Convocatoria Colaboración Público-Privada CPP2021-009025

DICYME:
Dynamic Industrial CYberrisk Modelling based on Evidence

Iniciativa Conjunta de:

DENEXUS

Universidad Rey Juan Carlos

ENTREGABLE 4.1:
Propuesta de un plan de integración, validación y evaluación

Coordinadores:

Romy R. Ravines (DeNexus Tech)
Isaac Martín de Diego (Universidad Rey Juan Carlos)
Alberto Fernández Isabel (Universidad Rey Juan Carlos)

MINISTERIO DE CIENCIA E INNOVACIÓN

Financiado por la Unión Europea NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

AGENCIA ESTATAL DE INVESTIGACIÓN

1 Introducción

En este documento se describe como se integrarán los indicadores y modelos diseñados en DICYME con DeRISK, que es el objetivo último del proyecto según se explica en la memoria técnica (ver Figura 1). Por otro lado, con la finalidad de disponer de un producto E2E que sea publicado como web app de DICYME, y proporcione una solución completa de cálculo del impacto financiero del riesgo cibernético, se usará un sistema alternativo a DeRISK. De esta manera se mostrará cómo funcionan las sugerencias derivadas de DICYME en el cálculo del ciber riesgo y el sistema de recomendación sin depender del acceso al código fuente del producto DeRISK. Por otro lado, DeNexus elaborará informes con resultados de DeRISK que incorporen los hallazgos de DICYME.

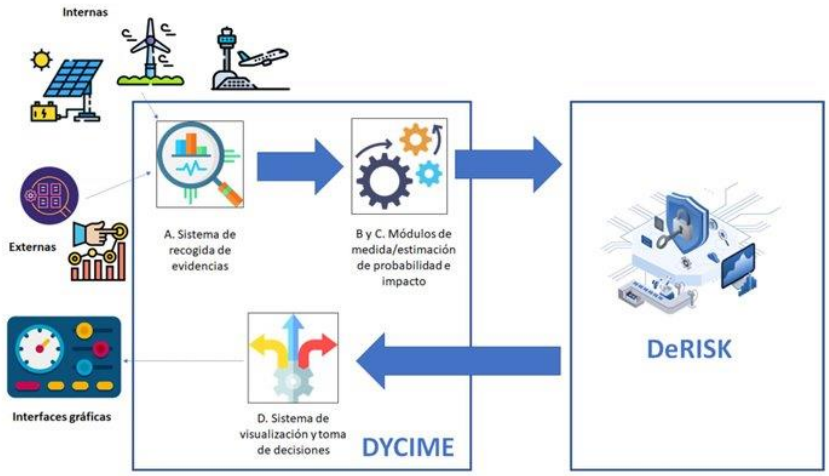

Figura 1. Relación entre el DICYME y el sistema de modelos de DeNexus (DeRISK) según la memoria técnica del Proyecto

2 Marco de Trabajo

Como se está explicando a lo largo del desarrollo del proyecto, en DICYME estamos elaborando indicadores basados en datos que proporcionan información relativa al ciber riesgo. Esos indicadores, algunos de ellos obtenidos como resultado de un sistema de *Machine Learning*, tienen significado propio, pero cobran mayor importancia cuando se usan de forma conjunta en un sistema que cuantifique el impacto financiero que incidentes de ciberseguridad pueden causar en una entidad. Para ello se necesita un sistema que reciba estos indicadores propuestos y los transforme en información para los que toman decisiones respecto al riesgo ciber.

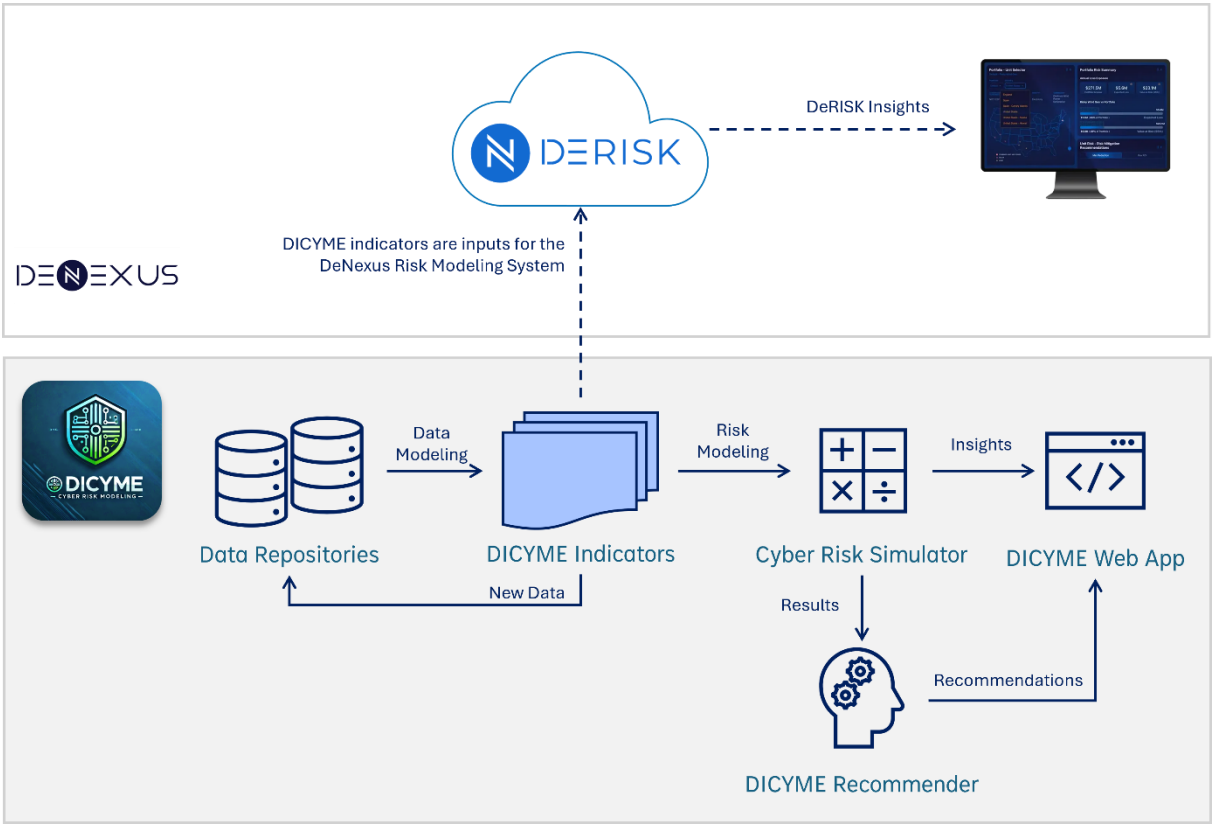

Figura 2. Marco de Trabajo del Proyecto DICYME.

La primera opción, tal como está reflejada en la memoria de DICYME es que DeRISK, el sistema de modelos propiedad de DeNexus, consuma los indicadores para estimar las pérdidas financieras y que su salida sea utilizada por DICYME para formular recomendaciones. Sin embargo, esta opción no es viable sin causar retrasos en el proyecto porque debido a restricciones de confidencialidad y cumplimiento (*Compliance*), el equipo de DICYME no tiene ni tendrá acceso al código fuente de DeRISK. Esta restricción se aplica también a los miembros de DICYME de DeNexus porque al ser personal investigador se limita al desarrollo de pruebas y prototipos y no está autorizado a modificar partes del código del producto como para hacer pruebas con los nuevos indicadores. En consecuencia, el marco de trabajo para la puesta en marcha de los entregables de DICYME, tal como se ilustra en la Figura 2 , tiene dos líneas de acción:

(1) **Integración con DeRISK**. A cargo del equipo DeNexus. Se ejecutará en función de los planes de trabajo de desarrollo de DeRISK, que no dependen ni pueden ser influenciados por DICYME. Sin embargo, antes de la finalización del proyecto, DeNexus informará como se han adaptado los resultados del proyecto y como se comportan los indicadores para los usuarios de DeRISK. Cabe destacar que esta integración requiere de cambios en las bases de datos y algunos procesos de recolección de datos de DeNexus por lo que su ejecución requiere de ingenieros y desarrolladores que no hacen parte de DICYME. DeNexus habilitará dichos recursos cuando sea viable y compatible con los compromisos internos.

(2) **Integración en la aplicación web de DICYME**. Esta línea de trabajo se ha propuesto porque permite tener un *end-to-end* funcional y una experiencia de usuario completa, de todos los indicadores y modelos desarrollados en DICYME. Permitiendo tener una visión integral del ciber riesgo, una estimación aproximada de su impacto financiero y recomendaciones para mitigar su nivel de riesgo. La aplicación web es el Entregable principal de este proyecto. En la Memoria técnica aparece como Interfaz gráfica y herramienta de soporte a la decisión. La novedad de esta línea de trabajo es que dicha herramienta será independiente de DeRISK. El equipo de la URJC dotará la aplicación del proyecto con un motor de cálculo sencillo pero eficaz, basado en la metodología FAIR.

3 El Sistema

Como comentado en la sección anterior, el resultado de DICYME será utilizado por DeNexus, dentro de su producto DeRISK. Sin embargo, esta línea de integración escapa al control y alcance de trabajo del equipo del Proyecto. Por este motivo, de aquí en adelante, el Sistema al que nos referimos corresponde al presentado en la Figura 2, dentro del recuadro gris.

Los principales componentes del Sistema DICYME son:

(1) Repositorio de datos
(2) Indicadores DICYME
(3) Simulador de ciber riesgo
(4) Recomendador DICYME
(5) Aplicación web

⌂ ***Importante***: El Sistema DICYME se encuentra en desarrollo. A la fecha de elaboración de este documento, apenas se dispone de un prototipo de este. Lo que se describe a continuación está sujeto a modificaciones según se vaya concretizando el desarrollo de la aplicación DICYME.

3.1 Repositorio de datos

Una de las principales contribuciones de DICYME es la **recolección automática de datos** con información sobre factores o drivers del ciber riesgo. Entre los desafíos que se han enfrentado está la heterogeneidad de las fuentes de datos y de los datos en sí mismos. Por ello, para cada fuente de dato se ha desarrollado un proceso de recolección específico, que se revisa regularmente y proporciona datos en un formato que se guarda en una base de datos Mongo DB.

A la fecha, junio 2024, los tipos de datos que se consultan y almacenan en los repositorios de datos son:

- Datos sobre ciber incidentes: información sobre incidentes hechos públicos y recopilados en 6 bases de datos accesibles de la web

- Datos sobre víctimas: características de la organización y presencia online
- Datos sobre actores de amenazas: información histórica sobre actores identificados en diferentes bases de datos públicas.
- Datos sobre vulnerabilidades: información detallada sobre vulnerabilidades existentes
- Datos sobre técnicas de ataque: información de la base de conocimiento MITRE ATT&CK.

⌂ ***Importante***: Todos los datos almacenados en el Sistema DICYME son públicos. La mayoría proviene de fuentes de libre acceso. El trabajo relacionado con fuentes de datos internas, y por lo tanto, privadas, se ha realizado con una muestra anonimizada y tratada para preservar la privacidad de los mismos.

La descripción de la estructura de cada base de datos, el volumen de información que almacenan y la frecuencia de actualización se describe en otros documentos del Proyecto. Cabe destacar:

- El código (*scripts*, algoritmos) de recolección y guardado de datos se encuentran en los repositorios del Proyecto, tanto en *Github* (repositorio público) como en *Gitlab* (repositorio privado).
- La infraestructura para estos datos es proporcionada por la URJC y es…..

3.2 Indicadores DICYME

La segunda gran contribución de DICYME es el diseño e implementación de nuevos indicadores relacionados con el ciber riesgo industrial. Este módulo contiene los procesos y resultados de los subsistemas desarrollados. Son subsistemas porque cada uno de ellos tiene un objetivo concreto, repositorio de datos, metodología propia y resultados (diferentes entre subsistemas).

⌂ ***Importante***: La transformación de los datos recolectados de fuentes diversas en indicadores que permitan reflejar de manera dinámica, algunos aspectos del perfil de las organizaciones con relación al riesgo ciber es una de la principales propuesta de valor de DICYME.

A la fecha, en este módulo se tiene:

1. **ATR2ATK**: Estimación del atractivo de una organización a ciber ataques basada en información pública, accesible por cualquier individuo o actor de amenazas. Índice que muestra cómo evoluciona el atractivo de una empresa a ser víctima de un ciber ataque.

2. **CVE2TTs**: Vulnerabilidades y las Técnicas de las matrices MITRE ATT&CK. Identificación de las técnicas que los actores de amenazas podrían explotar con mayor facilidad debido a la existencia de vulnerabilidades.

3. **THRACT:** Los actores de amenazas y las víctimas (organización). Índice que muestra cómo evoluciona un actor de amenazas respecto a un grupo objetivo y/o índice de cómo cambia o evoluciona el conjunto de actores de amenazas respecto a un potencial objetivo.

   En los entregables relacionados con cada subsistema se detallan los datos utilizados, código y resultados que se almacenan en los repositorios y sirven como principales inputs para la medición del ciber riesgo.

   Cabe destacar que cada subsistema tiene su propio método de validación. En general se trata de métodos de evaluación de resultados de la ciencia de datos, combinados con criterios de negocio y/o evaluación de expertos ya que en algunos casos el indicador está representando un concepto, no es producto de un modelo entrenado.

3.3 Simulador de ciber riesgo

Este módulo constituye el motor de cálculo de ciber riesgo que utiliza los datos e indicadores desarrollados en DICYME y sustituye al sistema de modelos DeRISK debido a las restricciones de acceso que se han explicado anteriormente.

Este simulador es una versión adaptada específicamente para DICYME de la metodología FAIR. Se trata de un sistema que simula la frecuencia y el impacto de incidentes relacionados con la ciberseguridad. El método original usa los conocimientos de expertos para especificar las distribuciones de probabilidad que se usan para simular las pérdidas financieras que una organización puede enfrentar en un año. En DICYME algunos de los parámetros de dichas distribuciones son determinados por los indicadores obtenidos para la organización a la que se le evalúa el riesgo.

Como comentado anteriormente, la forma como se integrarán los indicadores con este simulador aún no está definida.

⌂ ***Importante***: DICYME ha desarrollado un simulador de ciber riesgo capaz de aprovechar los datos e indicadores propuestos y servir de motor de cálculo para la evaluación de escenarios y selección de recomendaciones de mitigación del riesgo.

3.4 Recomendador DICYME

A la fecha no se ha trabajado ni en el diseño del mó de recomendaciones de DICYME. Esta sección se desarrollará en los próximos entregables.

3.5 Aplicación web

La aplicación web es el nombre que usamos para el sistema de visualización y herramienta de recomendaciones que se cita en la memoria técnica de DICYME.

Esta aplicación se está desarrollando y será alojada en los servidores de la URJC. En el Entregable E.4.1 se describe el trabajo realizado hasta la fecha.

Como expresado anteriormente, la aplicación web se ha empezado a desarrollar en mayo del 2024, luego apenas se han podido probar la tecnología seleccionada con algunos objetos en estado muy preliminar. Esta sección se desarrollará en los próximos entregables.

4 Funcionamiento del Sistema

El Sistema DICYME aún se encuentra en las primeras etapas de su desarrollo. El prototipo existente es muy incipiente y por lo tanto no se puede describir el funcionamiento del Sistema final. La Figura 3 muestra una idea de alto nivel de como funcionará el sistema, puede interpretarse como un flujo de datos donde en cada paso se ejecutan múltiples procesos que transforman el dato recibido en información de apoyo a la toma de decisiones.

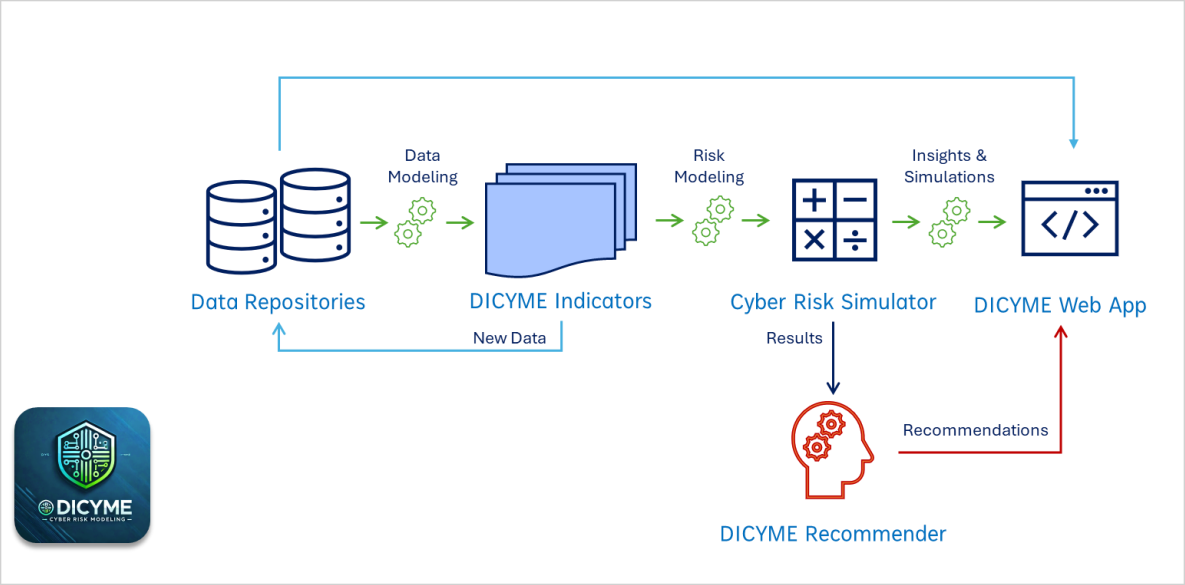

Figura 3. Esquema de Funcionamiento del Sistema DICYME

Algunos aspectos técnicos del sistema

- **Infraestructura**: Servidores de la URJC
- **Bases de datos**: MongoDB
- **Lenguajes**: Python, R, Shiny

Aun queda por definir las funcionalidades que tendrá el sistema. En principio, permitirá al usuario:

- Explorar los datos almacenados en los repositorios
- Calcular los indicadores propuestos por DICYME con datos nuevos

- Visualizar las curvas de pérdidas calculadas
- Visualizar recomendaciones

5 Comentarios Finales

En este documento destacamos asuntos importantes relacionados con el sisyema DICYME:

1. DICYME tendrá su propio motor de cálculo de riesgo para evitar demoras o bloqueos relacionados con el uso del código fuente de DeRISK.
2. El Sistema DICYME se encuentra en pleno diseño y por lo tanto no se puede describir su funcionamiento en detalle
3. Conforme avanza el proyecto, los subsistemas de indicadores van siendo entregados. Cada subsistema tiene su propio sistema de validación y evaluación por lo que los resultados parciales que formarán parte del sistema final ya están validados
4. No existen bases de datos publicas ni privadas con históricos de datos de pérdidas financieras enfrentadas por una misma entidad o grupo de entidades en los últimos años. Por tal motivo, no se pueden usar criterios de bondad de ajuste o precisión para evaluar la calidad de sus resultados. El sistema de simulación debe estar calibrado “por diseño”, de allí la relevancia de que cada subsistema esté debidamente validado.