que nuestro aporte en esta línea también puede ser significativo, ya que incluye una conceptualización de los elementos que es posible agregar a la predicción de procesos colaborativos, así como un prototipo desarrollado con base en una librería existente para predicción tradicional, utilizando también como entrada el log de eventos extendido con datos de participantes y mensajes intercambiados, es posible adaptar la entrada a la librería para obtener resultados de predicción asociado a participantes, sus actividades, sus tiempos, etc. El código fuente es abierto y se encuentra disponible en el repositorio open-coal en gitlab de FING, UdelaR. En este sentido tenemos previsto enviar un artículo de revista con la propuesta y el prototipo desarrollado, esperamos pueda ser publicado el próximo año.

Sobre la integración de datos de fuentes heterogéneas, al inicio del proyecto extendimos la propuesta del proyecto previo en este sentido, agregando fuentes NoSQL y proveyendo soporte a la extracción mediante una API genérica que permite conexión con cualquier fuente que implemente las operaciones definidas de extracción e integración. El código fuente es abierto y se encuentra disponible en el repositorio open-coal en gitlab de FING, UdelaR. Si bien la propuesta aporta a la integración de datos organizacionales y de procesos en el log de eventos extendido, en la comunidad se comenzó a definir un nuevo estándar de log de eventos Object-Centric Event Data (OCED), en la misma dirección (permitir integración más flexible de datos organizacionales al log de eventos). Si bien el grupo trabajó en el nuevo estándar esta nueva línea quedó por fuera del proyecto, ya que la propuesta de integración de datos fue cerrada al inicio del mismo.

En cuanto a la extensión de definiciones y aplicación de modelos definidos en proyecto previo de calidad de datos integrados y compliance de procesos, se inició un proyecto para predicción de desvíos de cumplimiento en procesos colaborativos aplicando técnicas de aprendizaje automático, que finalmente no siguió, pero se está trabajando el tema en una tesis de maestría de Ciencia de Datos y Aprendizaje Automático. El aporte creemos será importante ya que de la misma forma que se conceptualizó la predicción para procesos colaborativos, esperamos conceptualizar la predicción de desvíos de cumplimiento con base en el modelo de compliance definido previamente, agregando elementos específicos para procesos colaborativos.

Se trabajó en una plataforma de microservicios para ofrecer técnicas y algoritmos de minería de procesos en un único punto de acceso, una interfaz web que permite elegir distintas implementaciones (java, python, R) tanto de descubrimiento como de conformidad y extensión de modelos, que se agregó al framework definido en el proyecto previo PRICED. Si bien es un prototipo, y se requiere continuar el trabajo para madurar la herramienta, el aporte conceptual de la propuesta y la arquitectura definida puede ser un aporte importante para la adopción de minería de procesos sin tener que depender de una sola implementación de un vendedor. El código fuente es abierto y se encuentra disponible en el repositorio open-coal en gitlab de FING, UdelaR.

Adicionalmente, se comenzó a trabajar la línea de minería de procesos y sostenibilidad de procesos, en colaboración con la UCLM, España, evaluando dicha dimensión en los algoritmos de descubrimiento tradicionales, mediante la medición de consumo energético (disco, memoria, CPU) de algoritmos seleccionados en un benchmark existente. Creemos que este aporte fue sustantivo en la evaluación de los algoritmos no solo con las métricas de descubrimiento y evaluación de calidad de los modelos generados, sino también con el impacto ambiental que conlleva su ejecución. Este año, en el marco de ICPM 2024 se presentó el primer workshop al respecto de minería de procesos y sostenibilidad. Estamos trabajando en una extensión para revista con nuevas ejecuciones y análisis de parámetros optimizados para algoritmos que lo permiten. Adicionalmente seguiremos trabajando esta línea agregando elementos a ser considerados en el registro de procesos para análisis de sostenibilidad, incluyendo una pasantía de doctorado en UCLM, España.

En cuanto a los datos reales de ejecución de trámites provistos por AGESIC, creemos que los aportes han sido sustanciales para la mejora de los registros de los procesos asociados y el análisis de los procesos, incluyendo los colaborativos como residencia legal y certificado de antecedentes judiciales, y esperamos contar en el futuro con nuevos datos del sistema mejorado de trazabilidad para continuar aplicando los avances logrados en el proyecto y evaluando su aplicación en contexto de ejecuciones reales de organizaciones participantes. Adicionalmente estamos actualmente complementando con los datos de la PGE para las coreografías asociadas a los registros de trazabilidad, lo que proveerá una visión más completa al aplicar también el descubrimiento de coreografías definido.

La línea de procesos colaborativos está cobrando interés en la comunidad, con nuevas colaboraciones con un grupo de Italia (Universidad de Camerino) en el workshop de ICPM de procesos colaborativos desde 2023, y varios referentes de la comunidad (en particular en Alemania) que han manifestado interés y están trabajando en el tema, por lo que el futuro de la línea de investigación es prometedor, incluyendo varias posibilidades de colaboraciones. Podemos concluir que tanto la temática de procesos colaborativos como las distintas líneas que venimos trabajando en el proyecto, son de interés en la comunidad internacional y nuestros aportes han sido reconocidos con publicaciones en conferencias y revistas del área, y otras en escritura que esperamos publicar en el próximo año.

Productos derivados del proyecto

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
Publicación de trabajo en evento (artículo de conferencia)	Process and Organizational Data Integration from BPMS and Relational/NoSQL Sources for Process Mining	Andrea Delgado, Daniel Calegari	10.5220/0011322500003266	https://hdl.handle.net/20.500.12381/3700	Finalizado
Publicación de trabajo en evento (artículo de conferencia)	Adding the Sustainability Dimension in Process Mining Discovery Algorithms Evaluation	Andrea Delgado, Félix García, María de los Ángeles Moraga, Daniel Calegari, Alberto Gordillo, Leonel Peña	https://doi.org/10.1007/978-3-031-41623-1_10	https://hdl.handle.net/20.500.12381/3701	Finalizado
Artículo científico	Inter-organizational collaborative BPMN 2.0 business process discovery	Leonel Peña, Daniela Andrade, Andrea Delgado, Daniel Calegari	https://doi.org/10.1007/s10844-024-00876-7	https://hdl.handle.net/20.500.12381/3703	Finalizado
Publicación de trabajo en evento (artículo de conferencia)	An Approach for Discovering Inter-organizational Collaborative Business Processes in BPMN 2.0	Leonel Peña, Daniela Andrade, Andrea Delgado, Daniel Calegari	https://doi.org/10.1007/978-3-031-56107-8_37	https://hdl.handle.net/20.500.12381/3702	Finalizado
Tesis de grado/monografías	Minería de coreografías de procesos colaborativos en Gobierno Digital (e-Government)	Daniela Andrade, Andrea Delgado, Daniel Calegari	https://hdl.handle.net/20.500.12008/42546	https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/42546	Finalizado
Presentación en evento	Extending predictive process monitoring for collaborative processes	Daniel Calegari, Andrea Delgado	https://doi.org/10.48550/arXiv.2409.09212	https://hdl.handle.net/20.500.12381/3704	Finalizado
Tesis de grado/monografías	Integración de datos desde BPMS y fuentes NoSQL para minería de procesos de negocio	Santiago Sosa, Germán González, Andrea Delgado, Daniel Calegari	https://hdl.handle.net/20.500.12008/35408	https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/35408	Finalizado
Tesis de	Integración de	Álvaro	https://hdl.handle.net/20.500.12008/37989	https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/37989	Finalizado

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
grado/monografías	datos de procesos de negocio y organizacionales de bases de datos NoSQL	Vallvé, Andrea Delgado			
Tesis de grado/monografías	Plataforma de microservicios para minería de procesos de negocio	Nicolás Luraschi, Gonzalo Núñez, Damián Piccini, Andrea Delgado			En proceso
Tesis de grado/monografías	Predicción de la ejecución de procesos de negocio colaborativos inter-organizacionales	Carolina Espino, Nicolás Ribero, Andrea Delgado, Daniel Calegari			En proceso
Tesis de maestría	Predicción de la ejecución de coreografías de procesos colaborativos con minería de procesos y aprendizaje automático	Fabián Gambetta, Andrea Delgado			En proceso
Tesis de doctorado	Inteligencia artificial aplicada a la minería de procesos con integración de datos para el análisis y predicción de la ejecución de procesos colaborativos	Martín Rubio, Andrea Delgado			En proceso
Tesis de grado/monografías	Predicción de desvíos de cumplimiento en la ejecución de procesos de negocio colaborativos inter-organizacionales	Romina Parada, Damián Madeira, Federico Galiano, Andrea Delgado, Laura González			Finalizado
Otro	1era. Escuela Latinoamericana de Minería de Procesos (LASPM)	Andrea Delgado, Daniel Calegari, Barbara Weber, Hajo Reijers, Marcos Sepúlveda,	https://www.fing.edu.uy/inco/eventos/bpmuy/2022/index.html		Finalizado

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
		Jorge Munoz- Gama			
Otro	7th International Conference on Process Mining (ICPM) 2025, 20 al 24 de Octubre 2025 Montevideo, Uruguay	Andrea Delgado, Daniel Calegari, Martín Rubio, Leonel Peña, Daniela Andrade, Matías Esmoris, Fabián Gambetta	https://icpmconference.org/2025/		En proceso
Software	An Approach for Discovering Inter-organizational Collaborative Business Processes in BPMN 2.0	Leonel Peña, Andrea Delgado, Daniel Calegari	https://gitlab.fing.edu.uy/open-coal/bpmncollaborativepm		Finalizado
Tesis de grado/monografías	Minería de coreografías de procesos colaborativos en la Plataforma de Gobierno Electrónico (PGE) para mejora de los trámites del estado	Lucía Antunes, Andrea Delgado, Laura González			En proceso
Software	Process Mining on event logs from collaborative business processes for a choreography BPMN 2.0 model.	Daniela Andrade, Andrea Delgado, Daniel Calegari	https://gitlab.fing.edu.uy/open-coal/bpmnchoreographypm		Finalizado
Software	Predicción de la ejecución de procesos de negocio colaborativos inter-organizacionales	Carolina Espino, Nicolás Ribero, Andrea Delgado, Daniel Calegari			Finalizado
Artículo de divulgación	Minería de procesos y datos para la mejora de procesos colaborativos aplicada a e-Government	Andrea Delgado, Daniel Calegari, Leonel Peña, Daniela Andrade	https://idm.uy/2023/		Finalizado
Artículo de divulgación	Minería de procesos para	Leonel Peña,	https://www.fing.edu.uy/inco/eventos/jornadas-uruguayas-ciencias-computacion-2024/		En proceso

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
	descubrimiento y análisis de procesos colaborativos con aplicación en trámites de e-Government	Daniela Andrade, Andrea Delgado, Daniel Calegari			

Referencias bibliográficas

- [ABCD21] A. Artus, A. Borges, D. Calegari, A. Delgado, Integrated process data and organizational data analysis for BP improvement, DaWaK2021 Springer, 2021.
- [ACDLM19] A. Augusto, R.Conforti, M.Dumas, M.La Rosa, F.Maria Maggi, A.Marrella, M.Mecella, A.Soo, Automated Discovery of Process Models from Event Logs: Review and Benchmark, IEEE Transactions Knowledge and Data Eng., v. 31 (4), pp. 686-705, 2019
- [ADC23] Andrade, D., Delgado, A., Calegari, C, Minería de coreografías de procesos colaborativos en Gobierno Digital (e-Government), proyecto de grado Ingeniería en Computación, FING, UdelaR, 2023. Finalizado.
<https://gitlab.fing.edu.uy/open-coal/bpmnchoreographypm>
- [ADG24] Antunes, L., Delgado, A., González, L. Minería de coreografías de procesos colaborativos en la Plataforma de Gobierno Electrónico (PGE) para mejora de los trámites del estado, Proyecto de grado Ingeniería en Computación, FING, UdelaR, 2024. En curso.
- [BPM21] Betancor, F., Pérez, A. Marotta, A. Delgado, BP and organizational data quality model (BPODQM) for integrated process and data mining, QUATIC 2021. Springer, 2021.
- [BPM20] J. Chang, BPM Systems: Strategy and Implementation. CRC Press, 2016
- [CHQC20] Cong L., Huiling L., Qingtian Z., Ting L., Caihong L., "Cross-Organization Emergency Response Process Mining: An Approach Based on Petri Nets", Mathematical Problems in Eng., vol. 2020, 12, 2020.
- [CvDSW18] J. Carmona, B. van Dongen, A. Solti, M. Weidlich. Conformance Checking Relating Processes and Models, Springer, 2018
- [DC20] Delgado, A., Calegari, D., Towards a unified vision of business process and organizational data. XLVI CLEI, pages 108–117. IEEE, 2020
- [DC22] Delgado, A., Calegari, D. Process and Organizational Data Integration from BPMS and Relational/NoSQL Sources for Process Mining, ICSOFT, Lisboa, Portugal, 2022.
- [CD24] Calegari, D., Delgado, A. Extending predictive process monitoring for collaborative processes. COMINDS workshop in ICPM 2024, Copenhagen, Dinamarca, 2024
- [DC25] Delgado, A., Calegari, D., Predictive process monitoring for collaborative processes: concepts and tool support. Journal writing in progress.
- [DCMG21] Delgado, D. Calegari, A. Marotta, L. González, L. Tansini, A methodology for integrated process and data mining and analysis towards evidence-based processimprovement, ICSOFT 2021
- [DGC17] A. Delgado, L. González, D. Calegari: Towards Setting Up a Collaborative Environment to Support Collaborative BPs and Services with Social Interactions. ICSSOC Workshops 2017: 308-320, 2017
- [DGMCGP23] Delgado, A. García, F. Moraga, M.A., Calegari, D., Gordillo, A. Peña, L. Adding the Sustainability Dimension in Process Mining Discovery Algorithms Evaluation, BPM 2023, The netherlands
- [DGMCGP25] Delgado, A. García, F. Moraga, M.A., Calegari, D., Gordillo, A. Peña, L. Extending the sustainability dimension in Process Mining Discovery Algorithms Evaluation, Journal writing in progress
- [DMGTC20] Delgado, A., Marotta, A., González, L., Tansini, L., Calegari, D., Towards a data science framework integrating process and data mining for organizational improvement. ICSOFT 2020, pages 492–500, ScitePress, 2020
- [DUM18] M. Dumas, M. La Rosa, J. Mendling, H. Reijers, Fundamentals of BPM, 2nd Edition, Springer, 2018
- [DUMA05] M. Dumas, W. M. van der Aalst, and A. H. ter Hofstede, Process-aware Information Systems: Bridging People and Software Through Process Technology, Wiley & Sons, 2005
- [ERDC24] Espino C., Ribero N, Delgado, A., Calegari D., Predicción de la ejecución de procesos de negocio colaborativos inter-organizacionales, proyecto de grado Ingeniería en Computación, FING, UdelaR, 2024. Defensa diciembre 2024.
- [FGMM18] Di Francescomarino C., Ghidini, C., Maggi, F.M., Milani, F., Predictive Process Monitoring methods: which one suits me best? CoRR abs/1804.02422 (2018)
- [FURH16] Furht B., Villanustre F. Introduction to Big Data. Big Data Technologies and Applications. Springer, 2016
- [GD21] L. González, A. Delgado, Towards compliance requirements modeling and evaluation of e-Government inter-organizational collaborative BPs, HICSS 2021 ScholarSpace, pp. 2079–2088, 2021
- [GD24] Gambetta, F., Delgado, A. Predicción de la ejecución de coreografías de procesos colaborativos con minería de procesos y aprendizaje automático. Tesis de maestría CDyAA, FING, UdelaR, 2024. En curso
- [HERN4] J.Hernández, M.J. Ramírez, C.Ferri. Introducción a la minería de datos. Pearson, 2004
- [HMPR04] A. Hevner, S. March, J. Park, S. Ram. Design science in IS. Journal MIS Quarterly 28(1):75--105, 2004
- [IEEEEDS] IEEE Task Force on Data Science and Advanced Analytics, <http://www.dsaa.co/>
- [Kit04] B. Kitchenham. Procedures for Performing Systematic Reviews. SE Group, Dept. CS, Keele University and Empirical SE National ICT, 2004
- [KB07] B. Kitchenham, S. Charters. Guidelines for performing Systematic Literature Reviews in SE, Dept. CS, University of Durham, Keele, 2007
- [LV09] J. Iivari, J. Venable: Action research and design science research - Seemingly similar but decisively dissimilar. ECIS 2009: 1642-1653, 2009
- [OMG11] OMG. Business Process Model And Notation (BPMN) V.2.0, Object Management Group, 2011
- [ONG16] K. Ong, D. De Silva, Y.Ling Boo, E. Hui Lim, F. Bodi, D. Alahakoon, S. Leao, Big Data Applications in Engineering and Science. Big Data Concepts, Theories, and Applications. Springer, 2016
- [PADC24a] Peña, L., Andrade, D., Delgado, A. Calegari, D. An Approach for Discovering Inter-organizational Collaborative Business Processes in BPMN 2.0, COMINDS workshop in ICPM 2023, Rome, Italia, 2023
- [PADC24b] Peña, L., Andrade, D., Delgado, A. Calegari, D., Inter-organizational collaborative BPMN 2.0 business process discovery, Journal of Intelligent Information Systems (JIIS), Collection Best papers workshops ICPM 2023
- [PDC23] Peña, L., Delgado, A. Calegari, D., <https://gitlab.fing.edu.uy/open-coal/bpmncollaborativepm>
- [PLND24] Piccini, D., Luraschi N., Nuñez, G. Delgado, A., Plataforma de micros servicios para minería de procesos de negocio, proyecto de grado Ingeniería en Computación, FING, UdelaR, 2024. Defensa diciembre 2024.
- [RD24] Rubio, M. Delgado, A. Inteligencia artificial aplicada a la minería de procesos con integración de datos para el análisis y predicción de la ejecución de procesos colaborativos, Tesis de doctorado PEDECIBA, FING, UdelaR. 2024. En curso
- [RDT21] Rubio, M., Delgado, A., Tansini, L., Process mining and data mining integration frameworks for evidence-based business intelligence: a systematic review, CibSE 2021, Costa Rica, 2021
- [SGDC22] Sosa, S. González, G., Delgado A., Calegari, D. Integración de datos desde BPMS y fuentes NoSQL para minería de procesos de negocio, Proyecto de grado Ingeniería en Computación, FING, UdelaR, 2022. Finalizado.
- [SUMA06] S. Sumathi, S.N.Sivanandam. Introduction to Data Mining and its Applications. Springer, 2006
- [AHW03] W. van der Aalst, A. Hofstede, M. Weske. BPM: A Survey, Int. Conf. BPM, 2003
- [vaHW11] van der Aalst W.M.P. Intra- and Inter-Organizational Process Mining: Discovering Processes within and between Organizations. The Practice of Enterprise Modeling. PoEM 2011. LBIP, vol 92. Springer, 2011
- [vaHW16] W. M. P. van der Aalst. Process Mining: Data Science in Action, Springer, 2nd. Edition, 2016
- [Ven06] J. Venable. A Framework for Design Science Research Activities, Information Resource Management Association Conf., pp 21--24, 2006
- [VD22] Vallé, A., Delgado, A. Integración de datos de procesos de negocio y organizacionales de bases de datos NoSQL, Tesis licenciatura en Computación, FING, UdelaR, 2022, finalizado.

- [Wer14] R. Wieringa. Design Science Methodology for IS and SE. Springer, 2014
- [WES19] M. Weske. BPM Concepts, Languages, Architectures, 3rd Edition, Springer, 2019
- [Woh12] C. Wohlin, P. Runeson, M. H ̈ost, M. C. Ohlsson, and B. Regnell, Experimentation in SE. Springer, 2012.
- [Woo85] A. Wood-Harper. Research methods in IS: using action research. Research Methods in IS, 1985
- [Yin17] R. K. Yin, Case study research. Design and methods, 6th edition. Sage Publications, 2017.

Licenciamiento

Reconocimiento 4.0 Internacional. (CC BY)

