

# Herramientas y estándares para el aseguramiento de la información



## Proceso de Control de Divulgación Estadística en el INEC de Costa Rica

Catalina Ruiz Barquero

Área de Coordinación del Sistema de Estadística Nacional

INEC-Costa Rica

# Contenido

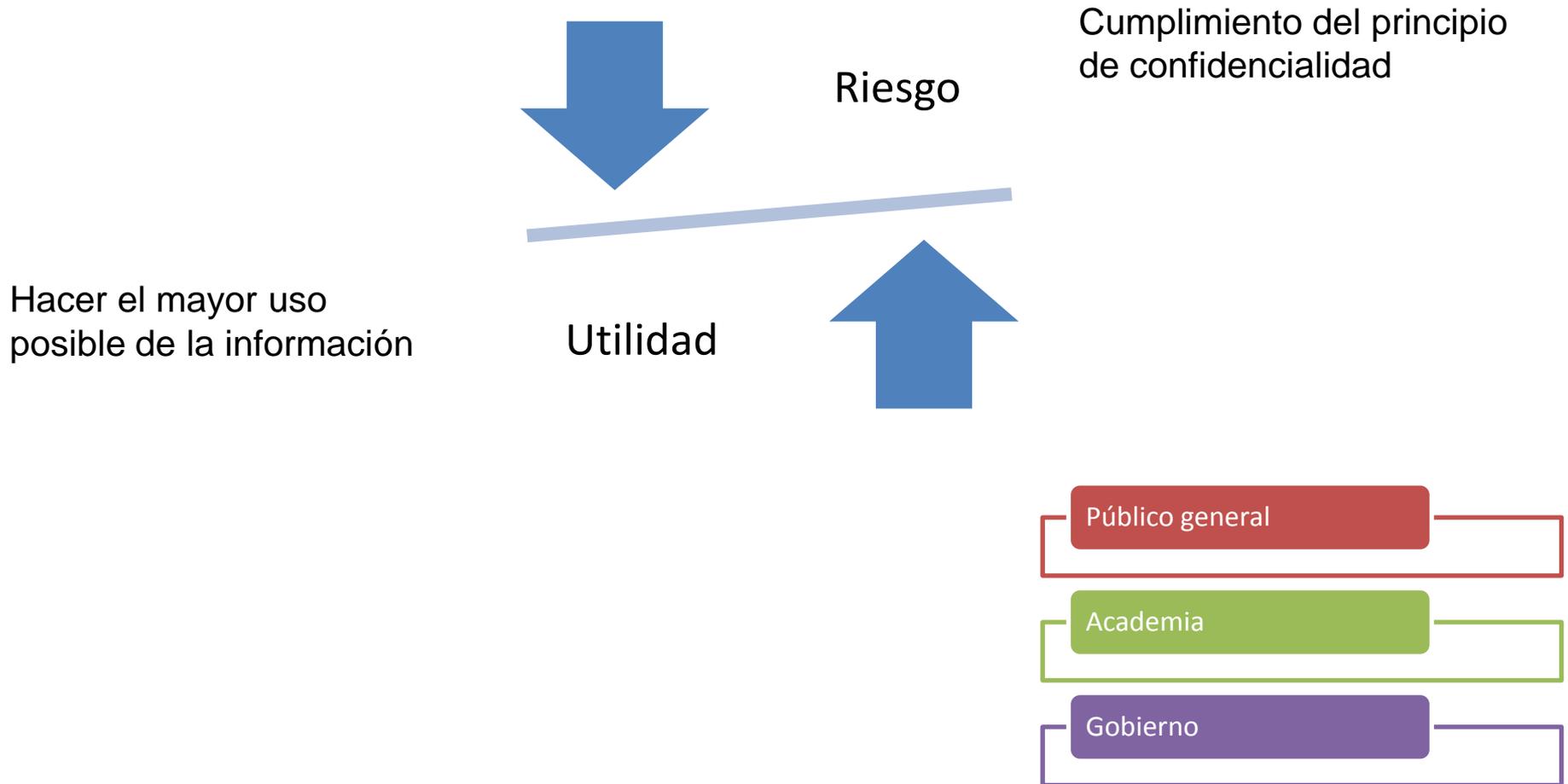
- El control de la divulgación estadística: entorno internacional y nacional
- La política de divulgación de estadísticas y entrega de microdatos en el INEC-CR
- La experiencia nacional en procesos de anonimización: Caso de encuestas, caso de registros

# Concepto

- Control de la divulgación estadística (SDC, por sus siglas en inglés), también conocido como anonimización:

Se refiere a las políticas que regulan la divulgación de los datos estadísticos, así como los métodos que se emplean para tratar los datos, con el fin de evitar el “desenmascaramiento” de la identidad de las unidades de estudio, sean estas personas, hogares, empresas, fincas, etc.

# Tipos de usuarios de la información y balance entre riesgo y utilidad



# Tendencia mundial

- Actualmente hay una tendencia a la publicación de microdatos estadísticos.
- Para esto, las agencias de estadística deben:
- Establecer políticas y procedimientos que definan las condiciones de acceso a los datos.
- Aplicar los métodos para el control de la divulgación estadística a los datos antes de entregarlos.
- Apoyo internacional para el desarrollo de procedimientos técnicos para la anonimización de datos y su aplicación, por parte de Organismos como Banco Mundial y OECD.

# Requerimientos internacionales

## OECD DRAFT RECOMMENDATION ON GOOD STATISTICAL PRACTICE

<p>4. Protect the privacy of data providers (including individuals, households, enterprises, administrations, and all level of government), and guarantee by law the confidentiality of the individual information provided and its use for statistical purposes only.</p>	<p>4.1. Statistical confidentiality is guaranteed by law (ECoP).</p> <p>4.2. Specific measures are in place to ensure the full protection of individual data from any potential disclosure without consent, with the aim to ensure the confidence of data providers in participating in statistical surveys: written instructions and internal guidelines are provided to statistical authority staff on the full protection of statistical confidentiality in the production and dissemination processes; appropriate penalties are prescribed for wilful breach of confidentiality and for any disclosure of individual data of a private nature that could infringe upon private life. These penalties are well-known to statistical staff and new employees sign legal confidentiality commitment upon appointment.</p> <p>4.3. The confidentiality policy is made known to the public (ECoP).</p> <p>4.4. Physical, technological, administrative and organisational provisions are in place to protect the security and integrity of statistical databases (ECoP).</p> <p>4.5. Provisions are in place and internal guidelines are available to allow external user access to micro-data for statistical research purposes under strict protocols and only after anonymisation of the data.</p>
--	---

## ¿Porqué se inicia con el proceso de *Control de Divulgación Estadística* en el INEC?

- Existen leyes y otros instrumentos normativos y técnicos que establecen principios básicos de acceso a la información y a la protección de la confidencialidad. En el caso de Costa Rica, entre otros, se destacan: :
  - Ley 7839 del Sistema de Estadística Nacional (1998)
  - Ley de Protección de la Persona frente al Tratamiento de sus Datos Personales (2010)
  - Política para la Divulgación de las Estadísticas y la Entrega de Microdatos del INEC (2014)
  - Código de Buenas Prácticas Estadísticas de Costa Rica (2014)
- Sin embargo, se ha identificado la ausencia de mecanismos y procedimientos claramente definidos para su aplicación y cumplimiento.



# Política de divulgación de resultados estadísticos y entrega de microdatos INEC-Costa Rica

- 35%
- 50%
- 75%

### Acceso

- Gratuidad por medio de la página WEB
- Cobro de productos y servicios especiales (se reglamenta) y se definen

### Oportunidad

- Se establece tiempo de entrega entre la finalización de la recolección o cierre y la divulgación, por tipo de operación estadística

### Calendario de divulgación

- Establece días en que se divulga y el mes en caso de Operación Estadística con periodicidad anual o mayor
- Debe informarse cualquier cambio con antelación
- Se debe publicar en la Gaceta

Entrega  
Adelantada

- Jerarca de Ministerio, BCCR según tema (24 horas antes )
- Emergencia Nacional o interés público
- Revisión expertos (contrato confidencialidad)
- Prensa si se considera oportuno (24 horas antes, contrato confidencialidad)

Metadatos

- Obligación de documentar (PAD) las Operaciones Estadísticas y divulgar (ANDA)
- Máximo un mes después de publicados resultados principales

Marcos  
Muestrales y  
Directorios

- No se entregan
- Se vende el servicio de muestras

# Tipos de archivos de microdatos

## Archivo de Uso Público

- Todo público
- Anonimizado
- Se descarga de la WEB
- Es gratuito

## Archivos con licencia o uso restringido

- Para usuarios de la academia, investigadores autorizados y de instituciones publicas autorizados
- Anonimizado
- Se debe solicitar el acceso (bajo contrato)
- Es gratuito

## Archivos no accesibles en sitio

- Uso restringido solo en oficina
- Solo a usuarios autorizados
- Se revisa las salidas de los datos para asegurar que no se obtenga información expuesta

## Archivos de acceso en línea

- Bases de datos para acceso público
- Mediante el sistema Redatam
- Gratuito



# Anonimización en Registros Administrativos

35%

50%

75%



# Registros Administrativos– Metodología UK

## Vitales

### -Usuarios públicos

Tabulados oficiales anonimizados

Redatam (anonimizado)

### -Usuarios científicos

Redatam (anonimizado a un menor nivel que el de usuarios públicos)

### -Gobierno

Bases originales

Documentación del proceso (usuarios externos)

Manuales para uso interno (técnicos que realizan el proceso y centro de información)

# Metodología para la anonimización de estadísticas vitales

1. Determinar los tipos de variables y categorizar según:
  1. Variables identificadoras
  2. Variables sensibles
  3. Variables de uso común (para publicación y para organismos internacionales)
  4. Variables de uso no-común
  
2. Definir las medidas de anonimización que se deben aplicar a cada variable según su nivel de riesgo

## Pasos para construir una tabla para publicación

Debe contestar las siguientes preguntas:

1. La tabla presenta la información de una manera comprensible para los usuarios?
2. La información que se presenta en la tabla es de calidad?
3. La información que se presenta en la tabla es útil para los usuarios?

## Selección del método de anonimización

- Diseño de las tablas
  - Modificación de celdas (en las tablas)
    - Modificaciones en las bases de datos para la posterior construcción de tablas



Proceso llevado a cabo hasta el momento...

- 35%
- 50%
- 75%

# Aplicación en el INEC

Base metodológica



Banco Mundial

Software



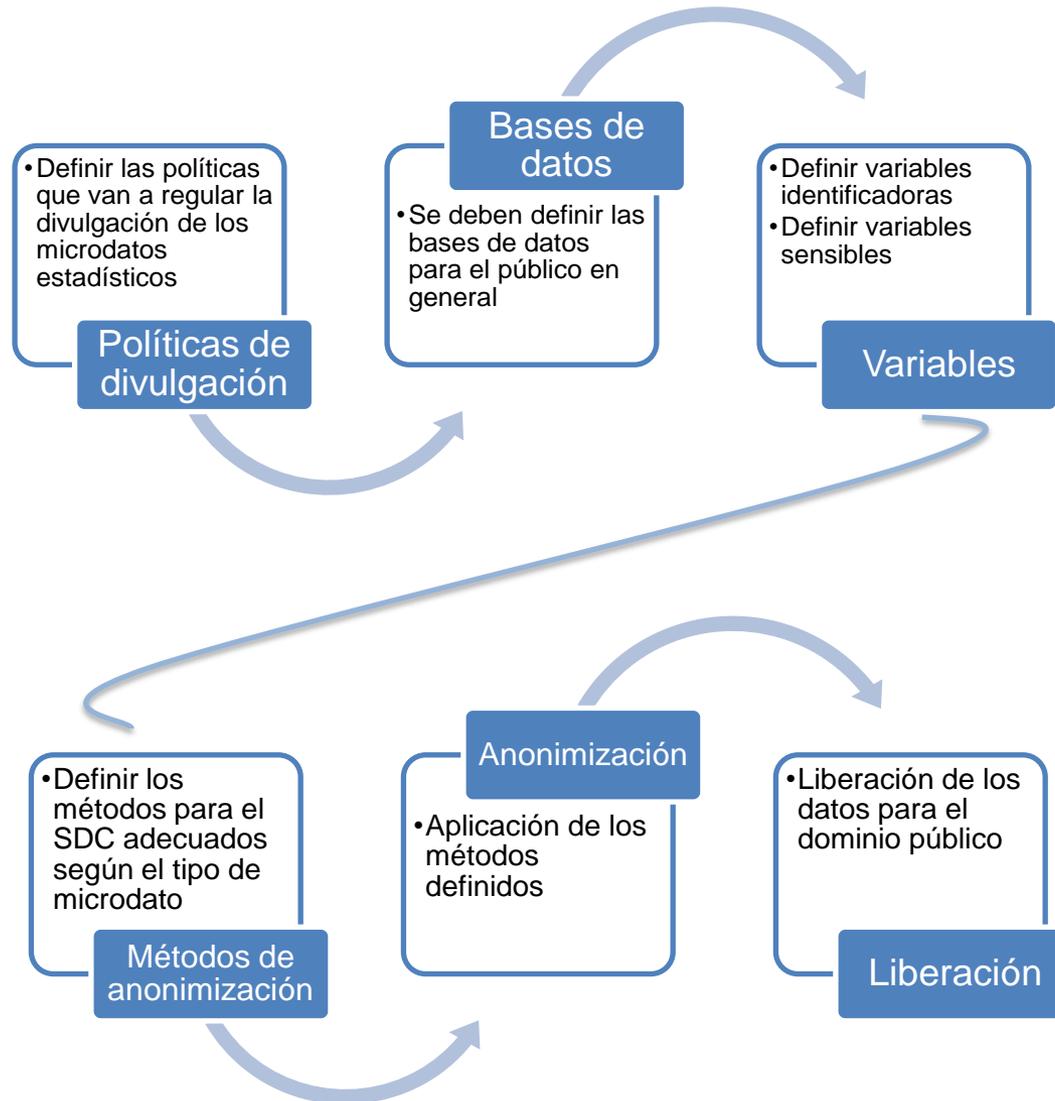
R (sdcmicro)

Metas



- Encuestas
- Registros Administrativos
- Censos

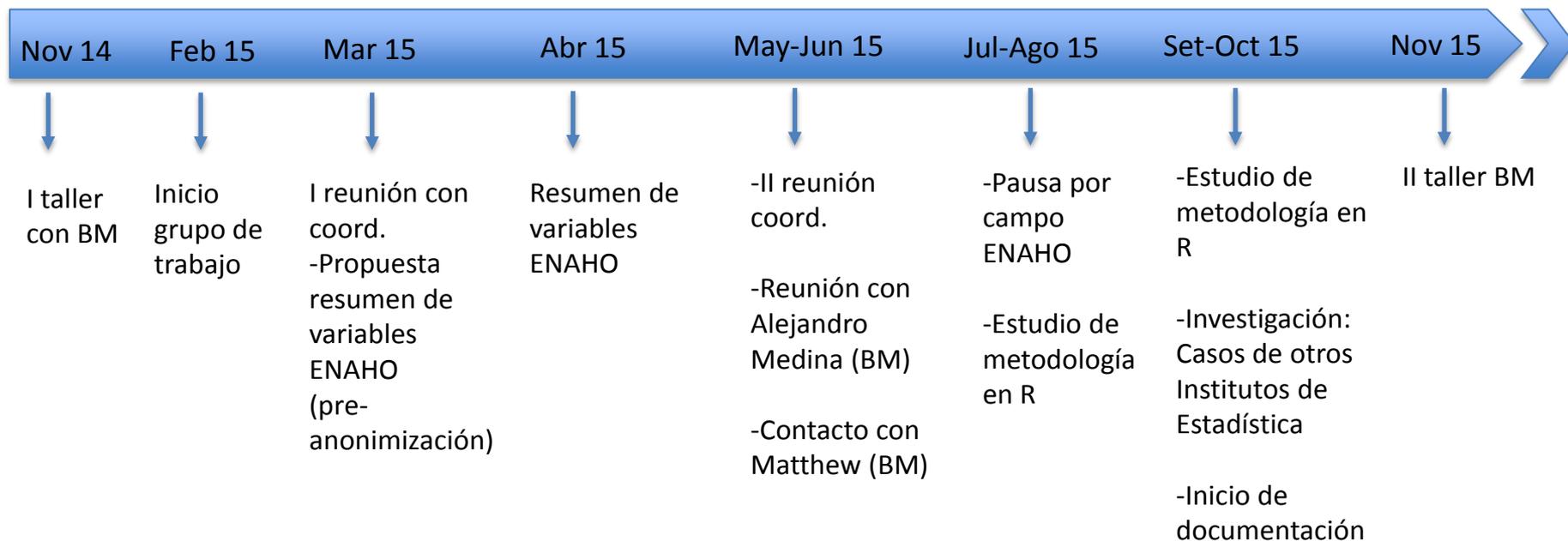
# Proceso para la liberación de bases anonimizadas



# Resultados del proceso



APRENDIZAJE!



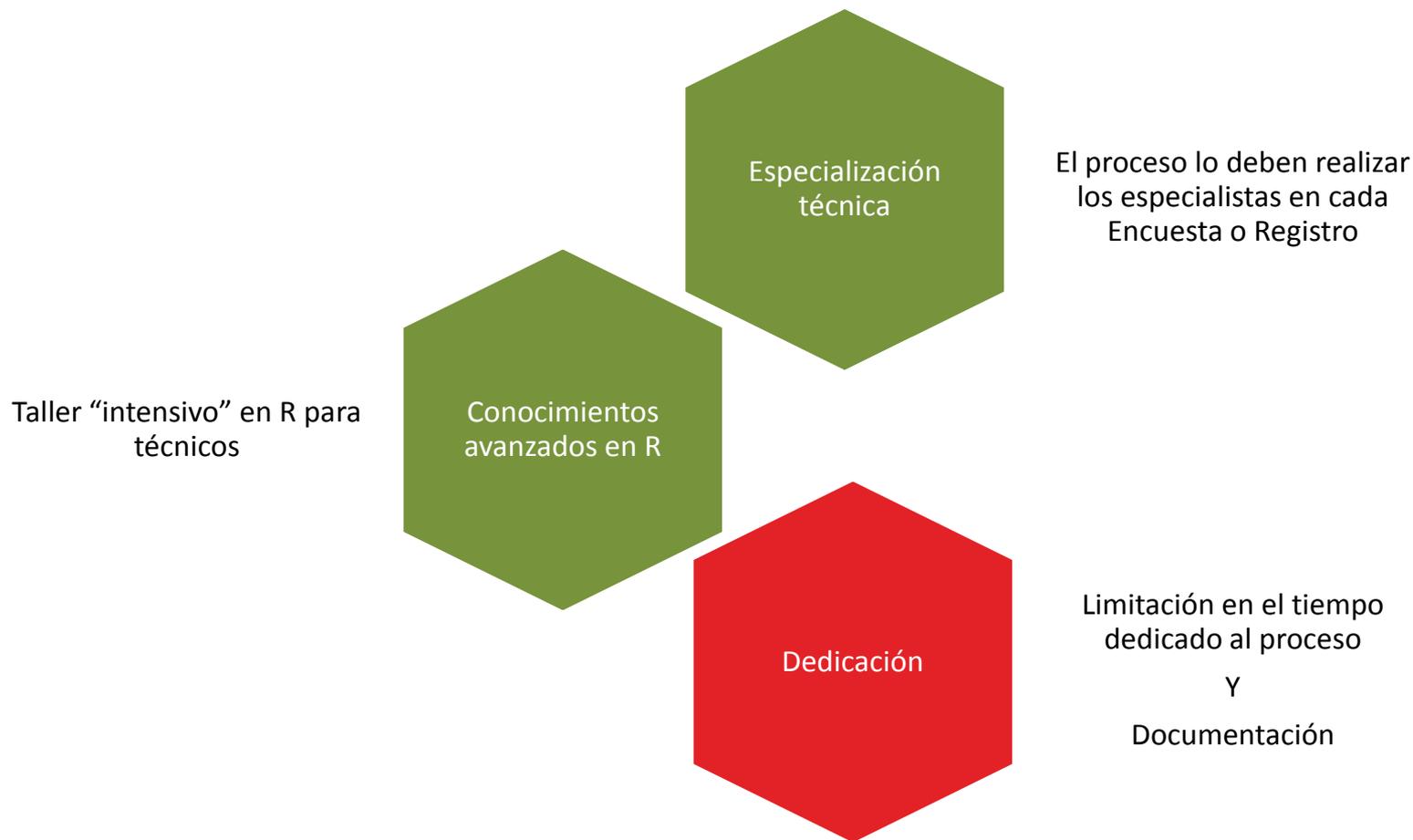


# Limitaciones

- 35%
- 50%
- 75%



# Limitaciones para implementar la anonimización:





# Sugerencias y comentarios

