



UNAM



RED·TIC

Red de Responsables TIC

U N A M



Recomendaciones generales de calidad de datos

(2ª versión)

Índice

INTRODUCCIÓN	3
PROPÓSITO	3
GLOSARIO	3
CARACTERÍSTICAS DE CALIDAD DE LOS DATOS	4
RECOMENDACIONES	6
<i>Documentación de los requisitos de calidad de datos</i>	6
<i>Información estadística universitaria</i>	8
<i>Monitoreo de la calidad de los datos</i>	8
<i>Diseño de bases de datos relacionales</i>	9
<i>Sistemas</i>	12
<i>Limpieza de los datos</i>	12
<i>Migración de datos</i>	15
<i>Calidad de datos abiertos</i>	17
<i>Indicadores de actividad y desempeño</i>	18
<i>La información confiable y comparable</i>	19
<i>El flujo sistemático de la información</i>	19
<i>Métricas de calidad de los indicadores</i>	20
ANEXO 1. DESCRIPCIÓN DE LOS ATRIBUTOS DE CALIDAD DE LOS DATOS	20
ANEXO 2. MÉTRICAS DE ACUERDO CON EL ESTÁNDAR ISO/IEC 25024:2015	22
ANEXO 3. EJEMPLOS DE FICHAS TÉCNICAS DE INDICADORES DE ACTIVIDAD Y DESEMPEÑO	26
REFERENCIAS	28
CRÉDITOS	29

Recomendaciones generales de calidad de datos

Introducción

Derivado de sus atribuciones y funciones, las diferentes áreas universitarias recopilan, generan, almacenan, transmiten y aseguran diariamente un gran número de datos recurrentes organizados a partir de diferentes fuentes, plataformas y servicios. La calidad de los datos es un elemento primordial para la planeación, seguimiento y ejecución de sus actividades tanto sustantivas como cotidianas, ya que permite la toma de decisiones a nivel operativo, táctico y estratégico a partir de esos datos.

La calidad de datos se refiere a los procesos y técnicas enfocadas a mejorar la eficacia, integridad, auditabilidad, oportunidad y confiabilidad de los datos existentes y se establece respecto al grado de cumplimiento en el que los datos satisfacen las necesidades del área universitaria y de la Universidad.

Administrar y mejorar la calidad de los datos puede tener un impacto directo en su uso tanto por personas como por los sistemas, al ofrecer una mayor coherencia semántica y sintáctica contribuyendo a contar con información útil, realizar los procesos académicos y administrativos de la Universidad, apoyar en la gestión y operación, así como facilitar la cooperación e intercambio de información entre las áreas universitarias aprovechando la infraestructura tecnológica existente.

Una mala calidad de los datos puede conducir a errores en la toma de decisiones al generarse análisis y resultados equivocados, pudiendo ocasionar un incumplimiento normativo.

Propósito

Este documento tiene la finalidad de servir como guía acerca del establecimiento de recomendaciones de calidad de los datos para los responsables TIC o aquel personal que interviene en la generación, actualización, recopilación, compartición y almacenamiento de la información en la Universidad con la intención de fortalecer sus características, como es su confiabilidad, disponibilidad y seguridad.

Glosario

Datos oscuros. Gartner en su glosario de tecnologías de información define los datos oscuros o *dark data* como los activos de información que las organizaciones recopilan, procesan y almacenan durante las actividades comerciales regulares, pero generalmente no los utilizan para otros fines (por ejemplo, análisis, relaciones comerciales y monetización directa). Las organizaciones a menudo almacenan estos datos solo con propósitos de cumplimiento.

Almacenar y proteger los datos habitualmente genera más gastos y a veces más riesgos que valor (Gartner, s.f.).

Datos sucios. Son aquellos que están incompletos, erróneos, inexactos o que no cumplen con cualquiera de las características de calidad que le son asignadas.

ETL - Extraer, Transformar y Cargar. (*Extract, Transform and Load* en inglés). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, *data mart*, o *data warehouse* para su análisis.

Indicadores de actividad. Datos cualitativos o cuantitativos (magnitud) que se consideran como signo o indicio de la presencia de una condición, con características o propiedades definidas. Para que un dato o un estadístico pueda ser considerado como indicador, dicho dato debe tener un punto de referencia contra el cual pueda juzgarse.

Indicadores de desempeño. Son razones o proporciones, porcentajes, tasas u otros valores cuantitativos que permiten a los tomadores de decisiones evaluar el desempeño en alguna actividad sustantiva o estratégica, así como para realizar análisis comparativos posteriores.

Limpieza de datos (*Data cleaning* en inglés). Se refiere a la depuración de datos que consiste en el proceso de detección, validación, corrección, sustitución o eliminación de datos incorrectos, inexactos, incompletos, inconsistentes, falsos o corruptos de una base de datos o un conjunto de datos.

Métricas. Medida de calidad en la que un dato posee un atributo dado, por ejemplo, los atributos o características de calidad de los datos como son exactitud, completitud, integridad, relevancia, por mencionar algunos.

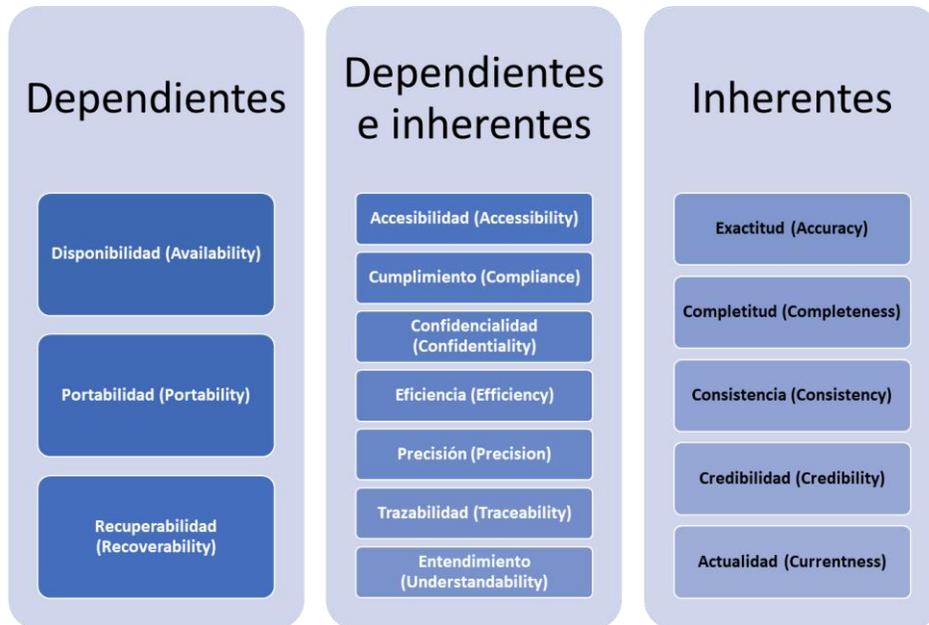
Perfilado. Se trata de una serie de técnicas para identificar datos erróneos, datos nulos, datos incompletos, datos sin integridad referencial, datos que no se ajustan al formato requerido, patrones de información de negocio, tendencias, medias, desviaciones estándares, entre otros.

Características de calidad de los datos

La norma *ISO 25012:2008* define la calidad de datos como el grado en el que los datos satisfacen las necesidades de los usuarios, establece un modelo de calidad para los datos conservados en un formato estructurado para ser usado por un sistema informático. Adicionalmente define 15 características de calidad de datos que reflejan un aspecto específico, agrupándolas en tres dimensiones:

- ◆ **Dependientes.** La calidad depende del sistema que hace uso de los datos.
- ◆ **Inherentes.** Tienen el potencial intrínseco para satisfacer las necesidades establecidas cuando los datos son utilizados bajo condiciones específicas, es decir, “se refiere al dato en sí y a la correspondencia del mismo con la información del mundo real” (Yanes, 2019).

- ◆ **Dependientes e inherentes (compartidas).** Son características de la calidad de datos comunes a ambos enfoques.



La descripción de cada atributo se encuentra en el Anexo 1: Descripción de los atributos de calidad de los datos.

De acuerdo con lo anterior, algunas características de la calidad de datos que son comunes a las áreas universitarias son:

◆ Disponibilidad

- **Accesibilidad:** Las interfaces de acceso a datos facilitan hacerlos públicos de manera que puedan ser consultados en condiciones de igualdad por diferentes tipos de perfiles de manera sencilla.
- **Oportunidad:** Los datos se actualizan regularmente y el intervalo de tiempo entre la recopilación y el procesamiento de los datos hasta la liberación cumple los requisitos para que resulten útiles.

◆ Usabilidad

- **Credibilidad:** La usabilidad se refiere a la cualidad de sí un dato puede ser utilizado en un contexto específico, para que esto suceda una variable importante es la credibilidad, en la cual los datos provienen de una fuente especializada de la Universidad o del país, son auditados regularmente, cumplen con normas generales y se puede verificar su exactitud en el contenido. Existe un rango de valores para los datos: conocidos o aceptables.

◆ Confiabilidad

- Exactitud: Los datos proporcionados son precisos, la representación de datos refleja el estado real de la información de origen y su representación no genera ambigüedades.
- Consistencia: Los datos coinciden en sus conceptos, dominios y formatos después de ser procesados. Son verificables durante un tiempo.
- Integridad: Los datos tienen un formato claro y que cumple con los criterios, tienen integridad estructural y de contenido.
- Completitud: Hasta qué punto no existe una deficiencia de un componente del sistema o de la base de datos que afectará la precisión e integridad de datos y su uso.

◆ Pertinencia

- Convivencia: Los datos recogidos presentan cierta relación y están dentro del ámbito del tema que los usuarios requieren o presentan coincidencias.

◆ Calidad de presentación

- Legibilidad: Los datos son claros y comprensibles, satisfacen las necesidades del usuario. Así como su descripción, clasificación y contenido son fáciles de entender (Jacintogr, 2020).

Los problemas de calidad de datos mayoritariamente se materializan en tres contextos diferentes dentro de las dimensiones como amenazas:

- ◆ Cuando existen anomalías en una única fuente de datos.
- ◆ Cuando datos no estructurados son migrados a una fuente de datos estructurada.
- ◆ Cuando se realizan procesos de integración en una única fuente de datos de información proveniente de diversas fuentes.

Recomendaciones

Documentación de los requisitos de calidad de datos

Se sugiere describir las características de los datos que sean relevantes para revisar su calidad, incluso establecer métricas “a través de las cuales se define la forma en la que cada dimensión es medida” (Yanes, 2019). Para ello se deben tomar en cuenta los requisitos identificados del usuario (anexo 2. Métricas del estándar ISO/IEC 25024:2015 *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality*).

Por ejemplo:

Característica	Descripción	Criterio de evaluación
Complejidad	Nombre: dato obligatorio Primer apellido: dato obligatorio Segundo apellido: dato opcional Número de cuenta: dato obligatorio	Se evalúa cada uno de los datos obligatorios. Si el dato obligatorio tiene un valor obtiene un punto, de lo contrario su calificación es cero.
Confidencialidad	Número de cuenta, nombre, primer y segundo apellido se les debe dar tratamiento de datos personales.	Si es tratado como dato personal en el sistema tiene un punto, de lo contrario la calificación es cero.
Exactitud - Semántica	El Número de cuenta: compuesto por nueve dígitos, en el caso de generaciones 1999 o anteriores el número de cuenta está compuesto por 8 dígitos.	Si el número de cuenta es de 9 dígitos o 8 dígitos si es generación 1999 o anterior tiene 1 punto, de lo contrario es cero.
Exactitud - Sintáctica	El número de cuenta está compuesto por valores numéricos sin guión como separador del dígito verificador.	Si el número de cuenta solo tiene valores numéricos obtiene un punto, de lo contrario tiene una calificación de cero.
Precisión	Entidad federativa de nacimiento de acuerdo con el catálogo de INEGI.	Si la clave de la entidad federativa corresponde al catálogo de INEGI obtiene un punto, de lo contrario su valoración es de cero.

En la definición de las métricas se pueden agrupar los criterios de evaluación, por ejemplo:

<p>Métrica: Complejidad.</p> <p>Propósito: Verificar que los requisitos de los usuarios se ven reflejados en los datos, identificando que no existan faltantes ni existan datos adicionales.</p> <p>Método de aplicación: Obtener el puntaje obtenido de los criterios de evaluación de complejidad y compararlo con el puntaje total que podrían obtener.</p> <p>Fórmula: puntaje obtenido por el registro/ puntaje total esperado.</p>
--

Información estadística universitaria

La calidad de datos tiene una relación cercana con la información estadística generada en la Universidad, debido a que es un aspecto fundamental para su producción al contribuir en el cumplimiento de las características de ser confiable, comparable y oportuna para la planeación, el seguimiento, la toma de decisiones, la transparencia y la rendición de cuentas. Los datos que se obtienen de las fuentes primarias son procesados y organizados con la finalidad de darles valor estadístico.

La información recabada se valida verificando el comportamiento histórico de las variables de interés, así como con la propia fuente primaria en el caso de haber recibido un archivo básico que requiere un proceso de agregación y presentación adicional.

Cualquier variación en la aplicación de los criterios estadísticos y/o fechas de corte puede volver a la estadística recabada y procesada no comparable con ejercicios anteriores. Todos los productos resultantes tienen origen en la base de datos que se construye con el acopio sistemático de información estadística.

Aun cuando los atributos homogéneos y consistentes de la información favorecen su comparabilidad en el tiempo, un mismo concepto con diferentes criterios estadísticos y fechas de corte que responden a distintos fines son igualmente válidos.

Es por ello que un dato estadístico puede ser diferente cuando se considera con atributos distintos. Por ejemplo: los períodos fiscales son distintos a los ciclos escolares, por lo cual los valores de algunos indicadores estadísticos pueden ser diferentes de acuerdo con el criterio con el cuál se están reportando. Del mismo modo, un indicador puede variar dependiendo de la fecha de corte, por ejemplo, el número de alumnos inscritos es probable que sea distinto durante el periodo inicial de inscripciones al obtenido después del proceso de altas, bajas y cambios o justo antes de la emisión de las actas de examen.

Monitoreo de la calidad de los datos

Para asegurar una correcta calidad de los datos es necesario realizar una gestión continua de la misma, definiendo niveles de aceptación para cada dato y puntos de control.

Llevar a cabo revisiones a los datos mediante el análisis de su situación contribuye a que se pueden identificar problemáticas en la calidad de los datos, por ejemplo: ausencia de valores, valores erróneos o imprecisos, errores ortográficos, violación a las restricciones de unicidad, integridad referencial o verificación, violación de los dominios de datos, errores de dependencias funcionales, duplicidad innecesaria de la información, datos duplicados incorrectos, errores en la sintaxis, inconsistencias en las unidades de medida, heterogeneidad en la captura por ejemplo

entre mayúsculas y minúsculas, inexistencia de catálogos o catálogos no definidos apropiadamente, entre otros.

Se sugiere que las revisiones y análisis las realice el dueño de la información acompañado del administrador de la base de datos.

Así mismo, a nivel de una base de datos se pueden desarrollar diferentes tipos de análisis de la calidad de los datos, por ejemplo: a nivel de columna, de tabla, de tablas cruzadas para evaluar la integridad referencial, así como también reglas de datos. Lo que permitirá identificar cuáles son las causas raíz de los problemas de calidad encontrados.

El monitoreo se puede realizar de modo que puedan ser identificadas las anomalías y se establezca un proceso de corrección sobre las mismas, evitando de esta forma que los procesos de toma de decisiones estén soportados sobre información incorrecta.

Por lo regular la tarea de medir la calidad de los datos en el entorno de bases de datos implica utilizar herramientas y métodos especializados como la valoración del nivel de madurez en la calidad de los datos.

Diseño de bases de datos relacionales

El diseño de las bases de datos relacionales es un elemento que puede ayudar a fortalecer la calidad de los datos. Este debe representar las necesidades de los usuarios que hayan sido identificadas cubriéndose en su totalidad para poder cumplir con la característica de completitud.

Cuando se diseña el modelo de datos por primera vez o se revisa un modelo existente se debe comprender la naturaleza de los datos con los que se va a trabajar y se sugiere verificar que se cumpla con los puntos siguientes:

- ◆ El diagrama debe estar completo y representar todos los requerimientos que se quieren modelar. Se sugiere detallar el diagrama para obtener el modelo de datos físico¹.
- ◆ El diagrama debe estar construido de acuerdo con el lenguaje utilizado, con estructuras adecuadas a las necesidades que se están modelando.
- ◆ El diagrama debe contener sólo las entidades necesarias y reducir al mínimo la repetición de datos, ser fácil de entender, representar la realidad que intenta modelar.

¹ Un modelo de datos físico es un modelo específico de bases de datos que representa objetos de datos relacionales (por ejemplo, tablas, columnas, claves principales y claves externas) y sus relaciones. Un modelo de datos físico se puede utilizar para generar sentencias DDL que, después se pueden desplegar en un servidor de base de datos (IBM, 2021).

- ◆ El diagrama debe ser legible, es decir, no se enciman los objetos ni se cruzan las líneas complicando su entendimiento.
- ◆ Se debe establecer una convención de nomenclatura para el nombrado de las entidades y campos, vistas y procedimientos que represente la información que tienen o manejan, para lo cual se puede considerar:
 - Es mejor tener un nombre descriptivo que un acrónimo que nadie recuerde o entienda.
 - Evitar tener campos que tengan el mismo nombre pero con diferentes tipos o significados.
 - Evitar ser redundante en el nombre de los campos, por ejemplo, “Nombre del artículo” y “Precio del artículo”, será suficiente con “Nombre” y “Precio”.
 - Evitar las palabras reservadas para el Motor de Base de Datos.
 - Es aconsejable utilizar la regla de la clave primaria simple (autogenerado único entero), asignándole el nombre o prefijo “id” en cada tabla (Martínez, 2017).
- ◆ Cada base de datos debe normalizarse a la tercera forma normal por lo menos².
- ◆ Todas las entidades deben tener una llave primaria.
- ◆ Los nombres de las entidades y atributos deben utilizar un estándar de nombrado definido por el área responsable de la información o en caso de no existir, por el equipo de trabajo y validado con dicha área.
- ◆ Los nombres de las entidades deben ser únicos.
- ◆ Los nombres de los atributos deben ser únicos dentro de las entidades.
- ◆ Utilizar restricciones de llave foránea cuando sean necesarias.
- ◆ Verificar que los tipos de datos sean adecuados a los valores que va a contener el atributo. Como ejemplo de una práctica incorrecta, si un atributo que va a almacenar fechas se diseña con un tipo de dato cadena puede ocasionar que se almacenen datos incorrectos o con formatos heterogéneos.
- ◆ Evitar en la medida de lo posible la captura libre de información que pueda proceder de catálogos. Por ejemplo, códigos postales se pueden obtener de un catálogo y de ahí derivar la colonia, el municipio y el estado. Otro caso es el caso del nombre de las áreas

² La normalización es el proceso que asigna atributos a entidades en forma tal que las redundancias de datos se reducen o eliminan. Se realiza con la finalidad de evitar redundancias, inconsistencias, y anomalías lógicas, así como proteger la integridad de los datos.

universitarias donde es preferible seleccionar los valores y usar los catálogos con nombres y claves que ayuden a la interoperabilidad entre sistemas.

- ◆ Revisar que los tipos de datos usados sean homogéneos, por ejemplo, para el primer apellido de alumnos, trabajadores u otro tipo de personas se podría usar un tipo de datos varchar(50) en sus respectivas entidades, en lugar de tener diferentes especificaciones como varchar(50) para alumnos y varchar(40) para trabajadores.
- ◆ La redundancia de los datos en tablas y campos debe utilizarse sólo en casos muy específicos y debe estar claramente documentada para tenerse en cuenta en futuros desarrollos.
- ◆ Considerar la creación de índices en todas las columnas diferentes a la clave principal que se utilizan de manera recurrente para consultar la tabla, ordenando las columnas de las más a las menos utilizadas. Asimismo, se recomienda usar el mismo orden al crear el índice y al consultar en la base de datos, de lo contrario pudiera resultar contraproducente al tiempo de respuesta.

Debido a que no todas las bases de datos existentes en la Universidad son relacionales, sin importar la tecnología empleada debe considerar al menos:

- ◆ Utilizar una herramienta de modelado y documentación correspondiente a la tecnología de almacenamiento y explotación de datos que se utilice, por ejemplo: base de grafos, bases de datos orientadas a objetos, documentos, textos, por mencionar algunos.
- ◆ El diseño se puede complementar con otros tipos de diagrama como el modelado de Unified Modeling Language (UML), sobre todo cuando no se está usando una base de datos relacional³ como MongoDB⁴.
- ◆ El diseño debe ser flexible, es decir, tiene “la capacidad de poder tolerar cambios en los requisitos y adaptarse a nuevas necesidades de los usuarios” (González, 2013).
- ◆ El diseño está desarrollado conforme a las reglas, estándares y convenciones establecidas por el área universitaria.
- ◆ El diseño está documentado y los documentos disponibles son entendibles y suficientes para comprender el modelo.
- ◆ Al modelar datos es importante recordar que la información contenida en los datos es la que le da sentido a las aplicaciones y no al contrario. Así mismo, la documentación tiene que ver con comunicar el diseño y hacerlo comprensible en el futuro considerando que lo puedan comprender otras personas (por ejemplo, dada una rotación de personal).

³ <https://eaminds.com/2018/08/03/modelando-nosql-data-bases/>

⁴ <https://www.mongodb.com/docs/manual/core/data-modeling-introduction/>

Sistemas

En los sistemas se define la calidad del dato como el grado en que los datos reflejan la verdadera información basada en exactitud, consistencia, oportunidad e integridad desde su captura hasta la generación de informes.

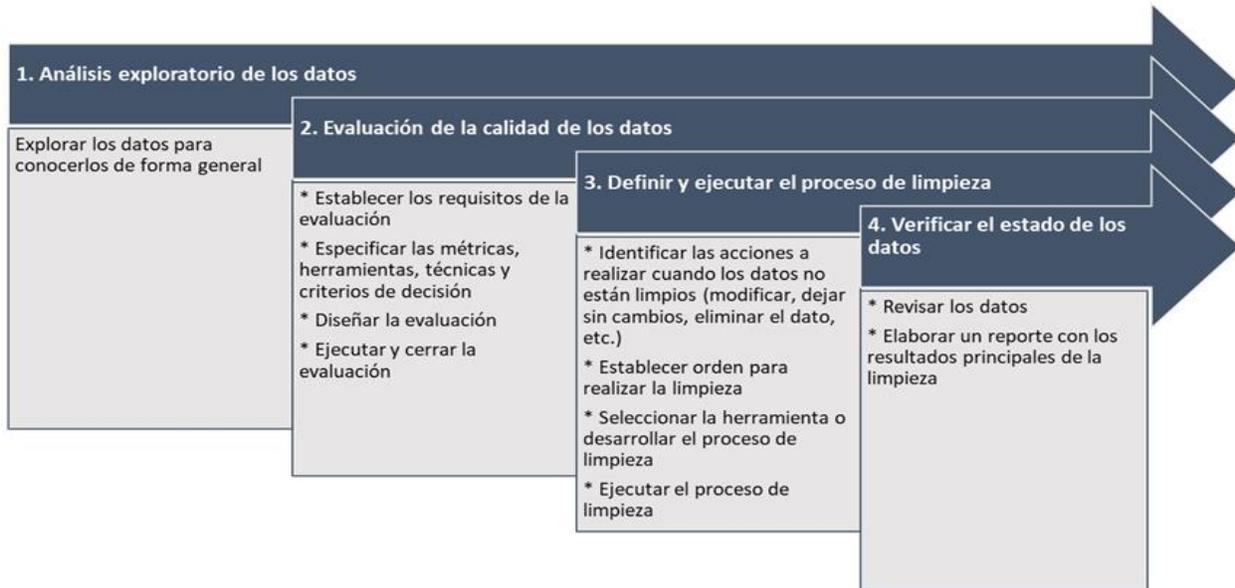
Los datos con baja calidad afectan el desempeño de los sistemas de información en las organizaciones. De igual forma si los datos recibidos por el sistema se almacenan en una base de datos sin validarlos implica en un futuro invertir tiempo y recursos en el proceso de detección y corrección de los mismos.

Es por ello que el diseño y validaciones de los sistemas tienen un alto impacto en la calidad de los datos, debido a que contribuyen mediante la implementación de controles a mantener la consistencia de los datos, cumplir las reglas de negocio, verificar que los datos obligatorios sean ingresados, emplear prácticas de programación segura, entre otros aspectos. Es por esto que en la definición de los requerimientos funcionales y no funcionales se sugiere incluir y documentar aquellos que contribuyen a mejorar la calidad de los datos.

El fortalecimiento de los sistemas junto con la capacitación de los usuarios que capturan información ayuda a mitigar los riesgos de tener información incorrecta e incompleta. También es muy importante considerar que antes de liberar un sistema de información se debe entregar la documentación de referencia del mismo, como puede ser un manual de usuario, manual de administración, manual técnico para el mantenimiento del sistema, por referir los más representativos.

Limpieza de los datos

Los procesos de limpieza se pueden hacer para: a) mejorar bases de datos que se encuentran en operación, b) cuando se integran datos de diferentes fuentes en una sola base de datos, c) en procesos de migración de bases de datos, d) cuando se obtienen datos de fuentes que no tienen un nivel de confiabilidad alto, entre otros. Ver la siguiente figura.



Proceso de limpieza de datos. Elaboración propia.

Con la finalidad de mejorar la calidad de los datos tratando de que aumente el nivel de cumplimiento de los criterios que se establezcan se pueden realizar procesos de limpieza de datos, para esto se pueden considerar los siguientes pasos:

- ◆ Análisis exploratorio de los datos para conocer el estado actual de la información.
- ◆ Evaluación de la calidad de los datos. De acuerdo con el estándar ISO/IEC 25040:2011 *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process*, se pueden realizar las actividades siguientes:
 1. Establecer los requisitos de la evaluación. Establecer el propósito de la evaluación, identificar los requisitos de calidad y las partes interesadas, los riesgos y el modelo de calidad a utilizar.
 2. Especificar la evaluación. Se especifican las métricas, herramientas, técnicas y criterios de decisión.
 3. Diseñar la evaluación. Se define el plan y las actividades de la evaluación.
 4. Ejecutar la evaluación. Se ejecuta la evaluación obteniendo las métricas de calidad y aplicando los criterios de decisión.
 5. Cierre de la evaluación. Se realiza un “informe de resultados finales y conclusiones con base a los valores obtenidos” (Calabrese, 2019).

- ◆ Definir y ejecutar el proceso que se usará para limpiar los datos de acuerdo con los hallazgos encontrados.
- ◆ Verificar el estado de los datos después de realizar la limpieza. En esta etapa se elabora un informe con los resultados principales obtenidos indicando aspectos como:
 - ¿Cuántos datos fueron modificados?
 - ¿Cuántos datos fueron eliminados?
 - ¿Cuántos datos necesitan limpieza manual?
 - Entre otros.

Para que la limpieza se lleve a cabo es necesario realizar una transformación de los datos que se determinen que deben ser actualizados de acuerdo con reglas establecidas definidas para el proceso de limpieza. Algunas modificaciones pueden resolverse de forma sencilla reemplazando valores, reemplazando subcadenas o haciendo cálculos sencillos para obtener un nuevo valor.

Otras modificaciones podrían ser más complicadas como modificar valores de acuerdo con catálogos establecidos que deben ser revisados y aprobados antes de realizar el proceso de limpieza. Otros pueden requerir de procesos más complejos e incluso es necesario que algunas modificaciones se realicen de forma manual por expertos en los datos que se están limpiando o mediante la comparación visual de documentos impresos.

Es importante mencionar que en los procesos de limpieza existen estrategias técnicas que se pueden considerar para detectar cual es el mejor valor a permanecer en caso de inconsistencias o registros duplicados. De manera alternativa, de no tener los elementos para implementar este tipo de estrategia, habría que documentar los elementos para determinar los datos que fueran candidatos a eliminarse, cuando la inconsistencia de los datos no permitieran saber cual es el valor correcto o se identifiquen valores incorrectos, por mencionar algunos casos.

Es importante considerar que existen diversas causas que pueden ensuciar los datos. Algunas de ellas son prevenibles, por lo que es recomendable hacer un análisis de riesgos con la finalidad de identificar acciones que puedan mitigar el impacto y fortalecer la calidad de los datos. Algunas de las causas que pueden ocasionar que los datos deban ser limpiados son:

- ◆ En los sistemas informáticos se presentan alguna de estas situaciones:
 - Validaciones insuficientes, es decir, el sistema no cuenta con validaciones de los datos que son ingresados o las validaciones son insuficientes.
 - El sistema cuenta con validaciones, pero no funcionan adecuadamente.
- ◆ Errores de captura de información como sintaxis no adecuada, errores ortográficos, entre otros.

- ◆ Modificación directa en la base de datos sin considerar las reglas de validación de los sistemas informáticos por personal con acceso y privilegios suficientes.
- ◆ Intrusión y modificación en las bases de datos por personal no autorizado.
- ◆ Escasa o inexistente definición de los dominios de datos.
- ◆ Falta o insuficiente definición de restricciones de integridad, llave primaria, unicidad y check en las bases de datos.
- ◆ Carga de datos incorrecta.
- ◆ Errores de diseño en la base de datos o en los sistemas.
- ◆ Datos calculados de forma errónea.
- ◆ Incumplimiento de reglas de negocio.
- ◆ Diferentes medios de ingreso de información y fuentes de datos inconsistentes.

Una línea de acción recomendada es aprender de los hallazgos identificados de manera proactiva, para detectar las oportunidades de mejora y atender de forma directa las causas raíz que están ocasionando los problemas en los datos anteriormente descritos, para que la próxima vez que se generen los datos estén más limpios y sean de mejor calidad para apoyar los fines de la institución.

Esto implica aplicar reingeniería, es decir aquellas reglas de negocio, omisiones o errores en la captura, producción, mantenimiento o archivado detectadas en el perfilado de datos y ejecutados en la limpieza, deberán ahora implantarse en el sistema original a fin de que se establezca una solución efectiva y evitar que el enfoque se centre en la limpieza de los datos.

Migración de datos

Las migraciones de datos consisten en realizar un proceso de extracción, transformación y carga (ETL por sus siglas en inglés) con la finalidad de desplazar datos de uno o varios orígenes hacia un nuevo destino. Por ejemplo: a) de un sistema de almacenamiento de datos a otro, b) de un formato de datos a otro o c) entre diferentes sistemas informáticos.

Los proyectos de migración de datos enfrentan varios retos, uno de ellos es la mala calidad de los datos, debido a que “es una de las razones por las que fracasan los proyectos de migración de datos, ya que estos proyectos implican el movimiento de grandes volúmenes de datos en formatos dispares” (Iqbal, 2019). Una inapropiada estrategia y/o planeación puede conllevar a

datos inexactos que contengan redundancias y valores desconocidos, a la pérdida de datos o, en su caso, a la superación de fechas o presupuestos.

“Según Gartner, la mala calidad de los datos puede tener un impacto financiero promedio de \$10,8 millones de dólares por año (2020). Además, los datos incorrectos significan que se está procesando información incorrecta, lo que podría implicar un nuevo trabajo” (Iqbal, 2019).

Entre los factores a considerar dentro de un proyecto de migración de datos están (Power Data, 2020):

- ◆ Tiempo que llevará realizar la migración completa.
- ◆ Cantidad de tiempo de inactividad que será necesaria.
- ◆ Riesgo derivado de problemas técnicos de compatibilidad, corrupción de datos, problemas de rendimiento de aplicaciones y pérdida u omisión de datos, entre otros.

Entre las estrategias de migración de los datos se encuentran:

- ◆ **Migraciones en *big bang*.** Implica realizar toda la migración dentro de una pequeña ventana de tiempo definida. Los sistemas conectados quedan inactivos mientras se someten los datos a procesos de ETL y se trasladan a la nueva base de datos.
 - Esta estrategia conlleva riesgos por lo que existe una intensa presión sobre la migración.
 - La verificación y aprobación de datos se encuentran en la ruta crítica.
- ◆ **Migraciones incrementales.** Implica tener ejecutándose los sistemas antiguos y nuevos en paralelo, lo que evita la inactividad y las interrupciones operativas que requieren las aplicaciones de misión crítica que operan 24/7.
 - La migración se realiza por fases y agrega cierta complejidad al diseño, ya que debe ser posible rastrear qué datos se han migrado y qué datos no, para redireccionar las peticiones cuando los sistemas operan en paralelo.

Para minimizar los riesgos relacionados a la migración de los datos, es necesario considerar lo siguiente:

- ◆ Entender qué datos se están migrando, de qué tipo son, cuál es su origen y qué formato adquirirán en destino, una vez completado el traslado.
- ◆ Establecer un metadatos o diccionario de datos que además del tipo de dato, origen, formato, entre otros elementos, explique qué es y la unidad de medida. Se debe tener bien clara la semántica del dato porque eso nos dirá qué hacer en la transformación del dato.

- ¿Qué pasa si se migran dos sistemas informáticos en uno solo? Se debe hacer correspondencia y mapeo de esquemas, así como la solución a problemas de heterogeneidad, sintáctica y semántica antes de transformarlo al modelo que contendrá lo de las dos fuentes de datos.
- ◆ Considerar una metodología/procedimiento de movimiento de datos que haya funcionado anteriormente o en otras áreas con necesidades semejantes.
- ◆ Definir lo que es viable y lo que no, en función de lo que las fuentes de datos admiten y lo que es razonable.
- ◆ Hacer una copia de seguridad de los datos antes de la ejecución de la migración.
- ◆ Deben definirse reglas de calidad de los datos para identificar y corregir cualquier error antes de que pueda llevarse a cabo la migración.
- ◆ Una vez identificada cualquier posible incidencia con sus datos de origen se debe proceder a la limpieza de los mismos.
- ◆ Aplicar los procesos ETL preferiblemente antes de proceder a la migración, como mínimo se deben considerar las etapas de transformación y de carga.
- ◆ Definir e implementar políticas de migración de datos para garantizar el orden necesario a lo largo de todo el proceso.
- ◆ Apostar por las pruebas y validación de los datos migrados, por ser una manera efectiva de asegurarse de que reúnen todos los atributos de calidad necesarios.
- ◆ El proceso de pruebas permite identificar mejoras en el proceso de migración y en los datos, además de contribuir a detallar la estimación del tiempo necesario para migrar, debido a que en algunas ocasiones podría ser más complejo de lo esperado inicialmente.

Se puede hacer uso de herramientas para apoyar la migración de los datos, ya sea que se construyan, lo que implica destinar recursos humanos y tiempo para desarrollarlas, o en su caso utilizar las desarrolladas por terceros o aquellas comerciales. Un ejemplo de su uso son las herramientas ETL, que son muy adecuadas para la tarea de migrar datos de una base de datos a otra, principalmente en proyectos en que existen pocas conexiones entre origen y destino.

Calidad de datos abiertos

Los datos que las áreas universitarias identifiquen como abiertos pueden ser reutilizados y puestos a disposición de la sociedad de acuerdo con las políticas de compartición que establezcan. Para mejorar la calidad de los datos abiertos se pueden seguir las pautas siguientes:

- ◆ Usar otras alternativas de búsqueda adicionales al uso de buscadores excesivamente guiados o acotados.
- ◆ Verificar que los datos no estén incompletos, demasiado limitados por el periodo de tiempo, por la frecuencia de la publicación o con información muy básica que “aporta poco valor y no cumple con los mínimos necesarios para obtener una representación completa” (Gobierno de España, 2017), es decir, que aporten el nivel de detalle adecuado.
- ◆ Verificar que los datos sean de fácil acceso.
- ◆ Corroborar que los datos estén actualizados y no sean obsoletos.
- ◆ Confirmar que los formatos usados sean reutilizables y estandarizados.
- ◆ Establecer claramente las licencias y condiciones de reutilización de los datos.
- ◆ Revisar la conformidad con la normatividad vigente.
- ◆ Evaluar la calidad de los datos de forma periódica.
- ◆ Generar metadatos que faciliten reutilizar los datos abiertos.

Indicadores de actividad y desempeño

Una tarea relevante de la gestión académico administrativa es dar seguimiento y evaluar de manera clara y objetiva los procesos y resultados de las acciones institucionales emprendidas. Al contar con información estadística confiable se tiene la posibilidad de construir indicadores de actividad o desempeño para describir y valorar a la institución o a una entidad académica en particular.

Para el titular de una facultad o escuela y su equipo de trabajo es importante conocer tanto el curso de las acciones emprendidas como los resultados obtenidos, así como tomar decisiones respecto a la continuidad o no de los proyectos, su fortalecimiento o modificación, si así fuera el caso.

En este sentido el indicador es crucial porque es uno de los elementos que nos dará la pauta del grado de cumplimiento. Desde décadas atrás, es frecuente encontrar en diversos estudios y reportes institucionales **indicadores de actividad** (descriptivos) “tales como la magnitud de la demanda, de la población escolar, el egreso, la titulación, el número de proyectos de investigación, el número y tipo de productos de investigación, la cantidad de asistentes a las actividades de difusión y extensión”, por mencionar algunos (DGPL, 2021:I).

También se han utilizado extensamente **indicadores de desempeño** (valorativos) tales como eficiencia terminal (porcentaje de alumnos egresados de una generación determinada), costo promedio por alumno, regularidad (porcentaje de alumnos regulares de una generación o de la totalidad de alumnos inscritos), relaciones demanda/cupo, índices de aprobación/reprobación,

libros o artículos publicados por investigador, variaciones de asistencia a actividades de difusión o extensión, entre otros (DGPL, 2021:1).

“La información obtenida a partir de indicadores de desempeño permite, asimismo, que las entidades (o la institución en su conjunto) puedan compararse entre sí o con otras instituciones; impulsar procesos de innovación educativa o académica; distribuir los recursos conforme a prioridades y rendir cuentas tanto a las propias comunidades como a instituciones públicas, organismos internacionales y, en general, a la sociedad en su conjunto” (DGPL, 2021:1).

La información confiable y comparable

“En el ámbito de las instituciones de educación superior existe un consenso en cuanto a los conceptos de relevancia estadística más utilizados para conocer y valorar su desempeño. Sin embargo, para que esta estadística adquiera la categoría de *confiable* y *comparable* se debe cuidar la especificación clara de lo siguiente” (CGPL, 2022:16) (Ver ejemplo en el Anexo 3):

- ◆ Los *conceptos* principales de interés estadístico.
- ◆ El establecimiento de las *fechas de corte* adecuadas para la generación y difusión de la información estadística institucional.
- ◆ La documentación de los *criterios estadísticos* para el tratamiento de los datos.
- ◆ La descripción de los contenidos, formatos o las dimensiones asociadas a los conceptos de interés (*niveles de agregación* o subgrupos y categorías) (CGPL, 2022:16).

El flujo sistemático de la información

“Las solicitudes formales de información a las *fuentes primarias* contienen *guías técnicas* en donde se describen los acuerdos sobre los conceptos y criterios que conforman la estadística institucional. Incluyen los protocolos de transferencia y demás aspectos técnicos que deben documentarse para asegurar su flujo sistemático” (CGPL, 2022:16). Las más de 150 entidades académicas y dependencias administrativas universitarias son potencialmente fuentes primarias de información.

“Conviene diferenciar dos tipos de fuentes de información: las **fuentes primarias** que son el origen de las estadísticas relevantes y las **fuentes responsables** que les otorgan valor estadístico a las variables mediante algún proceso de homologación o normalización. En ocasiones una fuente primaria es también una fuente responsable, sin embargo, puede no ser así, sobre todo si las fuentes primarias tienen objetivos de gestión distintos al interés estadístico institucional. En la administración central universitaria se tienen relevantes sistemas de gestión que se convierten en fuentes primarias de gran valor para resolver el flujo de información” (CGPL, 2022:17).

Métricas de calidad de los indicadores

Las métricas de calidad de los indicadores se describen en una *ficha* o *documentación técnica* que contiene lo siguiente (Ver ejemplo en el Anexo 3):

- ◆ Nombre del indicador
- ◆ Definición conceptual
- ◆ Unidad de medida
- ◆ Fórmula de cálculo (para indicadores de desempeño, como son la regularidad o la eficiencia terminal)
- ◆ Interpretación (indicadores de desempeño)
- ◆ Utilidad/fin/propósito
- ◆ Niveles de desagregación o subconjuntos posibles
- ◆ Categorías posibles/útiles
- ◆ Fuentes de información (primaria y responsable)
- ◆ Criterio estadístico o método de cálculo
- ◆ Periodicidad de medición
- ◆ Fecha de corte sugerida
- ◆ Observaciones

Se generarían tantas fichas técnicas para un mismo *concepto* como modalidades de cálculo y finalidades del indicador existan.

Anexo 1. Descripción de los atributos de calidad de los datos

El estándar ISO 25012:2008 *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model* presenta quince características de calidad que pueden ser abordadas desde los puntos de vista dependiente e inherente del sistema. Las características dependientes son aquellas cuya calidad depende del sistema que hace uso de los datos.

Características dependientes según ISO /IEC 25012:2008

Dimensión	Descripción
Disponibilidad (Availability)	El grado en el cual el dato tiene atributos que le permiten ser recuperados por usuarios autorizados o por aplicaciones en un contexto específico de uso.
Portabilidad (Portability)	El grado en el cual el dato tiene los atributos que le permiten ser instalado, sustituido o movido de un sistema a otro conservando la calidad existente en un contexto específico de uso.
Recuperabilidad (Recoverability)	El grado en el cual el dato puede mantener y conservar un nivel especificado de operaciones y calidad, aún en caso de falla.

(Pinzón, 2013:115)

Las características inherentes se refieren al “grado con el que las características de calidad de los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas cuando los datos son utilizados bajo condiciones específicas” (ISO 25000, 2008).

Características inherentes según ISO /IEC 25012:2008

Dimensión	Descripción
Exactitud (Accuracy)	El grado en el cual el dato tiene atributos que representan correctamente el valor del atributo intencionado de un concepto o evento en un contexto específico de empleo.
Complejitud (Completeness)	El grado al cual el dato del sujeto asociado con una entidad tiene valores para todos los atributos esperados e instancias de entidad relacionadas en un contexto específico de uso.
Consistencia (Consistency)	El grado en el cual el dato tiene los atributos que son libres de contradicción y son coherentes con otros datos en un contexto específico de uso.
Credibilidad (Credibility)	El grado en el cual el dato tiene atributos que son considerados verdaderos y creíbles por usuarios en un contexto específico de uso.
Actualidad (Currentness)	El grado en el cual el dato tiene los atributos que son del período correcto en un contexto específico de uso.

(Pinzón, 2013:116)

Existen características que son comunes tanto al enfoque inherente como al dependiente, las cuales se presentan en la siguiente tabla.

Características compartidas (inherentes y dependientes) según ISO /IEC 25012:2008

Dimensión	Descripción
Accesibilidad (Accessibility)	El grado en el cual el dato puede ser accedido en un contexto específico de uso, en particular, por personas en situación de discapacidad que necesitan el soporte de tecnología o una configuración especial.
Cumplimiento / Conformidad (Compliance)	El grado en el cual el dato tiene atributos que se adhieren a normas, convenciones o regulaciones vigentes y reglas similares relacionadas con la calidad de datos en un contexto específico de uso.
Confidencialidad (Confidentiality)	El grado en el cual el dato tiene los atributos que aseguran que sólo es accesible e interpretable por usuarios autorizados en un contexto específico de uso.
Eficiencia (Efficiency)	El grado en el cual el dato tiene los atributos que pueden ser procesados, y proporciona los niveles esperados de funcionamiento (desempeño) usando las cantidades y los tipos de recursos apropiados en un contexto específico de uso.
Precisión (Precision)	El grado en el cual el dato tiene atributos que son exactos o que proporcionan la discriminación en un contexto específico de uso.
Trazabilidad (Traceability)	El grado en el cual el dato tiene atributos que proporcionan un rastro de auditoría de acceso a los datos y de cualquier cambio hecho a los datos en un contexto específico de uso.
Entendimiento (Understandability)	El grado en el cual el dato tiene atributos que le permiten ser leído e interpretado por usuarios, y es expresado en lenguajes apropiados, símbolos y unidades en un contexto específico de uso.

(González, 2013:116)

Anexo 2. Métricas de acuerdo con el estándar ISO/IEC 25024:2015

Característica	Métrica
Exactitud	Exactitud sintáctica de los datos
	Exactitud semántica de los datos
	Aseguramiento de la exactitud de los datos
	Riesgo de inexactitud de un conjunto de datos

	Exactitud del modelo de datos
	Exactitud de los metadatos
	Rango de la exactitud de los datos
Compleitud	Compleitud de registros
	Compleitud de atributos
	Compleitud de archivos de datos
	Compleitud de valores de datos
	Registros vacíos en un archivo de datos
	Compleitud del modelo de datos conceptual
	Compleitud de los atributos del modelo de datos conceptual
	Compleitud de los metadatos
Consistencia	Integridad referencial
	Consistencia del formato de datos
	Riesgo de inconsistencia de datos
	Consistencia de la arquitectura
	Cobertura de la consistencia de valores de datos
	Consistencia semántica
Credibilidad	Credibilidad de los valores
	Credibilidad de la fuente
	Credibilidad del diccionario de datos
	Credibilidad del modelo de datos
Actualidad	Frecuencia de actualización
	Oportunidad de la actualización

	Requisito de actualización de los elementos (ítems)
Accesibilidad	Accesibilidad de usuario
	Accesibilidad de dispositivos
	Accesibilidad del formato de datos
Conformidad	Conformidad regulatoria de los valores y/o formato de datos
	Conformidad regulatoria debido a la tecnología
Confidencialidad	Uso de cifrado
	No vulnerabilidad
Eficiencia	Formato de los elementos (ítems) de datos eficiente
	Eficiencia usable
	Eficiencia en formato de datos
	Eficiencia en procesamiento de datos
	Riesgo de espacio desperdiciado
	Espacio ocupado por duplicación de registros
	Desfase temporal de actualización de datos
Precisión	Precisión de valores de datos
	Precisión del formato de datos
Trazabilidad	Trazabilidad de valores de datos inherentes
	Trazabilidad del acceso a los datos
	Trazabilidad de valores de datos dependientes del sistema
Entendibilidad	Entendibilidad de símbolos
	Entendibilidad semántica
	Entendibilidad de datos maestros

	Entendibilidad de valores de datos
	Entendibilidad del modelo de datos
	Entendibilidad de la representación de datos
	Entendibilidad de datos maestros enlazados
Disponibilidad	Tasa de disponibilidad de datos
	Probabilidad de datos disponibles
	Disponibilidad de elementos de arquitectura
Portabilidad	Ratio de portabilidad de datos
	Portabilidad de datos prospectivos
	Portabilidad de elementos de arquitectura
Recuperabilidad	Tasa de recuperabilidad de datos
	Respaldo de seguridad (Backup) periódico
	Recuperabilidad de la arquitectura

Métricas de la ISO/IEC 25024 (Adaptado de Fernández, 2018)

Anexo 3. Ejemplos de *fichas técnicas* de indicadores de actividad y desempeño

Indicador de actividad

UNAM. Docencia. Educación de bachillerato

Indicador 1a	Documentación técnica
Población escolar	
Ámbito: Alumnos	
Aplicación: Estadística Educativa Nacional SEP-ANUIES-INEGI y Agenda Estadística UNAM	

Definición conceptual

Total de **alumnos inscritos** en una facultad o escuela, carrera o programa de estudios de la UNAM, en un **ciclo escolar*** determinado.

Alumno inscrito

Persona registrada en una o más asignaturas establecidas en el plan de estudios, en un año o ciclo escolar determinado.

Ciclo escolar

Periodo o lapso de duración anual o semestral que señala el Consejo Universitario para realizar actividades académicas.

Unidad de medida

Alumno inscrito

Periodicidad de medición

Anual

Utilidad

Primordialmente, este indicador es una medida de contribución a la cobertura de los servicios de educación media superior (EMS) vinculados al Presupuesto Federal del Programa Presupuestario E007. Adicionalmente, permite conocer y valorar el esfuerzo universitario para atender a un número de personas en su calidad de alumnos en educación media superior en la UNAM.

Dimensión del desempeño

Cobertura

Naturaleza sistémica del indicador

Insumo

Niveles de agregación

- UNAM
- Plantel
- Turno

Categorías sugeridas ó subconjuntos

Primer ingreso, reingreso, sexo.

Criterio estadístico

Se contabilizan los alumnos inscritos en al menos una asignatura en examen ordinario en el primer semestre –plan semestral, Colegio de Ciencias y Humanidades– o año escolar –plan anual, Escuela Nacional Preparatoria–.

Fecha de corte (sugerida)

Inicio del ciclo escolar, una vez registrada la inscripción definitiva.

Fuentes de información

Primaria**

Dirección General de Administración Escolar

Responsable***

Coordinación General de Planeación y Simplificación de la Gestión Institucional

Observaciones

La comparabilidad en el tiempo de este indicador se asegura respetando el criterio estadístico descrito líneas arriba, acordado en 1986, año en que se inicia la publicación de la Agenda Estadística UNAM.

* También denominado **año escolar** o **año lectivo**. Ejemplo: ciclo escolar 2022-2023.

** Entidad de origen del registro de alumnos inscritos.

*** Entidad responsable de otorgarle valor estadístico al indicador.

Indicador de desempeño
Docencia. Educación de licenciatura

Indicador 3a	Documentación técnica
Regularidad (avance escolar)	
Ámbito: Alumnos	
Aplicación: Sistema de Integrado de Información Académica (SIIA), UNAM	

Definición conceptual

Proporción de alumnos regulares de una generación de ingreso con relación al total de alumnos de dicha generación.

Alumno regular

Alumno con un avance en créditos igual o superior al establecido en el plan de estudios en el que está inscrito de acuerdo con su generación de ingreso y al número de semestres o años cursados.

Fórmula de cálculo

$$\left(\frac{\text{alumnos regulares}}{\text{total de alumnos de la generación}} \right) * 100$$

Periodicidad de medición

Anual

Utilidad

Este indicador se utiliza en los diagnósticos, evaluaciones y procesos de planeación de los sistemas educativos. Permite conocer la eficiencia del sistema en el marco del avance de los alumnos en el tiempo establecido por el plan de estudios para conservar la calidad de alumnos regulares. Primordialmente, se utiliza como una medida del rendimiento de la entidad académica en términos de la relación insumo-producto.

Dimensión del desempeño

Eficiencia

Naturaleza sistémica del indicador

Proceso

Niveles de agregación

- Por carrera-generación del sistema escolarizado
- Por carrera-generación del SUAYED

Categorías sugeridas ó subconjuntos

Vía de ingreso (concurso de selección, pase reglamentado-plantel de origen), sexo.

Criterio estadístico

Se contabilizan los alumnos con un avance en créditos igual o superior al establecido en el plan de estudios en el que está inscrito de acuerdo con su generación de ingreso y al número de semestres o años cursados. Los alumnos con avance cero se excluyen de la generación.

Fecha de corte (sugerida)

Cierre del ciclo escolar, verificando la actualización completa de historias académicas.

Fuentes de información

Primaria**

Dirección General de Administración Escolar

Responsable***

Dirección General de Evaluación Institucional

Observaciones

En el caso de que en un plan de estudios contenga asignaturas optativas en los que los créditos sean variables, se contabiliza el número de créditos que deben ser cursados con la que tenga el menor, dejando la compensación correspondiente a la diferencia para el conteo del último semestre o año. En términos operativos, se cuenta con una base histórica de todos los planes de estudio impartidos en la universidad desde 2002 en el que se ha identificado el número de créditos que deben ser cursados para cada semestre o año, contra la que se compara el avance real de cada estudiante de acuerdo con su generación de ingreso y el número de semestres o años cursados.

* Entidad de origen del registro de alumnos inscritos.

** Entidad responsable de otorgarle valor estadístico al indicador.

Referencias

Calabrese (2019). Guía para evaluar calidad de datos basada en ISO/IEC 25012. Revisado el 26 de octubre de 2021. Disponible en: <https://core.ac.uk/download/pdf/301104068.pdf>

Fernández (2018). Desarrollo de un modelo de calidad de datos aplicado a una solución de inteligencia de negocios en una institución educativa: caso lambda. Tesis para el Título de Ingeniero Informático. Pontificia Universidad Católica del Perú. Disponible en: https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/12014/FERNANDEZ_MAR SHALL_CALIDAD_INTELIGENCIA_NEGOCIOS.pdf?sequence=1

Gartner (s.f.). Gartner Glossary. Dark Data. Revisado el 26 de octubre de 2021. <https://www.gartner.com/en/information-technology/glossary/dark-data>

Gobierno de España (2017). Manual práctico para mejorar la calidad de los datos abiertos. Revisado el 25 de octubre de 2021. Disponible en: https://datos.gob.es/sites/default/files/doc/file/manual_practico_para_mejorar_la_calidad_de_los_datos_abiertos_1_0.pdf

González (2013). Aplicación del estándar ISO/IEC 9126-3 en el modelo de datos conceptual entidad-relación. Revista Facultad de Ingeniería, UPTC, julio - diciembre de 2013, Vol. 22, No. 35. Revisado el 7 de octubre de 2021. Disponible en: <http://www.scielo.org.co/pdf/rfing/v22n35/v22n35a10.pdf>

IBM (2021). Modelos físicos de datos. Revisado el 11 de noviembre de 2022. Disponible en: https://www.ibm.com/docs/es/radfws/9.6?topic=SSRTLW_9.6.0/com.ibm.datatools.core.ui.doc/topics/cphysmod.htm

Iqbal (2019). Gestión de la calidad de los datos: ¿qué es y por qué es importante? Revisado el 26 de octubre de 2021. <https://www.astera.com/es/type/blog/data-quality-management/>

ISO 25000 (2008). ISO/IEC 25012. Revisado el 7 de octubre de 2021. Disponible en: <https://iso25000.com/index.php/normas-iso-25000/iso-25012>

Martínez (2017). Malas Prácticas en el Diseño de la Base de Datos: ¿Estás cometiendo estos errores?. Revisado el 29 de octubre de 2021. Disponible en: <https://www.toptal.com/database/malas-practicas-en-el-diseno-de-la-base-de-datos-estas-cometiendo-estos-errores>

Jacintogr (2020). La importancia de la calidad de los datos en las empresas. Revisado el 29 de octubre de 2021. Disponible en:

<https://itblogsogeti.com/2020/11/25/la-importancia-de-la-calidad-de-los-datos-en-las-empresas/>

Yanes (2019). La evaluación de la calidad de datos: una aproximación criptográfica. Computación y Sistemas, Vol. 23, No. 2, 2019, pp. 557–568 doi: 10.13053/CyS-23-2-2899. Disponible en: <http://www.scielo.org.mx/pdf/cys/v23n2/1405-5546-cys-23-02-557.pdf>

Power Data (2020). Migración de datos: definición, desafíos y mejores prácticas para afrontarla. Revisado el 3 de noviembre de 2021. Disponible en:

<https://www.powerdata.es/migracion-de-datos>

Dirección General de Planeación (DGPL), (2021). Indicadores de desempeño para escuelas y facultades de educación superior de la UNAM. UNAM, UNAM. Disponible en:

https://www.planeacion.unam.mx/Planeacion/Apoyo/IndDesFyE_nov2021.pdf

Coordinación General de Planeación y Simplificación de la Gestión Institucional (CGPL), (2022). Un modelo de planeación y los sistemas de información para la toma de decisiones en una organización compleja. UNAM. Disponible en:

<https://www.planeacion.unam.mx/Planeacion/Apoyo/UNAM-modeloplan-cgpl.pdf>

Créditos

Elaboración

Susana Laura Corona Correa (DGTIC), Alberto González Guízar (DGTIC), Jaime Escamilla Rivera (CGPL)

Revisión

María del Pilar Angeles (IIMAS), Hugo Alonso Reyes Herrera (DGTIC), Leticia Martínez Calixto (DGTIC), Fernando Israel González Trejo (FES Acatlán), Armando Vega (DGAE), Leonard Pulido Cauzard (DGAE), Ana Pérez Arteaga (IIMAS), Fernando Zaragoza Hernández (DGAE), Rogelio Palma Rodríguez (DGPo), José Gerardo Moreno Salinas (CUAIEED), Alejandra Fonseca Salazar (COUS)