

ACTA DE REUNIÓN N°3 - AYUDA DE MEMORIA

Ciudad: Bogotá

Lugar: Sesión virtual, plataforma Teams

Tema: Problemáticas que presenta el DANE
en temas de interoperabilidad

Hora: 9:30 a.m. a 11:25 p.m.

Fecha: 06/10/2020

Dependencia responsable: Secretaría
Técnica CASEN (DIRPEN)

Participantes

***Miembros de la Sala Especializada de
Modernización Tecnológica***

León Darío Parra Bernal

Jorge Andrés Gallego Durán

Valérie Gauthier Umaña

***Departamento Administrativo Nacional de
Estadística (DANE)***

Directora Técnica de Regulación, Planeación,
Estandarización y Normalización (DIRPEN) (E)
Elizabeth Moreno Barbosa

Jefe de la Oficina de Sistemas
Andrés Holguín Coral

Coordinadora Grupo Interno de Trabajo (GIT)
Planificación y Articulación Estadística (DIRPEN)
Mónica Patricia Pinzón Torres

Asesor Dirección General
Juan Sebastián Ordoñez

DIRPEN - Dirección Técnica
José Lobo Camargo
Cristian Leonardo Naranjo Puerta

Asesora Dirección DIRPEN
Mayra Ramírez

Profesional GIT Planificación y Articulación
Ruth Constanza Triana Acuña

Profesional GIT Planificación y Articulación
Estadística (DIRPEN)
María del Pilar Pulido

Orden del día

1. Solicitud autorización para grabar la reunión
 2. Verificación del quórum
 - I. Valérie Gauthier Umaña
 - II. León Darío Parra Bernal
 - III. Jorge Andrés Gallego Durán
 - IV. Andrés Holguín Coral
 - V. Juan Sebastián Ordoñez
 3. Resumen de la reunión anterior
 4. Presentación: problemática de interoperabilidad en el DANE
 5. Aportes de los expertos frente a los temas expuestos, de acuerdo a las problemáticas abordadas y el marco de interoperabilidad
 6. Conclusiones y cierre
-

Desarrollo

Objetivo

Dar a conocer las problemáticas actuales que presenta el DANE en relación a la interoperabilidad, con el fin de recibir sugerencias y comentarios por parte de los expertos.

1. Solicitud de autorización para grabar la reunión

Se inició la grabación de la sesión con la autorización de los participantes.

2. Verificación del quórum

Se verificó el quórum tanto de los expertos de la sala, como del Departamento Administrativo Nacional de Estadística (DANE).

3. Resumen de la reunión anterior

Pilar Pulido realizó la síntesis de los temas abordados en la sesión anterior (realizada el día 10 de septiembre), en la cual se revisó el documento del Plan Nacional de Desarrollo VII. "Pacto por la transformación digital de Colombia" y las líneas a trabajar.

Por otro lado, se contextualizó sobre los aspectos generales del Marco de Interoperabilidad para Gobierno Digital y el Marco General de Servicios Ciudadanos Digitales. Adicionalmente, se trabajó sobre las problemáticas presentadas en el DANE en cuanto a las brechas organizacionales, legales y tecnológicas y las dificultades presentadas por la ausencia de un Marco de Interoperabilidad, lineamientos de centralización, proceso estándar para compartir datos y diferentes interpretaciones por parte de las entidades frente a las disposiciones legales.

Finalmente, se recordaron los compromisos adquiridos frente al envío del acta y agendamiento de las siguientes sesiones, las cuales se llevarán a cabo el día martes, 27 de octubre a las 9:00 am, y el 17 de noviembre a las 9:00 am.

4. Presentación: problemática de interoperabilidad en el DANE

Juan Sebastián Ordoñez inició su presentación indicando que trabajará los temas de normativa, interoperabilidad, seguridad, calidad y vinculación de RRAA. Como contextualización, resaltó que la Sala de Modernización Tecnológica tiene como función:

- Asesorar al CAD en temas relacionados con interoperabilidad, calidad, seguridad, gestión de proveedores y lenguaje común
- Asesorar al CAD en la definición de roles y responsabilidades que garanticen el intercambio seguro
- Asesorar en la definición de los principios éticos, así como en lineamientos para la generación de información estadística a partir de la integración estadísticas tradicionales como censos, encuestas, registros administrativos y fuentes no tradicionales
- Asesorar la definición de esquemas para el intercambio de datos entre los miembros del SEN.

Por otro lado, indicó que en la integración (frente a los registros administrativos) se presentó la necesidad de explorar fuentes administrativas para mejorar y complementar la producción actual del DANE; como desafíos en el marco de la pandemia por el COVID-19, la OIT emitió una serie de sugerencias para garantizar el seguimiento al comportamiento del mercado laboral y reducir los posibles sesgos por los cambios en los métodos de captura de la información. En este sentido, el empleo de registros administrativos con propósitos estadísticos tiene grandes ventajas asociadas a la disponibilidad frecuente de la información, la granularidad de los datos, y su costo efectividad al ser recolectadas como parte de un proceso regular.

Enfaticó en la identificación de necesidades, aprovechamiento de la analítica, seguridad y privacidad, aseguramiento de la calidad, gestión de intercambio, almacenamiento y custodia, gestión de proveedores y coordinación y planificación.

Continuó su intervención tomando como ejemplo el trabajo de interoperabilidad realizado con el registro administrativo de relaciones laborales PILA, el cual utiliza integraciones con otros registros con información que PILA no trae, esto, con el fin de sacar indicadores complementarios del mercado laboral. PILA maneja consecutivos sintéticos, amarrado al error de la cédula, generando consecutivos diferentes, perdiendo calidad en la información, generando problemas de estandarización y consistencia. En este intercambio de información había necesidades del DANE que PILA no cumplía, como, por ejemplo, la entrega mensual de información, entonces, se realizó un acercamiento con ellos y, a través de un nuevo convenio, se estableció la periodicidad mensual.

Por otra parte, se socializó la situación en el intercambio de información entre el DANE y el Ministerio de Trabajo (dueños del registro), sin embargo, si en el cargue de datos se presentaba algún error o inconsistencias, el Ministerio solicitaba que no se colocara como fuente de la información PILA porque hacen operaciones, entonces nació el Registro Estadístico de Operaciones Laborales.

Frente a este mismo tema, indicó que el volumen de datos de PILA es bastante grande, más o menos 14 millones, por lo cual, el equipo que se necesitaba para procesar esta información era alto. Además, se necesita aprovechar elementos como la nube, sin embargo, es un desafío porque se deben anonimizar los registros previamente.

En cuanto a los aspectos asociados a la calidad, mencionó que, si se va a interoperar, es necesario corregir valores de las cédulas, esto a fin de asegurar la calidad de los datos y análisis de dichos valores; igualmente, se debe definir quién lo va a hacer, quién es el responsable de asegurar la calidad, si es la entidad o el DANE, o si la Superintendencia debe revisar los protocolos establecidos. También señaló que, si se aborda el tema de seguridad, se tiene en cuenta el manejo de identificadores sintéticos, por lo cual debe definirse cómo hacer para que estas variables o indicadores no se tengan que manipular, y determinar quién es el responsable de manejar las tablas de referencia.

A modo de complemento, indicó que se deberían garantizar accesos segmentados, por ejemplo, si DNP quiere revisar un dato, necesita determinar si necesita toda la información de PILA o solo información puntual, y es acá donde debe establecerse quién tiene acceso a qué.

En este intercambio de información para el aprovechamiento de registros (en donde se vinculan diferentes fuentes), es importante estandarizar el lenguaje entre registros con el fin de mejorar el vínculo entre una fuente y otra.

Juan Sebastián Ordoñez continuó su intervención con la normativa sobre la definición de 'Registros Administrativos' de acuerdo al decreto 2404 de 2019, señalando que hace referencia a un "Conjunto de datos que contiene la información recogida y conservada por entidades y organizaciones en el cumplimiento de sus funciones o competencias misionales u objetos sociales".

De igual forma, se consideran registros administrativos las bases de datos con identificadores únicos asociados a números de identificación personal, números de identificación tributaria u otros, los datos geográficos que permitan identificar o ubicar espacialmente los datos, así como los listados de unidades y transacciones administrados por los miembros del SEN. Respecto al término 'microdato' señaló que correspondía a "los datos sobre las características asociadas a las unidades de observación que se encuentran consolidadas en una base de datos."

Siguiendo con el punto de interoperabilidad, señaló que de acuerdo con MinTic, este término hace referencia a la "capacidad de las organizaciones para intercambiar información y conocimiento en el marco de sus procesos de negocio para interactuar hacia objetivos mutuamente beneficiosos, con el propósito de facilitar la entrega de servicios en línea a ciudadanos, empresas y a otras entidades, mediante el intercambio de datos entre sus sistemas". Lo anterior conlleva a formularse la pregunta sobre cómo hacer el proceso más eficiente.

De acuerdo a lo que se ha hecho en PILA, se tienen unos metadatos, sin embargo, hay aspectos que estos no explican, por lo cual, se deben realizar reuniones para entender el conjunto de metadatos.

Cuando se realiza el análisis de PILA se necesita generar interseccionalidades, porque cada registro genera actividades diferentes, esto puede ser porque se diligencia de manera rápida la planilla y no importa si queda o no bien, entonces, por ejemplo, ¿cómo saber cuál es la actividad principal de la empresa si se presenta este inconveniente?, ¿debería tomarse la actividad que le genera mayores utilidades o cuál?

Se socializó el proyecto piloto de intercambios de información e interoperabilidad de acuerdo con el piloto de análisis de mercado laboral PILA, el cual vincula: lectura de los datos, metadatos, confianza con los proveedores, revisión del sector económico PILA con el Ministerio, tiempos de llegada de datos y procesamiento de información y panorama de interoperabilidad.

Frente al tema de seguridad, se indicó que era necesario trabajar en un entorno de flexibilidad y movilidad organizativa, sin perder la confidencialidad, integridad o disponibilidad de los datos; conocer y comprender el cumplimiento normativo para evitar sanciones y consecuencias penales; conocer y aplicar las coberturas existentes ante el robo de información o ataque a los sistemas empresariales y, custodiar y mantener la confianza de los clientes y la reputación de la organización.

En relación a la calidad, Juan Sebastián Ordoñez señaló que el DANE había desarrollado el plan de fortalecimiento de los registros administrativos con el propósito de aprovechar de forma estadística la información administrativa dispuesta por entidades públicas y privadas. Con este propósito, se elaboró un análisis de diagnóstico en tres etapas: i) identificación de necesidades, ii) caracterización del registro administrativo y, iii) conclusiones y recomendaciones.

Finalmente, mencionó que en la vinculación de RRAA se deben considerar algunos aspectos como si el número de identificación es suficiente, si es o no necesario el uso de dígito de verificación y los cruces no determinísticos.

5. Aportes de los expertos frente a los temas expuestos, de acuerdo a las problemáticas abordadas y el Marco de Interoperabilidad

Jorge Andrés Gallego indicó que era interesante el ejemplo trabajado en cuanto al éxito de interoperabilidad a pesar de desafíos y cuellos de botella que se pudieron presentar, y preguntó, si generalmente los casos o eventos son exitosos, pues, de acuerdo a las necesidades el DANE acude o no a otra entidad. En muchos casos la relación va en sentido contrario, otras entidades requieren información del DANE y puede que no siempre se llegue a buen término. Señaló que, siendo un caso exitoso, cuántos casos no eran de mostrar, por ejemplo, por temas de acceso a información y sobre todo cuando el DANE es el proveedor.

También señaló que en el asesorar al CAD, había un principio que notó que estaba ausente y era el de la oportunidad de la información, por ejemplo, en el tema de la pandemia. Indicó que, desde su perspectiva, la sensación que se tenía afuera era que los procesos eran engorrosos y demorados. Comentó que en este momento se encuentra apoyando al DNP en la evaluación del impacto del Ingreso Solidario en la movilidad de personas y que los cruces con censos resultarían útiles. Recomendó revisar cómo funciona la institucionalidad, y qué cambios o mejoras debían realizarse.

Andrés Holguín, Jefe de la Oficina de Sistema, agregó que frente a la pregunta sobre cuántas solicitudes se recibían, DIRPEN tenía el dato exacto, sin embargo, se realizaban Comités de Reserva Estadística cada 15 días, y por lo general, se recibían de 3 a 6 solicitudes mensuales. Indicó que es difícil el proceso porque en la mayoría de los casos para poder delimitar lo que se necesita es complejo, y un ejemplo de eso es que el DANE lleva 6 meses negociando con la Contraloría, porque ellos no han podido delimitar la necesidad que tienen. Según ellos, necesitan el censo, pero no se puede entregar, es necesario definir las variables puntuales que requieren. De la información que llega al DANE, por ejemplo, Cuentas Nacionales, se construye con base en información de DIMPE, pero es el 30% de lo que necesitan, usan de manera masiva RRAA que entregan otras entidades. También se presentan problemas para recibir información, si el DANE necesita información, ejemplo puntual el caso de la DIAN, se requiere 2 años para solicitar información, pero se debe estructurar qué se necesita puntualmente por parte del DANE, pues, según la Ley 1581 se debe justificar la necesidad de cada variable que se solicita.

Igualmente indicó que con las entidades privadas es aún más complejo la solicitud de información. Se necesita un proceso más eficiente y determinar la forma como se hacen públicos los datos.

Valérie Gauthier, comentó que era interesante utilizar estándares, y propuso el uso de la privacidad diferencial como una nueva línea en lugar de anonimizar, adicionalmente dijo que se podían realizar cruces para sacar información, y de esta manera, se acceder a los datos sin la necesidad de anonimizar.

Andrés Holguín agregó que estaba de acuerdo y que habían varias alternativas en vez de anonimizar. Sin embargo, era importante tener en cuenta que se habían presentado inconvenientes a nivel mundial, como, por ejemplo, el censo en Australia.

Valérie Gauthier mencionó que sería interesante analizar cómo cruzar diferentes bases de datos con algunas bases, teniendo en cuenta que estamos amarrados a las disposiciones de MinTic y demás instituciones, hablando el mismo idioma. También preguntó sobre el proyecto de la cédula electrónica.

A esta duda, Andrés Holguín contestó que se reactivaba la cedula electrónica. Y que se contaba con las directrices que estableciera la Agencia Nacional Digital, entidad mediadora.

Valérie Gauthier señaló que sería oportuno trabajar con la Agencia Nacional Digital con el fin de encontrar lineamientos frente a los aspectos analizados en la sesión. Andrés Holguín aclaró que se estaba trabajando en el tema de lenguaje común con el fin de analizar cómo deberían guardarse los datos con el propósito de mantener la seguridad e interoperar sin anonimizar.

Valérie Gauthier, agregó que se podían manejar otras bases, pero no con información sensible, y mencionó como ejemplo cruzar información de compras de la canasta básica. Juan Sebastián Ordoñez, manifestó que en este punto debía contarse con un identificador para datos que protegiera la información.

Valérie Gauthier propuso el uso de un *Hash* con clave, y que mencionó, por ejemplo, que en cripto se usa el Oráculo (tiene toda la información) para buscar un prototipo.

León Darío Parra, indicó que el ejercicio trabajado era interesante y retomó el aporte de Jorge Andrés Gallego sobre la interoperabilidad en la práctica. Señaló que aspectos jurídicos y *Habeas Data* constituyen un problema a resolver y luego de esto, entraría la anonimización de datos. Cuando se cruzan datos en PILA, hay información basura, que no se puede utilizar o son datos perdidos, cuando es pequeño en imputación de datos, en menos del 5% hay varias metodologías para tener calidad del dato; asimismo, la información que se cruza no está limpia y no genera confiabilidad. Cuando se cruza información pueden quedar muchos datos perdidos que impiden realizar buenos análisis de datos.

Por otro lado, el experto preguntó cuál era el protocolo o procedimiento para realizar ejercicios de interoperabilidad y que no pasara lo que mencionó anteriormente, e indicó que, desde su perspectiva, hacía falta conocer buenas prácticas para evaluar la calidad del dato. A esto Juan Sebastián Gallego respondió que por debajo de determinado porcentaje no se usaba una variable y que se buscaba realizar ejercicios de consistencia con registros base, por ejemplo, limpiar la cedula para hacer análisis de consistencia, una actividad que constituía un gran esfuerzo; por lo tanto, se debía automatizar sobre cualquier registro para velocidad de procesamiento, intercambio y fomentar el uso de la nube.

6. Conclusiones y cierre

Juan Sebastián Ordoñez mencionó las funciones de asesoría de la Sala de Modernización Tecnológica dentro del marco del SEN y la interacción con el CAD.

Referente a la presentación PILA:

- Ministerio de Salud, dueño del registro administrativo PILA, que se utiliza como complemento de las operaciones estadísticas DANE dentro del marco de la pandemia.
- Desafíos: estandarización de las variables, las llaves de cruce, identificadores y realización de pruebas de consistencia,
- Se realizó gestión de proveedores para contar con la base semestral, trimestral y, finalmente, mensual dentro del marco de la pandemia.
- Luego de varias reuniones, y dado que al registro administrativo PILA se le realizan varios procesos, nace el registro estadístico RELAB-Registro de Relaciones Laborales. En la gestión de intercambio, almacenamiento y custodia se realiza la interoperabilidad, a través de nube/local y con niveles de acceso. En términos de calidad queda la consulta sobre quién asegura la calidad del registro.
- Seguridad, privacidad, ¿quién es el responsable de esta? ¿Se segmenta o no? ¿Qué algoritmos se deben utilizar para preservar las llaves de identificación sensibles para los proveedores y para garantizar la reserva? ¿Quién va a determinar cómo garantizar la seguridad y la privacidad de intercambios?
- Aprovechamiento de registros: ¿Cómo se realiza la vinculación de los registros? ¿Cómo se debe hacer? ¿Existen estándares para vincular los registros?
- Se presentó el Marco Conceptual De Interoperabilidad, se consultó cuál sería la mejor manera de realizar interoperabilidad garantizando la seguridad, la reserva, la calidad y, generar confianza con los proveedores para obtener los registros necesarios tales como el reporte de ingresos de la DIAN, por ser una variable sensible.
- Clasificación de actividad entre registros, existe diferencia de clasificación entre cada uno de los registros administrativos que afectan la interoperabilidad.
- Retos y obligaciones: Cómo realizar las sinergias entre el dueño del registro y el instituto estadístico para garantizar la calidad de los productos de salida.
- ¿Cómo garantizar la privacidad de entrada y de salida? ¿Quién es el encargado de realizar la auditoría de estos productos?
- Dimensiones de calidad: ¿Quién va a definir los indicadores de calidad?, (cobertura, integralidad, exactitud, cobertura, emparejamiento, valores faltantes por emparejamiento, errores de agregación) ¿Quién garantiza o cuáles son los estándares de calidad que se requieren para garantizar los outputs?
- Se presentó el proyecto piloto RELAB. Se expusieron algunas conclusiones y consideraciones para el tratamiento de datos. Finalmente, se mostraron los indicadores que se realizaron con este registro y los proyectos que se están realizando para utilizar este registro, con el propósito de fortalecer los grupos donantes y realizar imputaciones durante la encuesta.

Referente a los comentarios de expertos:

- Jorge Andrés Gallego:

La interoperabilidad de PILA es un caso exitoso, ¿cuántos casos no son exitosos y, en particular, cuando el DANE es el dueño del registro? Respuesta 6 mensuales.

La interoperabilidad debe ser en doble vía. Se debe trabajar en las mejoras, garantizando la reserva en términos oportunos, a nivel DANE.

Los requisitos para acceder a la interoperabilidad con el DANE deben tener en cuenta el principio de oportunidad.

- Valérie Gauthier:

Analizar la posibilidad de "privacidad diferencial"

Realizar cruces de bases de datos para obtener información de aquellos datos que no son sensibles, por ejemplo, en compras de la canasta familiar.

Realizar una alianza con la Agencia Nacional Digital.

- León Darío Parra:

Primero deben resolverse aspectos asociados al tema jurídico, luego al *Habeas Data* y después si abordar la anonimización de datos.

Al realizar cruces de información se puede presentar información basura o datos perdidos, por lo tanto, es importante tener en cuenta las buenas prácticas y calidad en el manejo de la información.

Compromisos

1. Tarea: Enviar acta

Responsable: DANE

Fecha entrega: Del 7 al 9 de octubre.

2. Tarea: Propuesta agenda cuarta reunión Sala de Modernización Tecnológica

Responsable: DANE

Fecha entrega: Antes de la próxima reunión

3. Tarea: Exposición por parte de los expertos en cuanto a aportes y recomendaciones frente a las problemáticas expuestas y Marco de Interoperabilidad

Responsable: León Darío Parra Bernal, Valérie Gauthier Umaña, Jorge Andrés Gallego Duran

Fecha entrega: Cuarta reunión 27 de octubre

Próxima reunión:

Responsable de convocar: Secretaría Técnica
CASEN

Fecha: 27 de octubre 9:00am
