

**INFORME DE AVANCE**

“Porcentaje de documentos metodológicos que limitan los avances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del Marco Maestro de Empresas”

Documento aprobado el 31 de diciembre de 2023 por:

Denisse López Arenas, jefe Departamento de Metodología e Innovación Estadística

**DEPARTAMENTO DE METODOLOGÍA E INNOVACIÓN ESTADÍSTICA**

**SUBDEPARTAMENTO DE DISEÑO DE MARCOS Y MUESTRAS**

**INSTITUTO NACIONAL DE ESTADÍSTICAS**

**Diciembre / 2023**

**DEPARTAMENTO DE METODOLOGÍAS E INNOVACIÓN ESTADÍSTICA**

**SUBDEPARTAMENTO DE DISEÑO DE MARCOS Y MUESTRAS**

**UNIDAD DE MARCOS MAESTROS**

Instituto Nacional de Estadísticas

Diciembre / 2023

Elaborado por: Pablo Farías Rubio, coordinador de Sección de Marcos Maestros.

Katherine Rojas Guerrero, analista Sección de Marcos Maestros.

Sebastián Matamala Valencia, analista Sección de Marcos Maestros.

Revisado por: Iván Touron Romero, jefe Subdepartamento de Diseño de Marcos y Muestras.

Aprobado por:

ÍNDICE

[1. INTRODUCCIÓN 5](#_Toc150248929)

[2. PROCESO DE TRABAJO 6](#_Toc150248930)

[3. ANTECEDENTES GENERALES 7](#_Toc150248931)

[4. OBJETIVOS 9](#_Toc150248932)

[4.1 Objetivos específicos 9](#_Toc150248933)

[5. MARCO DE REFERENCIA 10](#_Toc150248934)

[6. MARCO MAESTRO DE EMPRESAS 12](#_Toc150248935)

[6.1 Fuentes de información para la construcción y/o actualización del MME 17](#_Toc150248936)

[6.2 Priorización de fuentes de información para asignación 18](#_Toc150248937)

[7. VARIABLES SELECCIONADAS 19](#_Toc150248938)

[7.1 Tipos de procedencias 21](#_Toc150248939)

[7.2 Selección de variables a estudiar 21](#_Toc150248940)

[7.3 Consideraciones de actualización para las variables seleccionadas 25](#_Toc150248941)

[8. DESARROLLO Y RESULTADOS 25](#_Toc150248942)

[9. CONCLUSIÓN 31](#_Toc150248943)

[10. ANEXOS 32](#_Toc150248944)

[8.1 Arboles de asignación para la información del MME 33](#_Toc150248945)

[8.2 Arboles de asignación para variable actividad económica 33](#_Toc150248946)

[8.3 Arboles de asignación para variable ubicación geográfica 37](#_Toc150248947)

[8.4 Arboles de asignación para variable Medios de contacto 40](#_Toc150248948)

ESTADO DE CUMPLIMIENTO

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre indicador | Fórmula de cálculo | Meta | Avance del indicador |
| Porcentaje de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del Marco Maestro de Empresas. | (N° de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del MME elaborados en el año t/ N° total de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del Marco Maestro de Empresas comprometidos para el año t)\*100 | 100 % de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del Marco Maestro de Empresas. | (1/ 1) \* 100 = 100% |

IDENTIFICACIÓN DEL INDICADOR DE GESTIÓN

|  |  |
| --- | --- |
| **Nombre indicador** | Porcentaje de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del MME |
| **Fórmula de cálculo** | (N° de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del MME elaborados en el año t/ N° total de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del MME comprometidos para el año t)\*100 |
| **Meta** | 100% de documentos metodológicos que limitan los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del MME. |
| **Notas explicativas** | 1. Se entenderá por MME al Marco Maestro de Empresas.  2. El documento metodológico que limita los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del MME corresponde a uno.  3. El documento metodológico que limita los alcances y criterios que se utilizan para determinar la obsolescencia del microdato en variables seleccionadas del MME deberá ser enviado mediante correo electrónico a más tardar el 15 de diciembre de 2023 y deberá contener información respecto de las variables seleccionadas, criterios utilizados en su selección, análisis de obsolescencia del microdato y resultados obtenidos |
| **Plazo de entrega** | Diciembre 2023 |
| **Unidad responsable** | Subdepartamento de Diseño de Marcos y Muestras |

# INTRODUCCIÓN

Los marcos estadísticos desempeñan un papel esencial en la organización y gestión de datos en el ámbito empresarial, proporcionando la estructura necesaria para la recopilación y análisis sistemático de información. Estos marcos se utilizan para identificar las unidades de muestreo en una población objetivo y son fundamentales para diversas operaciones estadísticas. Además, funcionan como listas de unidades que almacenan datos relacionados con diferentes características, permitiendo la comparación y evaluación a lo largo del tiempo. Sin embargo, para mantener la relevancia y la utilidad de estos marcos, es crucial garantizar la constante actualización de la información contenida en ellos.

En el contexto específico de la construcción y actualización del Marco Maestro de Empresas (MME), surge el desafío de abordar la obsolescencia de los datos de manera efectiva. La integridad y actualización de los datos recopilados son fundamentales para asegurar la calidad y la precisión de los resultados esperados. La obsolescencia de los datos se refiere a la pérdida de vigencia con el tiempo, lo que puede afectar la validez de la información contenida en el MME. Por lo tanto, es imperativo establecer criterios claros y coherentes para determinar cuándo es necesario actualizar o reemplazar conjuntos de datos específicos.

La actualización constante de datos desempeña un papel crucial en la planificación y ejecución de operaciones estadísticas. La falta de actualización puede tener un impacto directo en la eficiencia operativa y, a su vez, afectar las metas y estrategias de la institución responsable de estos procesos. La no actualización de datos puede llevar a la carencia de información actualizada sobre elementos críticos, como medios de contacto o identificadores de unidades empresariales. Además, al analizar información desactualizada, pueden surgir imprecisiones en las estimaciones, lo que podría resultar en conclusiones erróneas.

El presente informe se estructura de la siguiente manera: se exhibe el proceso de trabajo y los antecedentes generales de la investigación, posteriormente se definen los objetivos, el marco de referencia y se entrega una contextualización de lo que respecta al MME y sus métodos de asignación de información. De forma subsiguiente, se exponen las variables seleccionadas y los tipos de procedencia que poseen estas, para así, finalmente presentar los análisis del estudio, las conclusiones y anexos.

# PROCESO DE TRABAJO

El proceso de trabajo para el presente estudio se originó con la detección de la problemática respecto a la presencia de datos obsoletos. Para abordar y delimitar esta dificultad, se determinó que la versión del Marco Maestro de Empresa (MME) a trabajar correspondería a la del año 2022, que es de uso interno en el Subdepartamento de Diseños Marcos y Muestras y contiene información relacionada a las procedencias de algunas variables.

Posterior a lo anterior, se definieron los criterios mediante los cuales se definieron los criterios mediante los cuales se seleccionarían las variables, así como también la clasificación de los tipos de procedencia de datos. Este último aspecto es relevante para la identificación de los datos obsoletos.

De forma paralela a la definición de criterios, se recopilaron antecedentes de recomendaciones internacionales que abordaran temáticas relacionados con frecuencia y periodicidad que se debe considerar para la actualización de variables en directorios de empresas y la obsolescencia de la información, para así generar una justificación robusta sobre los procedimientos a realizar

Con los criterios definidos, se procede a la identificación y selección de las variables mediante las cuales se abordaría el estudio, con la finalidad de determinar la obsolescencia de los datos contenidos.

Finalmente, con los pasos anteriores consumados, se efectuó el análisis de las variables identificadas respecto a sus años de referencia, mediante el uso de estadísticas descriptivas, las cuales se presentarán en los apartados correspondientes.

# ANTECEDENTES GENERALES

Los marcos estadísticos se comprenden como herramientas organizativas que proporcionan la estructura necesaria para la recopilación y el análisis sistemático de datos. Estos, se emplean – dentro de otros elementos – para identificar las unidades de muestreo de una población objetivo determinada. En otros términos, los marcos proporcionan una base sólida para la selección de las muestras que darán paso a diferentes operaciones estadísticas.

Además de lo descrito anteriormente, los marcos se pueden percibir como listas de unidades que poseen información relacionadas a diferentes características. En esta perspectiva, se conciben como una herramienta que permite la comparación de los datos recopilados a lo largo del tiempo. Para que aquello se cumpla, es relevante que la información que contengan se encuentre vigente, es decir, que los datos se estén constantemente actualizando para garantizar la precisión y relevancia de los marcos.

En el contexto de construcción y actualización del MME, es relevante considerar los aspectos relacionados con la obsolescencia de los datos para contar con una mayor robustez respecto al diseño del marco propiamente tal. La integridad y actualización de los datos que se logran recopilar son fundamentales para garantizar la calidad y precisión de los resultados esperados.

La obsolescencia de los datos se puede comprender como la pérdida de vigencia o relevancia con el tiempo de un grupo de datos determinado, lo que puede repercutir en la validez de la información que recopila un marco, en este caso particular el MME. En este contexto, se vuelve indispensable definir criterios claros y consistentes para determinar cuándo un conjunto de datos se considera obsoleto y requiere ser actualizado o reemplazado.

Mantener los datos actualizados en la planificación del levantamiento de información para las operaciones estadísticas es de suma importancia. Esto se debe a que cualquier desactualización puede afectar directamente la eficiencia operativa, y, a su vez, repercutir en las metas y planificaciones estratégicas de la institución encargada de ejecutar estos procesos. En este contexto, la falta de actualización de los datos puede repercutir negativamente al no poder contar con información vigente en aspectos como medios de contacto o elementos que permitan identificar la unidad a la cual se le debe recopilar su información. Por otro lado, al momento del análisis de la información que se capturó, los datos que se encuentren obsoletos pueden generar imprecisiones en las estimaciones, potencialmente resultando en conclusiones erróneas.

En suma, resulta crucial disponer de bases de datos con información actualizada para respaldar la toma de decisiones precisa. La actualización de datos es esencial para tomar decisiones informadas y precisas, dado que la omisión de esta actualización podría dar lugar a errores o a la utilización de datos irrelevantes o desactualizados como fundamento para dichas decisiones.

# OBJETIVOS

Identificar los datos obsoletos en las variables asociadas a un marco especifico, de acuerdo a los criterios metodológicos de actualización.

## 4.1 Objetivos específicos

* Conocer los niveles de completitud de las variables seleccionadas.
* Identificar los años de procedencia de las fuentes de información que se utilizan en la asignación de información para las variables.
* Definir criterios de obsolescencia
* Identificar y seleccionar variables para el análisis
* Identificar datos obsoletos

# MARCO DE REFERENCIA

Las unidades estadísticas que conforman el MME no son estáticas. Dichas entidades se crean, sufren diferentes modificaciones a lo largo de su existencia y – eventualmente – desaparecen, por lo que la información que se recopila de ella puede ir cambiando constantemente. Siguiendo esta lógica y tomando como parámetro el “Manual de recomendaciones sobre los registros de empresas de la Unión Europea”, elaborado en el contexto del proyecto de Cooperación UE-CAN (2010) en materia de estadística, se establece que los registros de empresas se deben actualizar tomando en consideración diferentes parámetros, los cuales se procederán a detallar a continuación.

Al establecer la naturaleza, la frecuencia y la periodicidad de las actualizaciones de las características o variables se debe tener en consideración las necesidades de los usuarios del MME, ya que, poco va a servir invertir en esfuerzos y recursos en la actualización de información que no presenten interés para los usuarios. Asimismo, existen otros factores que influyen al momento de definir la política de actualización de los datos de los registros estadísticos de empresas como el tipo de variable a tratar, el momento de la actualización, la fuente de información que se utiliza y el tamaño de la unidad de observación

Respecto al tipo de variable que se busca actualizar, las que tiene relación con nombres y direcciones se han de reflejar en los marcos tan pronto como se notifiquen, ya que esta información es relevante para el envío de formularios, cuestionarios, etc. De igual manera, los medios de contacto, como teléfono y correo electrónico, se deben actualizar en cuanto se notifica el cambio. En contraparte, las variables que tiene relación con actividad económica, la decisión de si realiza la actualización de forma inmediata o no es más compleja, ya que esto puede afectar que una unidad se incluya o quede relegada de la muestra de una encuesta. Bajo esta premisa, se debe evaluar si es más conveniente esperar a que termine el ciclo anual de una encuesta para realizar la actualización, con el fin de subsanar cualquier tipo de impacto en la calidad de los datos.

En relación con el momento de actualización, el manual recomienda que se tome en consideración las necesidades de los usuarios, ya que estos pueden plantear demandas antagónicas respecto al momento de introducción de los cambios, puesto que algunos desean que estas sean inmediatas, mientras que otros desean aplazarlas hasta que sus efectos en la coherencia de los datos sean mínimos. La solución frente a esto es gestionar las dos situaciones: la primera se actualiza en cuanto se cuente con la información y para los segundos permanecerá igual por un determinado rango de tiempo, hasta que se decida actualizar.

Existen múltiples fuentes de información que pueden proporcionar información relevante para la actualización de variables, lo que conlleva cierto grado de conflicto entre ellas. En este contexto, para discernir cual es la mejor fuente de recursos a utilizar, estas se deben conocer en profundidad, incluyendo aspectos como métodos de recogida y validación de los datos; el conocimiento de estos elementos permite establecer prioridades entre las distintas fuentes de información. Determinado este orden de importancia, se ha de establecer un modo de aplicarlas a los procesos de actualización de la información. Un método fiable es asignar a las características una fecha de procedencia, ya que este tipo de información puede decidir que los datos procedentes de una fuente con menor prioridad pueden corregir a otros procedentes de una fuente que en principio se consideraría mejor, pero que es considerablemente más antigua.

En cuanto al tamaño de la unidad estadística, el manual estipula que en aquellas de mayor tamaño es más complejo realizar el proceso de actualización en relación con una de menor tamaño. El interés por las unidades más grandes – generalmente - es mayor por parte de los usuarios, por lo que los procesos de actualización de sus datos deben ser correctos. Para realizar estos, se recomienda establecer contacto directo con el informante de la unidad para validar el cambio de información o comparar los datos de una serie de fuentes y comprobar si son coherentes con la información que se requiere actualizar.

# MARCO MAESTRO DE EMPRESAS

El MME corresponde a un registro con fines estadísticos que contiene las empresas que se encuentran dentro del territorio nacional y que realicen alguna actividad económica dentro de un período de tiempo acotado. Dichas unidades son identificadas a través de un Rol Único Tributario (RUT), los que se obtienen de una serie de registros administrativos y encuestas del INE, que mediante su depuración dan paso a registros con variables asignadas.

El MME cuenta con un set de variables en los ámbitos real y ficticio[[1]](#footnote-1), además de marcas adicionales que permiten identificar características de las UL que lo componen. El objetivo de estas es poder otorgar un mayor nivel de información a los usuarios del MME en caso de requerirla, además de poder ser utilizadas como filtros. En las siguientes tablas se presentan las variables por ámbito:

**Tabla 1: Descriptor de campos del Marco Maestro de Empresas ámbito real.**

|  |  |
| --- | --- |
| Nombre de la variable | Descripción |
| RUT\_REAL | Rol único tributario otorgado por la administración fiscal, que identifica a la unidad legal en forma oficial. |
| DV\_RUT | Dígito verificador del RUT |
| ID\_GRUPO\_EMPRESAS\_INE | Identificador(es) de cada Grupo según Grupos de Empresas INE. |
| ID\_GRUPO\_EMPRESAS\_CMF | Identificador de cada grupo según la Comisión del Mercado Financiero (CMF). |
| NOM\_GRUPO\_EMPRESAS\_INE | Nombre del grupo de empresas identificado por INE. |
| RAZON\_SOCIAL | Nombre o denominación de la unidad legal que desarrolla una actividad económica, sea persona natural o jurídica. |
| NOM\_REPRESENTANTE\_LEGAL | Nombre del representante legal de la empresa. |
| NOMBRE\_FANTASIA | Nombre o denominación comercial de la unidad legal que desarrolla una actividad económica, sea persona natural o jurídica. |
| REGION\_CUT\_2018 | Código de 2 dígitos correspondiente a la Región, de acuerdo con la división político-administrativa oficial (RPC), vigente desde el 06 de septiembre 2018. |
| PROVINCIA\_CUT\_2018 | Código de 3 dígitos correspondiente a la Provincia, de acuerdo con la división político-administrativa oficial (RPC), vigente desde el 06 de septiembre 2018. |
| COMUNA\_CUT\_2018 | Código de 5 dígitos correspondiente a la Comuna, de acuerdo con la división político-administrativa (RPC), vigente desde el 06 de septiembre 2018. |
| DIRECCION | Ubicación Geográfica asociada a la unidad local principal. |
| INICIO\_DE\_ACTIVIDADES | Fecha de inicio de actividades según el SII. |
| SINDICO\_DE\_QUIEBRAS | Marca de la nómina de síndico de quiebras. |
| SECCION\_CIIU4CL | Primer nivel de clasificación económica CIIU Rev.4. Adaptación nacional (letra). En caso de no tener actividad asignable, toman el valor 0. |
| DIVISION\_CIIU4CL | Segundo nivel de clasificación económica CIIU Rev.4. Adaptación nacional (2 dígitos). En caso de no tener actividad asignable, toman el valor 00. |
| GRUPO\_CIIU4CL | Tercer nivel de clasificación económica CIIU Rev.4. Adaptación nacional (3 dígitos). En caso de no tener actividad asignable, toman el valor 000. |
| CLASE\_CIIU4CL | Cuarto nivel de clasificación económica CIIU Rev.4. Adaptación nacional (4 dígitos). En caso de no tener actividad asignable, toman el valor 0000. |
| SUBCLASE\_CIIU4CL | Quinto nivel de clasificación económica CIIU Rev.4. Adaptación nacional (5 dígitos). En caso de no tener actividad asignable, toman el valor 00000. |
| ESTRATOS\_F29 | Estrato calculado según ventas anuales en UF obtenidas de la selección de algunos códigos del F29. |
| FONO | Teléfono principal de la empresa. |
| CORREO\_ELECTRONICO | Correo electrónico principal de la Unidad Legal. |
| PAGINA\_WEB | Página WEB principal de la Unidad Legal. |
| DIRECTORIO\_PRIORIZADO | Marca que indica 1 si cumple con ser Registros de tamaño grande (estrato 9 al 12) o registros coincidentes con la producción estadística o Grupos de empresas INE, y 0 para los no coincidentes. |
| GRANDES\_CONTRIBUYENTES | Marca de grandes contribuyentes, 1 para los grandes contribuyentes y 0 para los que no lo son. |
| REGIMEN\_TRIBUTARIO | Marca de la nómina de Régimen Tributario. |
| CATEGORIA\_TRIBUTARIA | Marca que indica con un 1 si la unidad legal es de primera categoría tributaria, 2 si es de segunda categoría, 3 si presenta ambas categorías y 0 si no tiene categoría asociada. |
| TIPO\_CONTRIBUYENTE | Tipo de sociedad constituida por el contribuyente según el SII. |
| SUBTIPO\_CONTRIBUYENTE | Mayor desagregación del tipo de sociedad declarada por el contribuyente según SII. |
| COMERCIO\_EXTERIOR | Marca de las unidades legales presentes en BBDD de Aduana. Solo exporta (1), solo importa (2) o ambas (3). |
| MONEDA\_EXTRANJERA | Marca de las unidades legales que declaran formularios del SII en moneda extranjera, 1 para los coincidentes y 0 para los no coincidentes. |
| AFECTO\_A\_IVA | Empresas tiene la obligación de para el Impuesto al Valor Agregado, 1 para afecto a IVA y 0 para los que no. |
| PERS\_NAT\_CEPAL[[2]](#footnote-2) | Corresponde a una marca de unidades legales, el cual toma valor 1 si tiene cinco o más personas empleadas, y toma valor 0 si la unidad legal tiene menos de cinco empleados a cargo. |
| TAMAÑO\_OCUPADOS\_AAAA | Tamaño de empresa según ocupados, regidos por la Ley N°20.416 del Estatuto Pyme. |
| TIPO\_DE\_REGISTRO | Característica del registro de acuerdo con su aparición en la ejecución del Directorio anterior como registro continuo (0) o registro nuevo (1). |
| PERS\_NAT\_SII | Marca de las unidades legales según criterio de corte de Rut menor a cincuenta millones (SII). |
| TERMINO\_GIRO\_AAAA | Marca de la nómina de registros que posee término de giro ante el SII. |
| ID | Identificador correlativo de uso interno. |
| NORMALIZACION | Marca que indica con un SI la dirección geográfica fue normalizada, NO si no lo fue y NA si no presenta dirección. |
| PEE\_AAAA | Marca que indica con un 1 si el registro fue levantado en la producción estadística estructural del año AAAA y con un 0 si no lo fue. |
| PEC\_AAAA | Marca que indica con un 1 si el registro fue levantado en la producción estadística coyuntural del año AAAA y con un 0 si no lo fue. |
| PRODUCCION\_ESTADISTICA | Marca que indica con un 1 si pertenece a la producción estadística estructural, 2 si pertenece a la producción coyuntural, 3 si pertenecen a ambas producciones estadísticas INE y, 0 si no pertenecen a los casos anteriores. |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Tabla 2: Descriptor de campos del Marco Maestro de Empresas ámbito ficticio**

|  |  |
| --- | --- |
| Nombre de la variable | Descripción |
| RUT\_FICTICIO | RUT ficticio asociado a una empresa. |
| REGION\_CUT\_2018 | Código de 2 dígitos correspondiente a la Región, de acuerdo con la división político-administrativa oficial (RPC), vigente desde el 06 de septiembre 2018. |
| PROVINCIA\_CUT\_2018 | Código de 3 dígitos correspondiente a la Provincia, de acuerdo con la división político-administrativa oficial (RPC), vigente desde el 06 de septiembre 2018. |
| COMUNA\_CUT\_2018 | Código de 5 dígitos correspondiente a la Comuna, de acuerdo con la división político-administrativa (RPC), vigente desde el 06 de septiembre 2018. |
| SECCION\_CIIU4CL | Primer nivel de clasificación económica CIIU Rev.4. Adaptación nacional (letra). En caso de no tener actividad asignable, toman el valor 0. |
| DIVISION\_CIIU4CL | Segundo nivel de clasificación económica CIIU Rev.4. Adaptación nacional (2 dígitos). En caso de no tener actividad asignable, toman el valor 00. |
| GRUPO\_CIIU4CL | Tercer nivel de clasificación económica CIIU Rev.4. Adaptación nacional (3 dígitos). En caso de no tener actividad asignable, toman el valor 000. |
| CLASE\_CIIU4CL | Cuarto nivel de clasificación económica CIIU Rev.4. Adaptación nacional (4 dígitos). En caso de no tener actividad asignable, toman el valor 0000. |
| SUBCLASE\_CIIU4CL | Quinto nivel de clasificación económica CIIU Rev.4. Adaptación nacional (5 dígitos). En caso de no tener actividad asignable, toman el valor 00000. |
| VTA\_$\_ENERO | Ventas del mes - 1 informado en el F29. |
| VTA\_$\_FEBRERO | Ventas del mes - 2 informado en el F29. |
| VTA\_$\_MARZO | Ventas del mes - 3 informado en el F29. |
| VTA\_$\_ABRIL | Ventas del mes - 4 informado en el F29. |
| VTA\_$\_MAYO | Ventas del mes - 5 informado en el F29. |
| VTA\_$\_JUNIO | Ventas del mes - 6 informado en el F29. |
| VTA\_$\_JULIO | Ventas del mes - 7 informado en el F29. |
| VTA\_$\_AGOSTO | Ventas del mes - 8 informado en el F29. |
| VTA\_$\_SEPTIEMBRE | Ventas del mes - 9 informado en el F29. |
| VTA\_$\_OCTUBRE | Ventas del mes - 10 informado en el F29. |
| VTA\_$\_NOVIEMBRE | Ventas del mes - 11 informado en el F29. |
| VTA\_$\_DICIEMBRE | Ventas del mes - 12 informado en el F29. |
| VTA\_ANUAL\_$\_F29\_AAAA | Ventas acumuladas de los últimos 12 meses según el F29. |
| VTA\_ANUAL\_UF\_F29\_AAAA | Ventas acumuladas en UF de los últimos 12 meses según el F29. |
| ESTRATOS\_F29 | Estrato calculado según ventas anuales en UF obtenidas de la selección de algunos códigos del F29. |
| TAMAÑO\_F29 | Tamaño de la unidad legal obtenido según el monto de la venta anual calculado desde el Formulario 29 en UF (micro, pequeña, mediana, grande). |
| CUENTA\_OCUPADOS\_AAAA | Total de trabajadores según F1887. |
| TAMAÑO\_OCUPADOS\_AAAA | Tamaño de empresa según ocupados, regidos por la Ley N°20.416 del Estatuto Pyme. |
| PROM\_TRAB\_OCUPADOS\_AAAA | Promedio anual de trabajadores según F1887. |
| PROM\_TRAB\_HONORARIO\_AAAA | Promedio anual de trabajadores a honorarios según F1879. |
| SINDICO\_DE\_QUIEBRAS | Marca de síndico de quiebras, 1 para los coincidentes y 0 para los que no lo son. |
| TERMINO\_GIRO\_AAAA | Marca de la nómina de registros que posee término de giro ante el SII. |
| GRUPO\_EMPRESAS\_INE | Marca de las unidades que pertenecen a algún grupo de empresas según levantamientos de encuestas INE. 1 para los coincidentes y 0 para los que no lo son. |
| MONEDA\_EXTRANJERA | Marca de moneda extranjera, 1 para los coincidentes y 0 para los que no lo son. |
| GRANDES\_CONTRIBUYENTES | Marca de grandes contribuyentes, 1 para los coincidentes y 0 para los que no lo son. |
| GRUPO\_EMPRESAS\_CMF | Marca de las empresas que pertenecen a un grupo empresarial según la CMF, 1 para los coincidentes y 0 para los que no lo son. |
| PRODUCCION\_ESTADISTICA | Marca que indica con un 1 si pertenece a la producción estadística estructural, 2 si pertenece a la producción coyuntural, 3 si pertenecen a ambas producciones estadísticas INE y, 0 si no pertenecen a los casos anteriores. |
| CATEGORIA\_TRIBUTARIA | Marca que indica con un 1 si la unidad legal es de primera categoría tributaria, 2 si es de segunda categoría, 3 si presenta ambas categorías y 0 si no tiene categoría asociada. |
| DIRECTORIO\_PRIORIZADO | Marca de directorio priorizado, 1 para los coincidentes y 0 para los que no lo son. |
| REGIMEN\_TRIBUTARIO | Marca de la nómina de Régimen Tributario. |
| COMERCIO\_EXTERIOR | Marca de las unidades legales presentes en BBDD de Aduana. Solo exporta (1), solo importa (2) o ambas (3). |
| AFECTO\_A\_IVA | Empresas tiene la obligación de para el Impuesto al Valor Agregado, 1 para afecto a IVA y 0 para los que no. |
| PERS\_NAT\_SII | Marca de las unidades legales según criterio de corte de Rut menor a cincuenta millones (SII). |
| PERS\_NAT\_CEPAL[[3]](#footnote-3) | Marca de las unidades legales según CEPAL,2013. "COMPENDIO DE RECOMENDACIONES TÉCNICAS Y BUENAS PRÁCTICAS PARA EL DESARROLLO DE LOS DIRECTORIOS DE EMPRESAS Y ESTABLECIMIENTOS ", pág.6. |
| ID | Identificador correlativo de uso interno. |
| TIPO\_DE\_REGISTRO | Indicador de si el registro es continuo o nuevo. |
| PEE\_AAAA | Marca que indica con un 1 si el registro fue levantado en la producción estadística estructural del año AAAA y con un 0 si no lo fue. |
| PEC\_AAAA | Marca que indica con un 1 si el registro fue levantado en la producción estadística coyuntural del año AAAA y con un 0 si no lo fue. |
| NOMINA\_F22\_AAAA | Marca de registros que están contenidos en el Formulario de declaración de Renta (F22). |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

## 6.1 Fuentes de información para la construcción y/o actualización del MME

Las fuentes de información que se emplean en la construcción del MME pueden proceder de distintas fuentes de origen. Estas fuentes sirven para evidenciar la procedencia del microdato cada vez que se realiza el proceso de asignación.

A continuación, las tablas 3 y 4 muestran la fuente administrativa del origen de la información que alimentan las variables del MME, el nombre de las fuentes de información y su periodicidad de difusión para con INE.

**Tabla 3: Fuentes de información externas para la construcción y actualización del MME (ambos ámbitos)**

|  |  |  |
| --- | --- | --- |
| Fuente Externa | Fuente de información | Periodicidad |
| Servicios de Impuestos Internos | Formulario N°29 | Mensual |
| Formulario N°22 | Anual |
| Declaraciones juradas (DDJJ) N°1887 | Anual |
| DDJJ N°1879 | Anual |
| Catastro de contribuyentes | Anual |
| Catastro de direcciones | Anual |
| Catastro de giros | Anual |
| Nómina de grandes contribuyentes | Anual |
| Nómina de término de giro | Anual |
| Nómina representante legal | Anual |
| Nómina de moneda extranjera | Anual |
| Nómina de régimen tributario | Anual |
| Municipalidades | Patentes municipales | Anual |
| Servicio Nacional de Aduanas | Importaciones | Mensual |
| Exportaciones | Mensual |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Tabla 4: Fuentes de información internas para la construcción y actualización del MME (ambos ámbitos)**

|  |  |  |
| --- | --- | --- |
| Fuente interna | Fuente de información | Periodicidad |
| Instituto Nacional de Estadísticas (INE) | Producción estadística | Anual |
| Base de datos (BBDD) Sector público y sectoriales | Anual |
| BBDD Grupo de empresas | Anual |
| Verificación telefónica | Anual |
| MME anterior | Anual |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

## 6.2 Priorización de fuentes de información para asignación

En esta sección, se procederá a presentar la asignación de datos en conformidad con la priorización de estas fuentes utilizados con este propósito. Esta decisión se basa en el hecho de que las fuentes están vinculados a distintos años, lo cual enriquece la rastreabilidad de la procedencia de los datos en el MME.

La priorización de las fuentes de información desempeña un papel crucial en el proceso de asignación de microdato en el MME. Cada vez que se efectúa una actualización de estos basado en una fuente, se registra en una variable la información sobre la procedencia de los datos. No obstante, es importante señalar que no todas las variables cuentan con una variable interna asociada que pueda proporcionar este tipo de información, lo cual plantea una problemática que debe ser abordada.

En la tabla 5 se presenta de manera detallada la priorización de las fuentes de información utilizados para la asignación de datos en cada variable del Marco Maestro de Empresas. En esta lógica, se comprende que para una variable, aquellas fuentes que presenten valor 1 son los primeros seleccionados para asignar información, en tanto que los que tengan una valoración de 2 serán los segundos en ser empleados y así sucesivamente. A modo de ejemplo, para construir la variable “medios de contactos” se utiliza como primera opción la información de Verificación Telefónica (n°1), para posteriormente emplear la Producción Estadística (n°2) y así prosiguiendo hasta completar la información requerida.

**Tabla 5.Priorización de fuentes de información según cada variable del MME (ambos ámbitos).**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Asignación de datos según fuente de información | | | | | | |
| Dimensión  y/o  variables  Fuentes de  información | **Actividad económica** | **Ubicación geográfica** | **Medios de Contacto** | **Razón social** | **Representante legal** | **Nombre de fantasía** |
| Verificación Telefónica | 0 | 1 | 1 | 1 | 1 | 1 |
| BBDD Grupo de Empresas INE | 3 | 3 | 4 | 4 | 0 | 0 |
| BBDD Sectoriales y Sector Público | 1 | 2 | 3 | 5 | 0 | 2 |
| Producción Estadística | 0 | 4 | 2 | 2 | 3 | 3 |
| MME (ejecución anterior) | 2 | 6 | 5 | 6 | 0 | 0 |
| Grupos Económicos CMF | 0 | 0 | 0 | 3 | 0 | 0 |
| Catastro de Giros | 5 | 0 | 0 | 0 | 0 | 0 |
| Formulario 29 | 6 | 0 | 0 | 0 | 0 | 0 |
| Formulario 22 | 4 | 0 | 0 | 0 | 0 | 0 |
| Síndico de Quiebras | 0 | 0 | 0 | 0 | 0 | 0 |
| Catastro Razón Social | 0 | 0 | 0 | 7 | 0 | 0 |
| Catastro Representante legales | 0 | 0 | 0 | 0 | 2 | 0 |
| Catastro Direcciones | 0 | 5 | 0 | 0 | 0 | 0 |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

La Tabla 5, muestra la asignación de los datos en las dimensiones y/o variables del MME de acuerdo a la priorización de las fuentes de información. Cabe destacar, que las dimensiones para la asignación son Actividad económica (contempla las variables de sección, división, grupo, clase y subclase asociadas a CIIU4CL 2012), Ubicación geográfica (representada por la dirección, región, provincia, comunas asociadas a la CUT 2018) y Medios de contacto (compuesta por teléfono, página web y correo electrónico).

Por ejemplo, en el caso de la razón social, la asignación del dato sigue un procedimiento específico. En primera instancia, se utiliza la información de verificación telefónica como opción principal. En caso de que esta fuente no proporcione la información completa, se recurre a la información de producción estadística, que puede ser de carácter estructural o coyuntural. Si aún persisten estos vacíos de información, se procede a emplear la información de grupos empresariales proporcionado por la Comisión del Mercado Financiero (CMF), seguido por los grupos empresariales del INE, bases de datos sectoriales y fuentes del sector público. Este proceso culmina con la consulta del Marco Maestro de Empresas previamente actualizado y, finalmente, con el catastro de razón social, con el objetivo de completar los datos faltantes en la variable correspondiente.

# VARIABLES SELECCIONADAS

En el presente capítulo se aborda la selección de las variables del MME a considerar en el estudio, definiciones utilizadas y la priorización realizada.

Con la finalidad de analizar la obsolescencia de los datos contenidos en el MME se utilizaron las siguientes variables:

* REGION\_CUT\_2018
* SECCION\_CIIU4CL
* FONO
* CORREO\_ELECTRÓNICO
* PAGINA\_WEB

La selección de estas variables se realizó en base al cumplimiento de los siguientes criterios definidos:

* Fuentes de datos con año de referencia
* Procedencia explícita en el MME
* Tipo de procedencia mixta.

El primero de los criterios de selección, hace referencia a que las fuentes utilizadas deben contar con el año de referencia, ya que, es un dato vital para determinar la procedencia de la información contenidos en las variables del MME, así como también su obsolescencia.

El segundo criterio indica que la procedencia debe estar en forma explícita como una variable en el MME, información que se encuentra en la versión del MME que se utiliza en forma interna en el Subdepartamento de Diseño de Marcos y Muestras.

Finalmente, el tercer criterio utilizado se refiere al tipo de procedencia debe ser mixta. Cabe señalar que en el punto 7.1 del presente documento se especifican los tipos de procedencia utilizados.

## 7.1 Tipos de procedencias

Es importante señalar que para el presente estudio se entiende por procedencia como el origen de los datos asignados a una variable determinada, identificando las fuentes de datos utilizadas y el año de referencia respectivo. En el presente punto se describen los tipos de procedencia que se definieron para clasificar las variables contenidas en el MME, los cuales se pueden categorizar en tres tipos: fija, mixta o sin procedencia específica. A continuación, se detalla que comprende cada una de estas:

* **Fija:** Considera que la variable se deriva de una fuente de información específica correspondiente a un año determinado, el cual no necesariamente coincide con el año del MME.
* **Mixta:** Considera aquellas variables que cuentan con múltiples fuentes de información provenientes de diferentes años. Se pueden observar tres variantes en este tipo de procedencia:
  1. **n – 1:** Existen varias fuentes de información con el mismo año de procedencia, que se utilizaron para asignar el mismo dato.
  2. **1 – n:** Existe un mismo nombre de fuente de información para asignar el dato, pero con distintos años de procedencia.
  3. **n – n:** Existe una variedad de fuentes de información asociados a distintos años de procedencia. Este caso es el más complejo.
* **Sin procedencia específica:** En este caso, la variable se asigna un valor utilizando criterios metodológicos sin una fuente de procedencia claramente definida.

## 7.2 Selección de variables a estudiar

Para realizar la selección de variables, se utilizó la versión final del MME 2022 (versión metodología[[4]](#footnote-4)) y se analizaron las variables contenidas con la finalidad de identificar y clasificar la procedencia de los datos.

Se analizó la procedencia de los datos en ambos ámbitos del MME (real y ficticia), identificando las variables, su tipo de procedencia, así como también si se encuentra contenida en forma explícita como variable en este.

Los resultados obtenidos producto del análisis realizado se encuentran en las tablas 6 y 7, para los ámbitos real y ficticio respectivamente.

**Tabla 6. Clasificación de procedencias de datos para variables del MME (ámbito real).**

|  |  |  |
| --- | --- | --- |
| Nombre de la variable | Tipo de procedencia | \*Existencia explicita de la procedencia |
| RUT\_REAL | FIJA | NO |
| DV\_RUT | - | NO |
| ID\_GRUPO\_EMPRESAS\_INE | FIJA | NO |
| ID\_GRUPO\_EMPRESAS\_CMF | FIJA | NO |
| NOM\_GRUPO\_EMPRESAS\_INE | FIJA | NO |
| RAZON\_SOCIAL | MIXTA | NO |
| NOM\_REPRESENTANTE\_LEGAL | MIXTA | NO |
| NOMBRE\_FANTASIA | MIXTA | NO |
| REGION\_CUT\_2018 | MIXTA | SI |
| PROVINCIA\_CUT\_2018 | MIXTA | NO |
| COMUNA\_CUT\_2018 | MIXTA | NO |
| DIRECCION | MIXTA | NO |
| INICIO\_DE\_ACTIVIDADES | FIJA | NO |
| SINDICO\_DE\_QUIEBRAS | - | - |
| SECCION\_CIIU4CL | MIXTA | SI |
| DIVISION\_CIIU4CL | MIXTA | NO |
| GRUPO\_CIIU4CL | MIXTA | NO |
| CLASE\_CIIU4CL | MIXTA | NO |
| SUBCLASE\_CIIU4CL | MIXTA | NO |
| ESTRATOS\_F29 | FIJA | NO |
| FONO | MIXTA | SI |
| CORREO\_ELECTRONICO | MIXTA | SI |
| PAGINA\_WEB | MIXTA | SI |
| DIRECTORIO\_PRIORIZADO | - | - |
| GRANDES\_CONTRIBUYENTES | FIJA | NO |
| REGIMEN\_TRIBUTARIO | FIJA | NO |
| CATEGORIA\_TRIBUTARIA | FIJA | NO |
| TIPO\_CONTRIBUYENTE | FIJA | SI |
| SUBTIPO\_CONTRIBUYENTE | FIJA | NO |
| COMERCIO\_EXTERIOR | FIJA | NO |
| MONEDA\_EXTRANJERA | FIJA | NO |
| AFECTO\_A\_IVA | FIJA | NO |
| PERS\_NAT\_CEPAL | - | - |
| TAMAÑO\_OCUPADOS\_2021 | FIJA | NO |
| TIPO\_DE\_REGISTRO | - | - |
| PERS\_NAT\_SII | - | - |
| TERMINO\_GIRO\_20 | FIJA | NO |
| ID | - | - |
| NORMALIZACION | - | - |
| PEE\_2021 | FIJA | NO |
| PEC\_2021 | FIJA | NO |
| PRODUCCION\_ESTADISTICA | FIJA | NO |
| VERIFICACION\_TELEFONICA | FIJA | NO |
| SECTORIAL | FIJA | NO |

\*La existencia explicita de procedencia hace referencia exista una variable especifica llamada procedencia, que detalle el origen del dato de una variable determinada

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Tabla 7. Clasificación de procedencias de datos para variables del MME (ámbito ficticio).**

|  |  |  |
| --- | --- | --- |
| Nombre de la variable | Tipo de procedencia | \*Existencia explicita de la procedencia |
| RUT\_FICTICIO | FIJA | NO |
| TIPO\_DE\_REGISTRO | - | NO |
| MONEDA\_EXTRANJERA | FIJA | NO |
| GRANDES\_CONTRIBUYENTES | FIJA | NO |
| GRUPO\_EMPRESAS\_CMF | FIJA | NO |
| SINDICO\_DE\_QUIEBRAS | FIJA | NO |
| ESTRUCTURALES | FIJA | NO |
| SEGUNDA\_CATEGORIA | FIJA | NO |
| GRUPO\_EMPRESAS\_INE | FIJA | NO |
| REGION\_CUT\_2018 | MIXTA | SI |
| PROVINCIA\_CUT\_2018 | MIXTA | NO |
| COMUNA\_CUT\_2018 | MIXTA | NO |
| SECCION\_CIIU4CL | MIXTA | SI |
| DIVISION\_CIIU4CL | MIXTA | NO |
| GRUPO\_CIIU4CL | MIXTA | NO |
| CLASE\_CIIU4CL | MIXTA | NO |
| SUBCLASE\_CIIU4CL | MIXTA | NO |
| VTA\_$\_ENERO | FIJA | NO |
| VTA\_$\_FEBRERO | FIJA | NO |
| VTA\_$\_MARZO | FIJA | NO |
| VTA\_$\_ABRIL | FIJA | NO |
| VTA\_$\_MAYO | FIJA | NO |
| VTA\_$\_JUNIO | FIJA | NO |
| VTA\_$\_JULIO | FIJA | NO |
| VTA\_$\_AGOSTO | FIJA | NO |
| VTA\_$\_SEPTIEMBRE | FIJA | NO |
| VTA\_$\_OCTUBRE | FIJA | NO |
| VTA\_$\_NOVIEMBRE | FIJA | NO |
| VTA\_$\_DICIEMBRE | FIJA | NO |
| VTA\_ANUAL\_$\_F29\_2022 | - | NO |
| VTA\_ANUAL\_UF\_F29\_2022 | - | NO |
| PRODUCCION\_ESTADISTICA | FIJA | NO |
| COMERCIO\_EXTERIOR | FIJA | NO |
| REGIMEN\_TRIBUTARIO | FIJA | NO |
| DIRECTORIO\_PRIORIZADO | - | NO |
| AFECTO\_A\_IVA | FIJA | NO |
| ESTRATOS\_F29 | - | NO |
| TAMAÑO\_F29 | - | NO |
| CUENTA\_OCUPADOS\_2021 | FIJA | NO |
| TAMAÑO\_OCUPADOS\_2021 | FIJA | NO |
| PERS\_NAT\_CEPAL | - | NO |
| PERS\_NAT\_SII | - | NO |
| CATEGORIA\_TRIBUTARIA | FIJA | NO |
| TERMINO\_GIRO\_20 | FIJA | NO |
| ID | - | NO |
| NOMINA F29 2020 | FIJA | NO |
| NOMINA\_F22\_2020 | FIJA | NO |
| PROM\_TRAB\_HONORARIO\_2021 | FIJA | NO |
| PROM\_TRAB\_OCUPADOS\_2021 | FIJA | NO |
| PEE\_2021 | FIJA | NO |
| PEC\_2021 | FIJA | NO |

\*La existencia explicita de procedencia hace referencia exista una variable especifica llamada procedencia, que detalle el origen del dato de una variable determinada

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

En la tabla 6, se presentan las 44 variables contenidas en el ámbito real del MME, donde se observa que para 36 de ellas se pudo identificar su tipo de procedencia, específicamente 21 fija y 15 mixta. De estas últimas, solamente 5 variables indican su procedencia en forma explícita en el MME, las cuales corresponden a REGION\_CUT\_2018, SECCION\_CIIU4CL, FONO, CORREO\_ELECTRONICO y PAGINA\_WEB.

Respecto a la información contenida en la tabla 7, se observa que de las 51 variables que contienen al MME en el ámbito ficticio, 42 tienen información del tipo de procedencia, de las cuales 34 corresponden a procedencia fija y 8 a mixta. De estas últimas se identifican solamente dos con procedencia explícita, REGION\_CUT\_2018 y SECCION\_CIIU4CL.

Luego de comparar los resultados obtenidos en ambos ámbitos de MME, se observó que las variables identificadas en el ámbito ficticio y que cumplían con los criterios definidos para la selección estaban contenidas en aquellas identificadas en el ámbito real. Las variables REGION\_CUT\_2018 y SECCION\_CIIU4CL son idénticas en ambos ámbitos.

Por lo anterior, se definió seleccionar las variables REGION\_CUT\_2018, SECCION\_CIIU4CL, FONO, CORREO\_ELECTRONICO y PAGINA\_WEB, para el desarrollo del estudio.

## 7.3 Consideraciones de actualización para las variables seleccionadas

Respecto a las variables seleccionadas, en el presente apartado se señalan los periodos de tiempos que se han definido para los procesos de actualización de estas.

Para la variable "sección de actividad económica", las recomendaciones bibliográficas señalan que la actualización debe realizarse – idealmente – una vez al año o cuando se cuente con información actualizada para asignar en esta variable. Para efectos prácticos del MME, este proceso se realizará cada 3 años, en base, a que a la dificultad que presenta la naturaleza de los registros administrativos. Cuando un contribuyente realiza su declaración, este debe realizar un proceso de autoclasificación de la actividad económica principal que realiza, lo que puede provocar errores al no contar con los conocimientos sobre los clasificadores estadísticos que se utilizan para esto, y por tanto, una clasificación errónea.

Otro aspecto que determina el periodo de 3 años para la actualización de la información respecto a la variable de sección de actividad económica tiene relación con la precisión del dato. No se cuenta con un proceso constante de verificación directa con los informantes de las empresas, lo que dificulta la ratificación o no de la actividad económica principal, que es obtenida mediante registros administrativos.

En la misma lógica de los puntos anteriores, otra justificación para los años de actualización de la variable sección de actividad económica es la variedad de información que puede presentar una empresa y sus distintas unidades sobre esta característica. Una casa matriz puede declarar una actividad totalmente distinta que la unidad legal de la misma empresa manifieste. Asimismo, una misma unidad legal puede presentar variadas declaraciones en las cuales expresa que posee actividad económica distintas a las que ya se tenía registrada.

En contraparte, para el resto de las variables que han sido objeto de estudio (teléfono, página web, correo electrónico, sección ciiu4.cl 2012, región cut), se determina que la actualización de los datos se realizará una vez al año, como lo sugiere las recomendaciones internacionales. Cabe señalar que en lo que respecta a la variable “región cut” se aplica el criterio antes señalado, no obstante, si se cuenta con información válida para actualizar antes que se cumpla el año, se deben renovar los datos de forma inmediata.

# DESARROLLO Y RESULTADOS

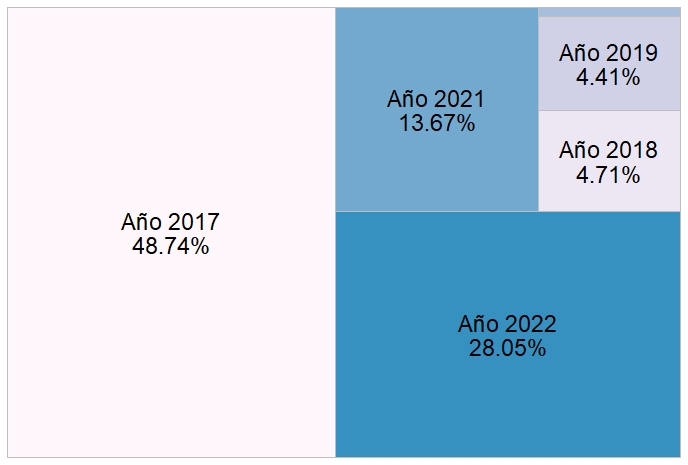
El marco de referencia con el cual se trabajó para el presente estudio consta de un total de 1.254.554 registros. En cuanto a la completitud de la información de cada variable, se observa que para la variable de sección de actividad económica se muestra que no existe información sin procedencia con respecto al total de unidades legales del marco de referencia. Mientras que para el año de referencia 2022 se tienen 351.939 secciones de actividad económica asociados, correspondientes al 28,05% del total. Mientras que, para el año 2021, se tienen 171.537 datos inscritos, representando el 13,67%, además de observar que para el año 2020 se registraron 5.212 actividades económicas asignadas para este año, representando el 0,42%. La participación porcentual se visualiza en la figura 1.

**Tabla 8. Clasificación de procedencias de datos para variable sección de actividad económica del marco de referencia 2022.**

|  |  |  |
| --- | --- | --- |
| **Act. Económica (sección)** | **Cantidad de datos** | **Participación  (%)** |
| Año 2022 | 351.939 | 28,05% |
| Año 2021 | 171.537 | 13,67% |
| Año 2020 | 5.212 | 0,42% |
| Año 2019 | 55.272 | 4,41% |
| Año 2018 | 59.107 | 4,71% |
| Año 2017 | 611.487 | 48,74% |
| **Total** | **1.254.554** | **100%** |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Figura 1. Porcentaje de participación del año de procedencia en la variable sección de actividad económica en el marco de referencia 2022.**



**Año 2020 0.42%**

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

De acuerdo con el punto 7.3, la obsolescencia del dato para la sección de actividad económica en el marco de referencia se localiza entre los años 2017 al 2019 con 725.866 datos representando el 57,86% del total. Esta obsolescencia es debido a la gran masa de registros que contiene el marco de referencia y a la escaza cantidad de analistas que desarrollan la asignación precisa de la actividad económica principal, provocando una dependencia de otra unidad operativa que tiene el contacto directo con el informante. Por otro lado, la impuntualidad de la llegada de las fuentes de información hacia la institución provoca un retraso en la asignación del dato.

Con respecto a la variable teléfono se muestra que entre los años de referencia 2015 a 2022 existe información de 261.096 asociados a teléfono, representando el 20,81% del marco de referencia. Se puede observar en la tabla 9, que para el año de referencia 2022 existen 27.048 datos correspondientes al 10,36%. La gran masa de información corresponde hace referencia al año 2020, el cual se asocian 95.221 datos, representando el 36,47%.

La visualización de la participación de los años de referencias para el dato de teléfono se encuentra en la figura 2.

Tabla 9. Clasificación de procedencias de datos para variable teléfono del marco de referencia 2022.

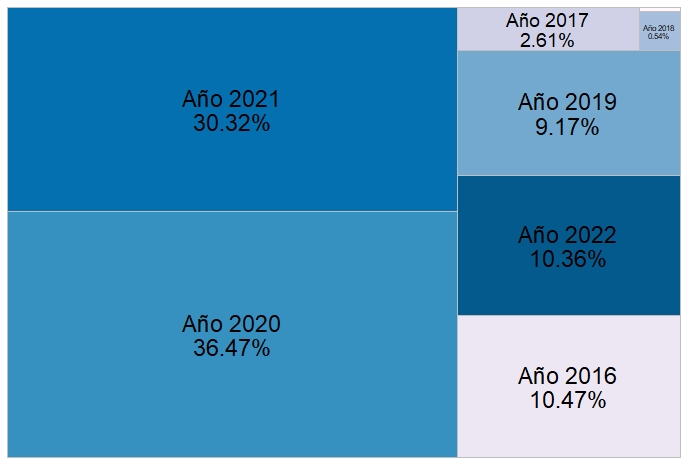
|  |  |  |
| --- | --- | --- |
| **Teléfono** | **Cantidad de datos** | **Participación  (%)** |
| Año 2022 | 27.048 | 10,36% |
| Año 2021 | 79.160 | 30,32% |
| Año 2020 | 95.221 | 36,47% |
| Año 2019 | 23.955 | 9,17% |
| Año 2018 | 1.407 | 0,54% |
| Año 2017 | 6.809 | 2,61% |
| Año 2016 | 27.347 | 10,47% |
| Año 2015 | 149 | 0,06% |
| **Total** | **261.096** | **100%** |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Figura 2. Porcentaje de participación del año de procedencia en la variable teléfono en el marco de referencia 2022.**

**Año 2015**

**0.06%**



Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

Cabe destacar, que la poca información actualizada es debido a la escaza obtención de fuentes de información que permitan asignar datos para la variable de teléfono, repercutiendo directamente en la obsolescencia del dato.

La variable "correo electrónico" tiene una cobertura de 7,30% correspondiente 91.609 datos que oscila entre los años 2015 a 2022, esto se puede observar en la tabla 10. De aquí se desprende que para 5.308 (5,79%) datos corresponden a la asignación del año 2022. El gran volumen de información lo contiene el año 2021 con 51.549 datos, representando el 56,27% del total de datos con correo electrónico 40.675 (11,65%) datos. Para mayor detalle de visualización, se expone en la figura 3.

Tabla 10. Clasificación de procedencias de datos para variable correo electrónico del marco de referencia 2022.

|  |  |  |
| --- | --- | --- |
| **Correo electrónico** | **Cantidad de datos** | **Participación  (%)** |
| **Año 2022** | 5.308 | 5,79% |
| **Año 2021** | 51.549 | 56,27% |
| **Año 2020** | 7.919 | 8,64% |
| **Año 2019** | 9.480 | 10,35% |
| **Año 2018** | 1.808 | 1,97% |
| **Año 2017** | 10.675 | 11,65% |
| **Año 2016** | 4.849 | 5,29% |
| **Año 2015** | 21 | 0,02% |
| **Total** | **91.609** | **100%** |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Figura *3*. Porcentaje de participación del año de procedencia en la variable correo electrónico en el marco de referencia 2022.**

***Gráfico, Gráfico de rectángulos

Descripción generada automáticamente***

**Año 2015**

**0.02%**

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

La poca cobertura de información en la actualización del dato se debe a la escasa información que existe para este tipo de variable.

Para la variable de “región”, asociado a la ubicación geográfica, se tiene una cobertura de información del 92,98% en la asignación del año del dato, es decir 1.166.446 unidades. Se puede observar que en la tabla 11, para el año de referencia 2022 se tienen 1.310 datos correspondientes al 0,11% de la cobertura total de la variable. Por otro lado, el gran volumen de información se asocia al año de referencia 2018 con 863.611 (74,04%) unidades. La visualización de la participación del año de procedencia para esta variable se encuentra en la figura 4.

Tabla 11. Clasificación de procedencias de datos para variable región del marco de referencia 2022.

|  |  |  |
| --- | --- | --- |
| **Región** | **Cantidad de datos** | **Participación  (%)** |
| **Año 2022** | 1.310 | 0,11% |
| **Año 2021** | 69.342 | 5,94% |
| **Año 2020** | 131.646 | 11,29% |
| **Año 2019** | 99.790 | 8,56% |
| **Año 2018** | 863.611 | 74,04% |
| **Año 2017** | 747 | 0,06% |
| **Total** | **1.166.446** | **100%** |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Figura 4. Porcentaje de participación del año de procedencia en la variable región en el marco de referencia 2022.**

***Gráfico, Gráfico de rectángulos

Descripción generada automáticamente***

**Año 2017 0.06%**

**Año 2022 0.11%**

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

Los inconvenientes directos para esta variable se deben a la exigua cantidad de analistas que asignan esta variable. Cabe destacar que para el marco de referencia del año 2018 fue la primera vez que se realiza un marco maestro agregando varias fuentes de información y esto provocó una sobre cobertura en las unidades legales asociadas al producto.

Para la variable “página web” se tiene una cobertura de información del 3,67% en la asignación de las procedencias del dato, es decir, 46.012 unidades con años oscilando entre 2016 al 2022. De lo anterior, se desprende que 3.333 datos corresponden al año 2022, representando el 7,24% del total de la información. Por otra parte, la robustez de la cantidad de información asociada a la página web se encuentra para el año de referencia 2021, con 16.439 (35,73%) datos.

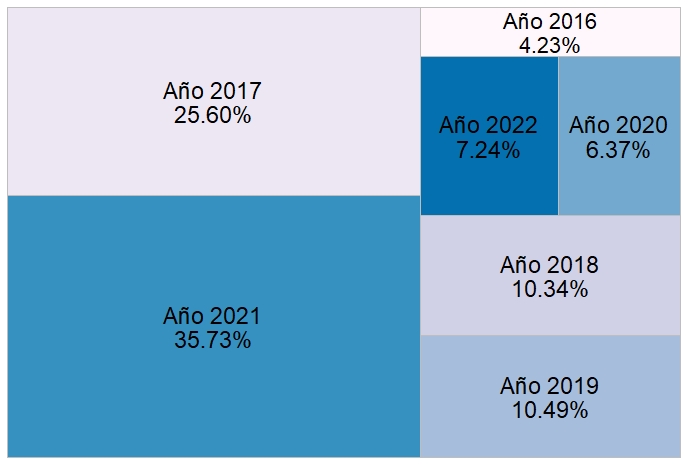
Los porcentajes son expuestos a través de la figura 5 y la información al detalle se encuentra en la tabla 12.

Tabla 12. Clasificación de procedencias de datos para variable de página web del marco de referencia 2022.

|  |  |  |
| --- | --- | --- |
| **Página web** | **Cantidad de datos** | **Participación  (%)** |
| **Año 2022** | 3.333 | 7,24% |
| **Año 2021** | 16.439 | 35,73% |
| **Año 2020** | 2.929 | 6,37% |
| **Año 2019** | 4.826 | 10,49% |
| **Año 2018** | 4.756 | 10,34% |
| **Año 2017** | 11.781 | 25,60% |
| **Año 2016** | 1.948 | 4,23% |
| **Total** | **46.012** | **100%** |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

**Figura 5. Porcentaje de participación del año de procedencia en la variable de página web en el marco de referencia 2022.**



Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

En suma, a lo expuesto anteriormente, la presente tabla resume los resultados obtenidos a partir del estudio realizado.

**Tabla 9. Tabla resumen de los resultados del análisis de obsolescencia.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | Cant. Datos obsoletos | Cant. Datos no obsoletos | Criterio de actualización |
| Act. Económica  (sección) | 725.866 | 528.688 | Trianual |
| Teléfono | 234.048 | 27.048 | De acuerdo con la llegada de las fuentes de información  (anual) |
| Correo electrónico | 86.301 | 5.308 |
| Región | 1.165.136 | 1.310 |
| Página web | 42.679 | 3.333 |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

# CONCLUSIÓN

En base al análisis de las variables elegidas en el marco de referencia 2022 y orientado hacia el objetivo de este documento se observa que de las cinco variables analizadas, solo “sección” de actividad económica y “región” de ubicación geográfica tienen una cobertura superior al 90% en los años de referencia en las procedencias del dato. Para el caso de actividad económica, y de acuerdo con las recomendaciones internacionales, tiene un 58% de unidades con información obsoleta, oscilando entre los años 2017 a 2019, llegando al peak crítico de mostrar 611.487 datos correspondientes al año 2017. Por otra parte, esta variable se asocia directamente a la precisión del dato, por lo que se recomienda actualizar la información de menor proporción, es decir, para el año 2019 se actualizarían 55.272 datos, seguido de los asociados al año 2018 siendo estos 59.107 registros y respecto al caso crítico de información del año 2017, se sugiere como última instancia actualizarlos debido al volumen de cobertura.

La variable “región” muestra una cobertura del 92,98% respecto al marco de referencia 2022, sin embargo, casi la totalidad de la información se encuentra obsoleta, oscilando entre los años 2017 y 2021. No obstante, la realidad institucional que recopila la información (INE), en variadas ocasiones presenta problemáticas con las fuentes de información que actualizan el dato, ya que llegan con retraso y no son oportunas. Bajo este contexto, se recomienda actualizar el dato de acuerdo a la llegada de las fuentes de información declaradas, y en caso de no contar con la condición anterior, se propone que se aborde de la siguiente manera: datos asociados al año 2017 (0,06%), año 2021 (5,94%), año 2019 (8,56%), año 2020 (11,29%) y finalmente año 2018 (74,04%).

Por otro lado, para la variable “teléfono” se muestra una cobertura del 57,86% respecto al marco de referencia 2022, los datos obsoletos abarcan el 89,64% correspondiente 234.048 unidades y oscilan entre los años 2015 al 2021. Este caso es en mayor medida crítico, ya que no existe información precisa en las escasas fuentes de información utilizadas para la actualización del dato, además, no existen protocolos para imputar el dato obsoleto. Se recomienda actualizar la información de acuerdo a la indicación sugerida en la variable región.

Para las variables “correo electrónico” y “página web”, ambas con baja cobertura (7,30% y 3,67% respectivamente) respecto del marco de referencia 2022. Para el caso de la variable primera variable, la obsolescencia del dato oscila entre los años 2015 al 2021 con 86.301 datos, representando el 94,19% del total. Mientras que, para la segunda variable, los datos obsoletos fluctúan entre el año 2016 y 2021 representando el 92,76% del total, es decir, 42.679 datos.

Se recomienda actualizar el dato de las variables asociadas a oportunidad, de acuerdo con la llegada de la fuente de información, de lo contrario, seguir pasos anteriores.

De manera general, se sugiere poseer una matriz de procedencias por cada una de las variables contenidas en el marco de referencia, o en el marco maestro de empresas (ambos ámbitos), para que, de esta manera, se pueda realizar el análisis de obsolescencia del dato con la totalidad de las variables.

# ANEXOS

## 8.1 Arboles de asignación para la información del MME

Para que las variables del MME contenga información relevante – en sus ámbitos reales y ficticios – se debe emplear los árboles de decisión. Estos, se comprenden como una herramienta para priorizar la utilización de diferentes fuentes de información en la asignación de información para las variables que conforman el marco.

En primera instancia, se debe tener en cuenta que se identifican distintos tipos de registros al momento de realizar la asignación. En primer lugar, se encuentran los registros nuevos, estos registros corresponden a UL que se agregaron en el trimestre, es decir, no estaban incluidas en versiones previas del MME. En segundo lugar, se encuentran los registros continuos, los cuales, a diferencia de los anteriores, existen en el MME, al menos, desde la versión anterior y finalmente se identifican los registros priorizados

## 8.2 Arboles de asignación para variable actividad económica

A continuación, se muestra el flujo del proceso de asignación de información de la variable actividad económica y luego se explica la forma en que el proceso es ejecutado.

Diagrama

Descripción generada automáticamente**Figura 1: Flujo del proceso de asignación de la variable Actividad Económica**

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

Para asignar la información de la variable de actividad económica, se deben identificar los registros nuevos y continuos y, al mismo tiempo, identificar las personas naturales y jurídicas. Con los registros identificados se crea una matriz de información en SQL en la que se despliega toda la información que aportan las fuentes de información disponibles. La matriz contiene las siguientes columnas:

* RUT
* Dígito verificador
* Razón social asignada desde el último Catastro de Contribuyentes disponible
* Actividad económica desde los 2 últimos Formularios N°29 disponibles
* Actividad económica desde los 2 últimos Formularios N°22 disponibles
* Primeras 10 actividades económicas desde último Catastro de Contribuyentes
* Actividad económica de la Bases Sectoriales
* Actividad económica de las Encuestas Estructurales
* Actividad económica de la base de Grupo de Empresas.

La matriz creada es exportada a MS Access en donde se incluyen marcas que identifiquen las unidades legales que corresponden a personas jurídicas y naturales, lo que permitirá utilizar una serie de criterios que se detallarán más adelante en este mismo punto.

Las fuentes de información considerados para la asignación de la actividad económica corresponden, en su mayoría, a registros administrativos del SII. Estos registros contienen autoclasificaciones de actividad económica realizadas por cada contribuyente y, en consecuencia, pueden presentar errores desde su origen. En el caso del registro administrativo catastro de giros, contiene más de una actividad económica sin un orden de prioridad, por lo tanto, no es posible saber cuál es su actividad principal. Considerando ambos casos, no es posible aplicar el árbol de asignación a todos los registros como en la mayor parte de las variables del MME, siendo necesario aplicar una serie de criterios y comparaciones que permitan determinar la mejor opción disponible para actualizar el marco.

Para la asignación de registros nuevos se identifican las ofertas de datos para cada unidad, en caso de existir más de una, se compara la información disponible entre las fuentes del SII y se determina el dato a utilizar. Además de las comparaciones anteriores, se aplican los siguientes criterios expresados en las siguientes tablas y listados:

**Tabla 10: Listado de actividades económicas exclusivas de personas jurídicas con respaldo legal.**

|  |  |  |
| --- | --- | --- |
| Actividades Económicas | CIIU4.CL 2012 | Respaldo ley |
| Comité de agua potable | 36000 | Ley 19.418, Art. 4° |
| Junta de vecinos | 94990 | Ley 19.418, Art. 4° |
| Unidades vecinales | 94990 | Ley 19.418, Art. 4° |
| Sindicatos | 94200 | Ley 19.069, Art. 11° |
| Colegio de profesionales | 94120 | Decreto 3274 |
| Bancos | 64191 | D.F.L. Núm. 3 |
| Organizaciones políticas | 94920 | Ley 18.603, Art. 1° |
| Asociaciones gremiales | 94110 | Decreto Ley 2757 |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

Criterios de asignación:

1. Congruencia entre razón social y código de actividad.
2. Congruencia entre nombre de fantasía y código de actividad.
3. Priorizar códigos distintos a los residuales (Ej. 96090).
4. Consistencia entre personas naturales y actividades económicas cuestionables (Hospitales, Fabricación de gas, Transporte urbano y suburbano de pasajeros vía metro y metro tren, entre otros).

**Tabla 11: Listado de actividades económicas exclusivas de personas jurídicas y entidades con actividades únicas.**

|  |  |  |
| --- | --- | --- |
| Sector Económico | CIIU4CL 2012 | Justificación |
| Actividades de la administración pública en general | 84110 |  |
| Regulación de las actividades de organismos que prestan servicios sanitarios | 84120 |  |
| Regulación y facilitación de la actividad económica | 84130 |  |
| Relaciones exteriores | 84210 |  |
| Actividades de defensa | 84220 |  |
| Actividades de mantenimiento del orden público y de seguridad | 84230 |  |
| Fondo Nacional de Salud  (Fonasa 61.603.000-0) | 84301 |  |
| Instituto de Previsión Social  (IPS 61.979.440-0) | 84302 |  |
| Otros planes de seguridad social de afiliación obligatoria n.c.p. | 84309 |  |
| Actividades de organizaciones y órganos extraterritoriales | 99000 |  |
| Banco Central de Chile (BCCH 97.029.000-1) | 64110 | Única entidad que debe poseer esta Actividad Económica |
| Universidades Privadas o Estatales | 8531 |  |
| Correos de Chile (60.503.000-9) | 53100 | Única entidad que debe poseer esta Actividad Económica |

Fuente: Elaboración del Subdepartamento de Diseños de Marcos y Muestras.

Continuando con la asignación de registros nuevos, para las personas naturales, se utilizan las fuentes de información del árbol de asignación que no corresponden a RR.AA., los que fueron auditados por la Sección de Nomenclaturas (Encuestas Estructurales) y los que fueron revisados por el SMM (Base sectorial y Grupo de Empresas), estos mantienen el mismo orden de prioridad que en el árbol de asignación. Luego, se asignan los registros que tengan solo una oferta de actividad económica, independiente de que la misma actividad se encuentre en una o más fuentes de información. Esta asignación, se valida utilizando los criterios que se indicaron anteriormente, aunque, también existe la posibilidad de que se utilice la información a un menor nivel. Para los registros que quedan sin asignación, se continúa el proceso, esta vez, considerando la mayor frecuencia con que aparecen las actividades. En caso de no conseguir asignación, los registros quedan con la nomenclatura 00000. La procedencia del dato se verá reflejada según sea el caso, con una o más fuentes.

Para las personas jurídicas se consideran las mismas fuentes de información que para las personas naturales y en el mismo orden, luego se asignan los registros que resulten con una única oferta de actividad, la que puede estar en una o más fuentes. Además de la selección anterior, la actividad económica se compara con la razón social cuando esta haga alguna referencia clara a la actividad económica desarrollada. Al igual que para las personas naturales, la actividad podría ser asignada parcialmente, esto significa, disminuir el nivel según sea el caso.

Posteriormente, se realiza una asignación para los registros nuevos que queden sin dato utilizando la producción estadística coyuntural, la razón social y/o el nombre de fantasía, esto considerando las características de las fuentes de información y su imposibilidad de utilizarlo directamente. La procedencia reflejará la o las fuentes que contengan el dato asignado.

Como segunda fase se realiza la asignación de los registros continuos, los que son revisados de forma paulatina en los trimestres sucesivos cuando es posible hacerlo, en este caso, también se crea una matriz con las ofertas de datos considerando los dos últimos años disponibles de las siguientes fuentes de información:

* Grupo de Empresas
* F29
* F22
* Catastro de Contribuyentes
* Producción Estadística Coyuntural
* Bases Sectoriales
* Verificación Telefónica

Además de las fuentes anteriores, se utiliza como referencia la actividad económica que estos registros ya tienen asignada en el MME. De igual manera que en el paso anterior, se utilizan variables de referencia para validar u orientar la selección del dato. Estas variables son:

* Razón Social.
* Nombre de fantasía.
* Estrato de Venta.
* Marca directorio priorizado.
* procedencia de la actividad económica ya existente en el MME.

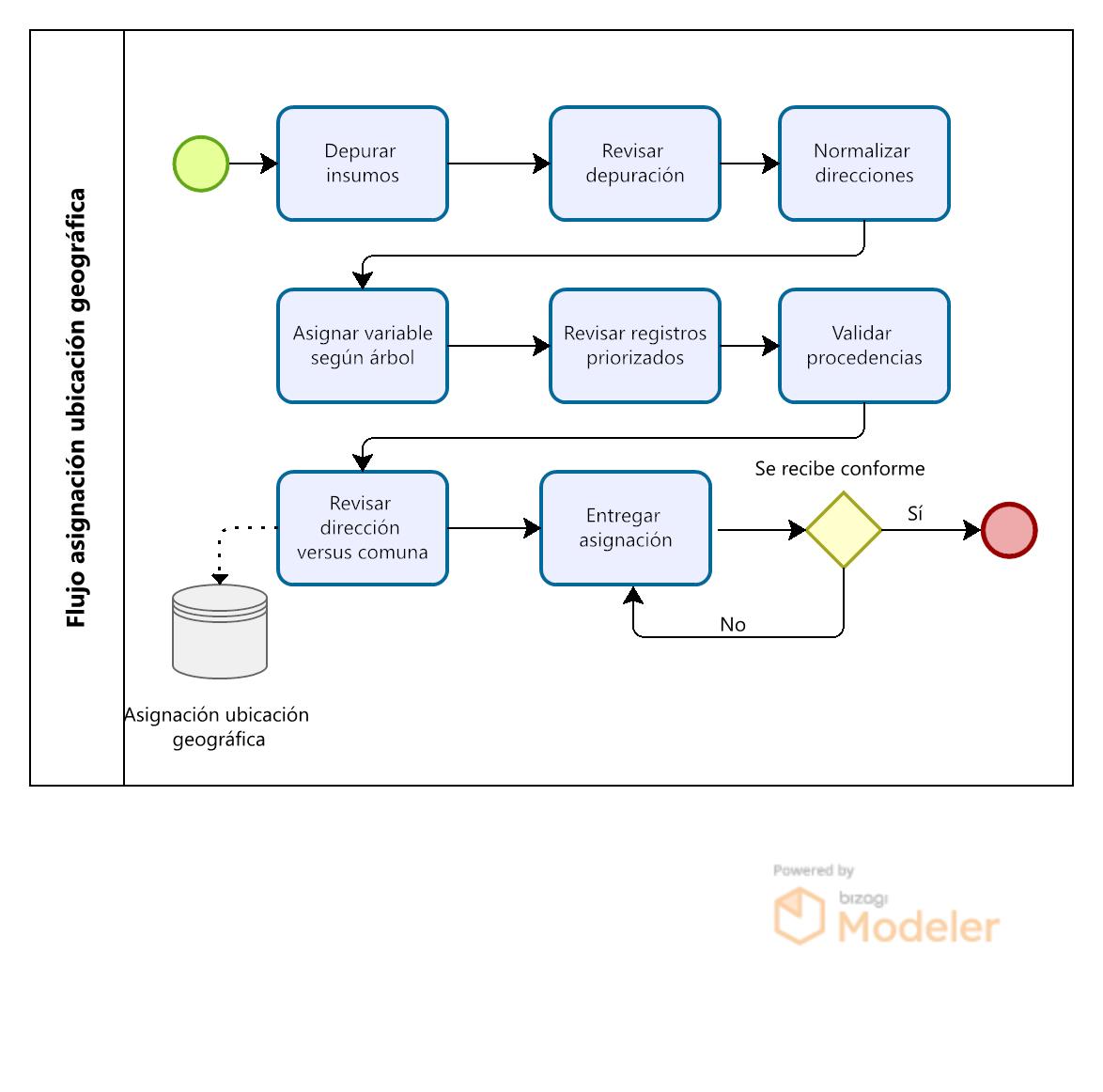
La matriz de asignación para los registros continuos se analiza en MS Access y se crean distintas columnas adicionales dependiendo de los cruces y comparaciones que se realicen, la clasificación existente se puede confirmar en caso que las fuentes entreguen el mismo dato ya asignado, se mantiene en caso que las fuentes de información no logren aportar un nuevo dato adecuado para ser incluido en el MME, se actualiza en caso que las nuevas fuentes coincidan en la nueva clasificación o se complementa debido a que una o varias fuentes presentan un nivel de desagregación del dato más alto.

Una vez finalizado el proceso de asignación de actividad económica para todos los registros, se realiza una revisión de lo obtenido cuidando la validez del dato según la tabla de correspondencia del clasificador CIIU4.CL 2012 además de la congruencia entre niveles de códigos y sección.

Es importante mencionar que el proceso de asignación de actividad económica se realiza en el ámbito real, debido a que solo de esta forma es posible validar y contrastar la información con otras variables como las de identificación. Una vez cerrado el proceso se solicita una conversión del listado de RUT y su clasificación final, de esta forma es posible reflejar resultados idénticos en los dos ámbitos del MME.

## 8.3 Arboles de asignación para variable ubicación geográfica

A continuación, se muestra el flujo del proceso de asignación de información de la variable ubicación geográfica y luego se explica la forma en que el proceso es ejecutado.

**Figura 2: Flujo de asignación ubicación geográfica.**

Fuente: generación propia Sección Marcos Maestros.

La Ubicación Geográfica se conforma por las variables Dirección[[5]](#footnote-5) y CUT[[6]](#footnote-6) desagregado, viéndose reflejada en el MME en los siguientes campos:

* Dirección
* Region\_CUT
* Provincia\_CUT
* Comuna\_CUT

El proceso de asignación de Ubicación Geográfica, al igual que el resto de las variables, comienza por la limpieza de las fuentes de información, esta limpieza consiste en identificar y eliminar los siguientes elementos:

* Espacios al inicio y final del dato.
* Doble o triple espacios.
* Tildes.
* Mantener texto en mayúsculas.
* Caracteres que no formen parte real de la dirección como comas, guiones, ceros no válidos, etc.

Posterior a este proceso, se realiza una revisión de las fuentes de información, no obstante, considerando el tiempo disponible en el trimestre, el proceso se modifica con la finalidad de priorizar el procesamiento de los datos que serán utilizados finalmente en el MME y no de las fuentes de información de forma particular e integral. Para facilitar reducir el proceso, mediante el árbol de asignación, es posible identificar los registros nuevos coincidentes con cada una de las fuentes que serán asignadas al MME, además de identificar los registros priorizados del marco.

Identificados los registros, se valida que en las fuentes de información los RUT sean únicos. Si hay RUT repetidos y no son demasiados, se revisan para dejarlos en únicos. Si son muchos, se identifican para, posteriormente, tratar de corregirlos, de lo contrario solo se trabaja con los RUT únicos.

Por el lado de la variable propiamente tal, se revisa que los códigos CUT sean válidos, y se debe eliminar cualquier simbología extraña. Si el registro no tiene dato, se actualiza con la nomenclatura “S/I”. Finalizado este proceso, se continúa con la siguiente etapa que es la normalización.

El proceso de normalización es realizado por la Unidad de Geografía y Actualización Cartográfica y consiste en estandarizar el dato contenido en el campo Dirección, corrigiendo la escritura y cuando es posible identificando el tipo de vía de la ubicación geográfica. El normalizador de direcciones funciona de forma semiautomatizada y utiliza un diccionario de datos para la detección y corrección de cadenas de texto (direcciones). Luego de la normalización, se debe realizar una revisión del resultado obtenido que consiste en lo siguiente:

* Validar que los RUT enviados sean los mismos que se reciben.
* Detectar la eliminación, reemplazo o adición de datos que puedan alterar la dirección original.
* Validar el CUT.
* Revisar que los nombres numéricos de las calles se escriban con letras para evitar confusión con la numeración de la dirección.
* Revisar que los tipos de vía estén estandarizados y exista solo uno por dirección.

Existe un delta de direcciones que no serán normalizadas, ya que estos errores no se encuentran en el diccionario de datos, este debe ser revisado y estandarizado de forma manual cuando es posible.

Finalizado el proceso de normalización, corresponde realizar una asignación según el árbol de asignación de variables, en caso de haber realizado una asignación previa a la normalización, esta será validada. A continuación, se revisan uno a uno los registros priorizados y se estandarizan las direcciones que no fueron cubiertas por el proceso de normalización.

Posterior a la asignación de los registros nuevos, se validan las procedencias de los registros continuos, si la dirección y CUT es igual a la que indican las fuentes actuales, se indica como procedencia la fuente correspondiente según el orden de asignación dictado por el árbol. El proceso requiere ser realizado de forma manual ya que cualquier carácter, visible o no, creará una diferencia entre las cadenas de texto, lo que impediría la actualización automatizada de la fuente de forma efectiva.

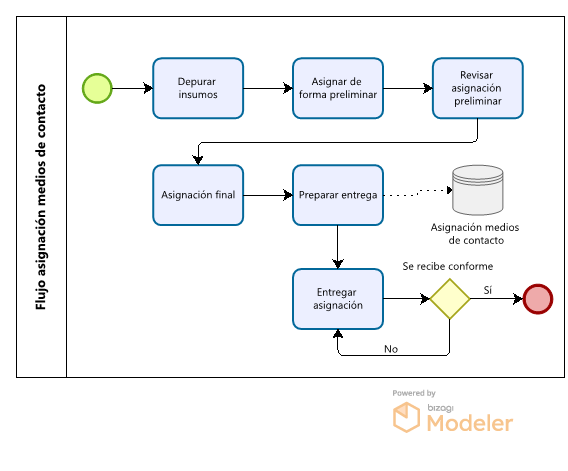
Como paso final se seleccionan los registros priorizados y se valida que las direcciones y el CUT sean coherentes, para esto, se cuenta con el conocimiento del propio analista además de apoyarse en herramientas como Google Maps, Here y WeGo. Una vez finalizada la revisión, se realiza la entrega de la BBDD para ser actualizada en el MME.

Al igual que para la variable actividad económica, se realizan análisis de la ubicación geográfica con la finalidad de mejorar la calidad del dato, estos estudios apuntan principalmente a reconocer y estandarizar direcciones que cumplen características similares, para esto, se identifican direcciones que correspondan a algún centro comercial (mall), aeropuertos, locales comerciales que se encuentran en estaciones de metro, direcciones rurales y luego se estandariza la dirección, permitiendo uniformar la información. Además del análisis anterior, se creó una guía de apoyo que es útil para identificar distintas rutas que son conocidas por un nombre informal como “Autopista del Maipo” y agregar su nombre formal, en este caso “Ruta 5”. Esta información se comparte con el equipo de Verificación Telefónica con la finalidad de que puedan registrar las direcciones con una estructura estandarizada que permita disminuir las correcciones y tener una mayor claridad en la información.

## 8.4 Arboles de asignación para variable Medios de contacto

Es el siguiente punto, se presenta el flujo de trabajo utilizado. Es un flujo común para las tres variables, solo difiere en las utilizadas y en la variable a actualizar.

**Figura 9: Flujo del Proceso de Asignación de Variables de Medios de Contacto**



Fuente: generación propia Sección Marcos Maestros.

El proceso de asignación de información de medios de contacto comienza por realizar una revisión inicial que consiste en seleccionar las fuentes a ser utilizadas, verificar que todos los registros contengan RUT, que estos sean válidos y ocultar otros campos que no tengan relación con las variables a utilizar. Como segundo paso, las fuentes son separadas por variable (Fono, Correo Electrónico y/o Página Web) dependiendo del dato que contengan. Posteriormente, se realizan validaciones por variable orientadas a cuidar la integridad, consistencia y formato de la información.

Con las bases separadas por variable se realizan depuraciones específicas. En el caso de la variable fono se revisa que la información no contenga puntos, guiones, letras, ceros al comienzo del campo, símbolos, espacios en cualquier parte del campo, se debe unir el código de área con el número de teléfono.

En el caso del correo electrónico, se revisa que la información no contenga espacios, se valida que el dato no corresponda a otras variables tales como número de teléfono, dirección, página web y se revisa que el dato no contenga dos o más arrobas, la extensión del mail debe estar completa y finalmente todos los correos deben estar escritos en minúscula.

Finalmente, para la variable Página Web se revisa que no existan espacios, se valida que no haya información que no corresponda al campo como números de teléfono, direcciones y correos electrónicos. Para terminar, se deja la información del campo Página Web en minúscula.

Las fuentes de información pueden contener RUT repetidos, por lo que estos se separan de los RUT únicos con la finalidad de verificar si la información que entregan los RUT repetidos también se repite o es distinta. Para una correcta asignación, es necesario tener RUT únicos de forma de tener el control de la información que se incluirá en el marco, en caso contrario, cada software que se utilice para realizar la asignación seleccionará, por defecto, alguno de los datos que ofrecen los RUT repetidos.

Cuando existen RUT repetidos con información distinta para una misma variable, los datos se ordenan como una matriz, de la siguiente forma: RUT, DV, Dato\_1, Dato\_2, etc. solo el Dato\_1 es el que se asigna al MME, el resto de la información se reserva y se incluye en una tabla satélite que contiene la información disponible que no se verá reflejada en el MME (solo variable fono).

Una vez ordenados los datos se realiza una asignación preliminar siguiendo el árbol para cada una de las variables. Posterior a la asignación realizada, se realiza una revisión del resultado, en caso de encontrar inconsistencias se deberá determinar la solución adecuada, esta considera distintas soluciones, que van desde mejorar la fuente de información depurada, hasta re ejecutar el proceso completo o parte de él.

Luego de la revisión y mejora del proceso, se considera que la asignación final está terminada y se cuantifica la información resultante. Finalmente, se revisa el formato de la base de datos y se realiza la entrega para su actualización en el MME.

Una vez finalizado al proceso, se crean productos para ser gestionados por Verificación Telefónica, estos consideran registros que no obtuvieron asignación de información, otros que puedan tener procedencias antiguas y finalmente registros con datos que pueden estar incompletos. Estos productos se generan con el fin de conseguir información actualizada.

1. MME ámbito real se define principalmente porque la nómina del MME es el RUT del contribuyente proporcionado en el SII. Mientras que el MME ámbito ficticio, la nómina del contribuyente es encriptada por el SII para guardar el secreto tributario del contribuyente. [↑](#footnote-ref-1)
2. Esta marca se basa a que la CEPAL recomienda que un directorio común básico deberá contener todas las empresas o personas jurídicas (sociedades, organismos sin fines de lucro y demás), así como todas las unidades del Estado y de las administraciones regionales y locales, que producen bienes o servicios y que tienen cinco o más personas empleadas. CEPAL,2013. "COMPENDIO DE RECOMENDACIONES TÉCNICAS Y BUENAS PRÁCTICAS PARA EL DESARROLLO DE LOS DIRECTORIOS DE EMPRESAS Y ESTABLECIMIENTOS ", pág.6. [↑](#footnote-ref-2)
3. Esta marca se basa a que CEPAL recomienda que un directorio común básico deberá contener todas las empresas o personas jurídicas (sociedades, organismos sin fines de lucro y demás), así como todas las unidades del Estado y de las administraciones regionales y locales, que producen bienes o servicios y que tienen cinco o más personas empleadas. [↑](#footnote-ref-3)
4. MME – 2022 (metodología) es un marco maestro de empresas que contiene variables de procedencias explícitas de algunas variables contenidas dentro del MME, siendo éstas de uso reservado del departamento que construye el MME. [↑](#footnote-ref-4)
5. contiene el tipo de vía (Avenida, pasaje, calle, etc.), nombre, numeración y aclaratoria (depto., lote, parcela, etc.) asociada a la unidad legal [↑](#footnote-ref-5)
6. Se utiliza la última versión publicada por la SUBDERE. Actualmente versión 2017. [↑](#footnote-ref-6)