

Sistema de Datos Digitales de Salud



23 marzo, 2024

UDLA, Quito



Información Práctica

- Baños
- Agua
- Wifi
- Café

La Agenda de la Mañana

Horas	Tema
9:00–9:15	Bienvenidos y Revisa la Agenda
9:15–9:45	Las infraestructuras de datos
9:45–10:00	Descanso
10:00–11:00	La estandarización de datos
11:00–11:15	Descanso
11:15–11:45	Introducción a la ciencia de datos
11:45–12:15	Datos y Ciudadanos
12:15–1:15	Almuerzos

La Agenda de la Tarde

Horas	Tema
12:15–1:15	Almuerzos
1:15–2:45	Visualización de datos
2:45–3:00	Descanso
3:00–4:30	La situación de los sistemas digitales de salud en Ecuador: avances y carencias
4:30–5:00	Preguntas y debate final

La Estandarización de Datos



¿Por qué necesitamos normalizar los datos?

Datos para Computadoras

Para utilizar las computadoras en el tratamiento y el análisis, éstos deben ser capaces de leer y procesar nuestros datos

A las computadoras puede resultarles muy difícil entender algunos tipos de datos

Los datos deben estar en formatos legibles **por máquina**

Estandarización de datos

Datos de Fuentes Múltiples

- Utilizamos datos del sistema de salud
- El sistema de salud consta de muchos subsistemas de salud
- La información sobre riesgos sanitarios procede de otros sectores

Datos de Fuentes Múltiples

- Utilizamos datos del sistema de salud
- El sistema de salud consta de muchos subsistemas de salud
- La información sobre riesgos sanitarios procede de otros sectores
 - Determinantes de la salud
 - Genética
 - Comportamiento
 - Influencias ambientales y físicas
 - Atención médica
 - Factores sociales

Gobernanza de Datos



Almacenamiento de Datos
Seguridad

Preguntas de Salud

Ingresos

Datos brutos

Estandarización

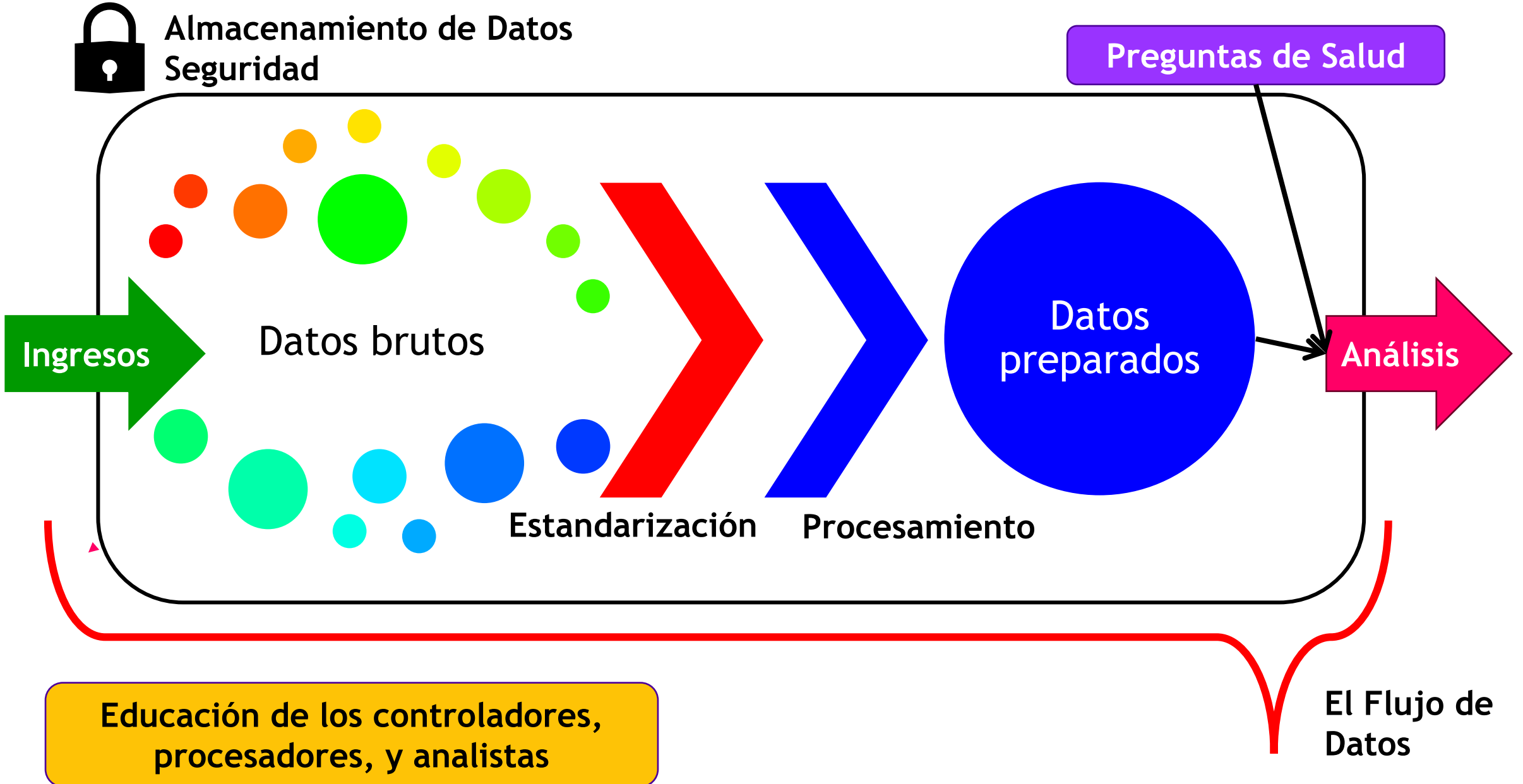
Procesamiento

Datos
preparados

Análisis

Educación de los controladores,
procesadores, y analistas

El Flujo de
Datos



El Flujo de Datos

- Para realizar un análisis, primero hay que preparar los datos
- La preparación de los datos supone un enorme esfuerzo
- Gran parte de la preparación de datos tiene que ver con **problemas de normalización y armonización**
- Si incluimos la normalización y la armonización como parte del sistema, aliviaremos la carga de los usuarios
- No elimina el procesamiento, pero mejoraría la velocidad del flujo de datos.

¿Por qué necesitamos normalizar los datos?

- Para garantizar que los datos que creamos y procesamos sean utilizables
- Facilitar la armonización de datos procedentes de distintas fuentes
- Mejorar el tratamiento posterior

Modelo Analítico Compartido

- una forma de normalizar y facilitar el intercambio, la puesta en común, el uso compartido o el almacenamiento de datos procedentes de múltiples fuentes
- Common Data Model (CDM) en ingles

Modelo Analítico Compartido

- especifican:
 - la estructura
 - el formato
 - el contenido de los datos que se van a poner en común o compartir

Observational Medical Outcomes Partnership (OMOP)

- OMOP es un MAC
- adoptado por OHDSI: una colaboración interdisciplinaria de múltiples partes interesadas dedicada a la investigación y a la mejora de la calidad

Observational Medical Outcomes Partnership (OMOP)

- OMOP es un MAC
- adoptado por OHDSI: una colaboración interdisciplinaria de múltiples partes interesadas dedicada a la investigación y a la mejora de la calidad



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Observational Medical Outcomes Partnership (OMOP)

- El grupo OHDSI crean soluciones de código abierto
- El MAC OMOP es un recurso abierto
- Desarrollado originalmente para dar cabida tanto a las reclamaciones administrativas como a los historiales médicos electrónicos
- Se ha ampliado para incluir muchas otras fuentes de datos sanitarios y relacionados con la salud

Objetivos de OMOP

- Apoye la investigación colaborativa entre fuentes de datos dentro y fuera de EE.UU.
- Pueda ser manejable para los propietarios de los datos y útil para los usuarios (eficiente para introducir y obtener datos)
- Permitir la normalización de la estructura, el contenido y los análisis centrados en casos de uso específicos

Adopción de OMOP

- 534 fuentes de datos
- 49 países diferentes



¿Cómo funciona el OMOP?

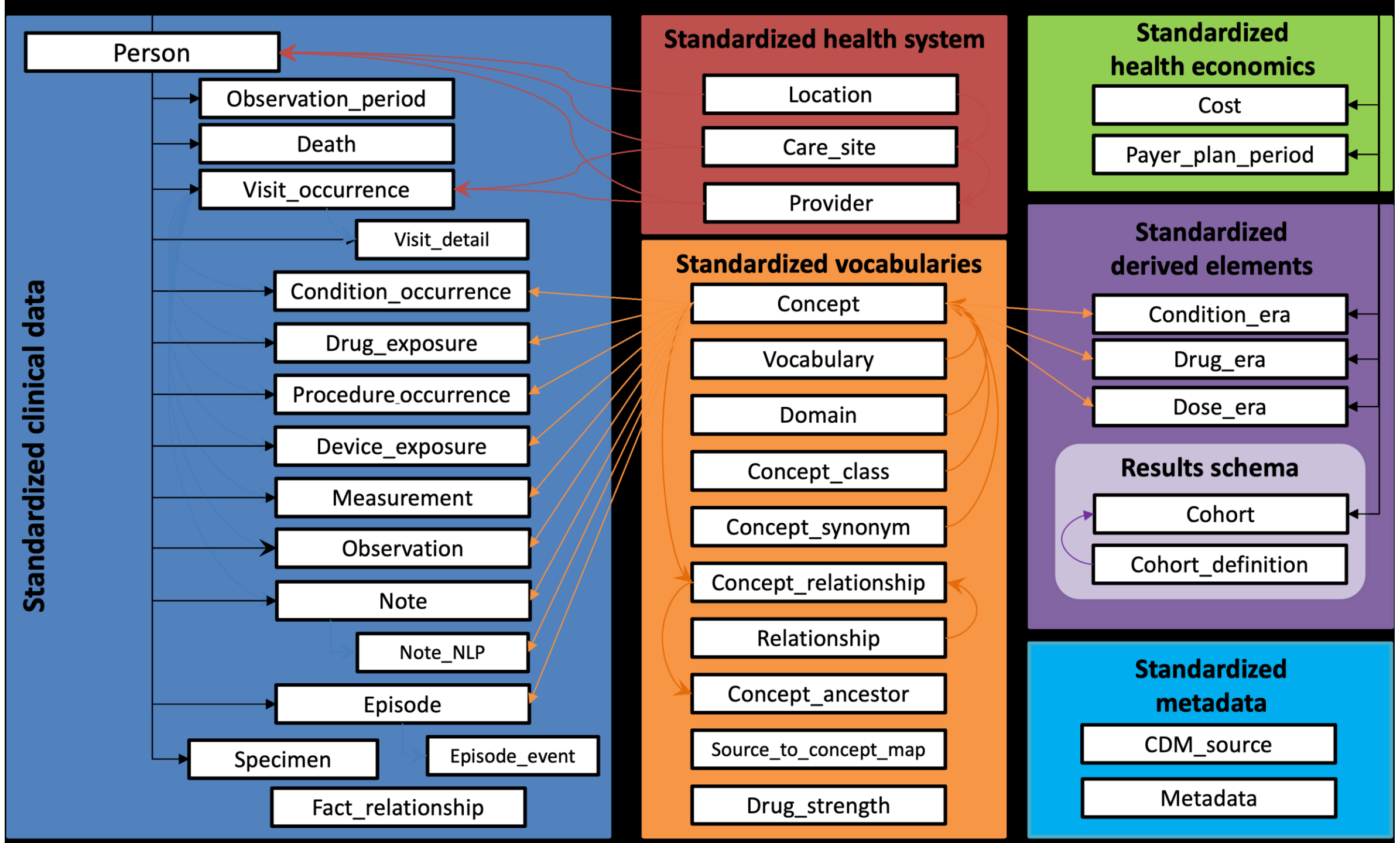
Componentes

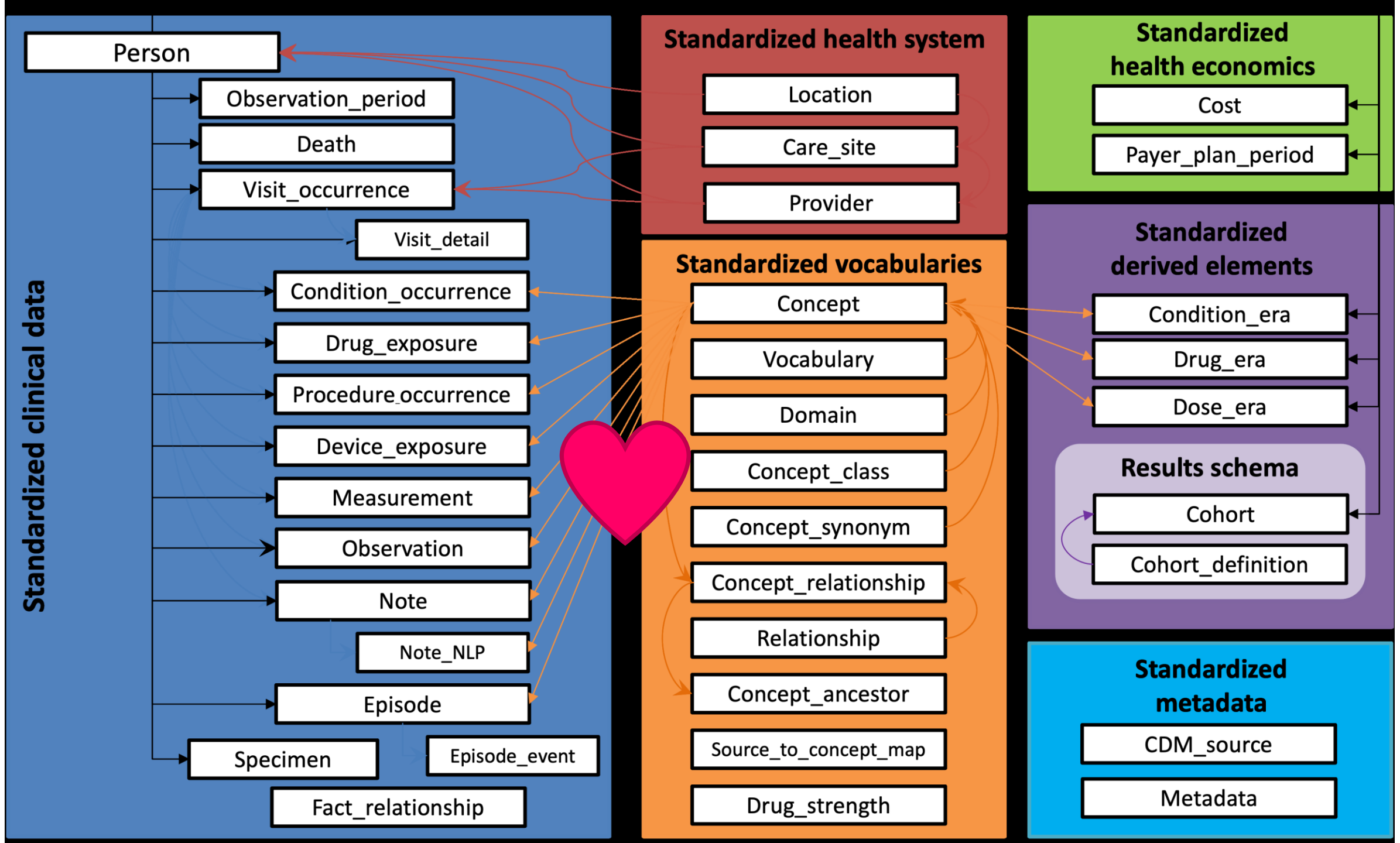
Matrices

Vocabularios

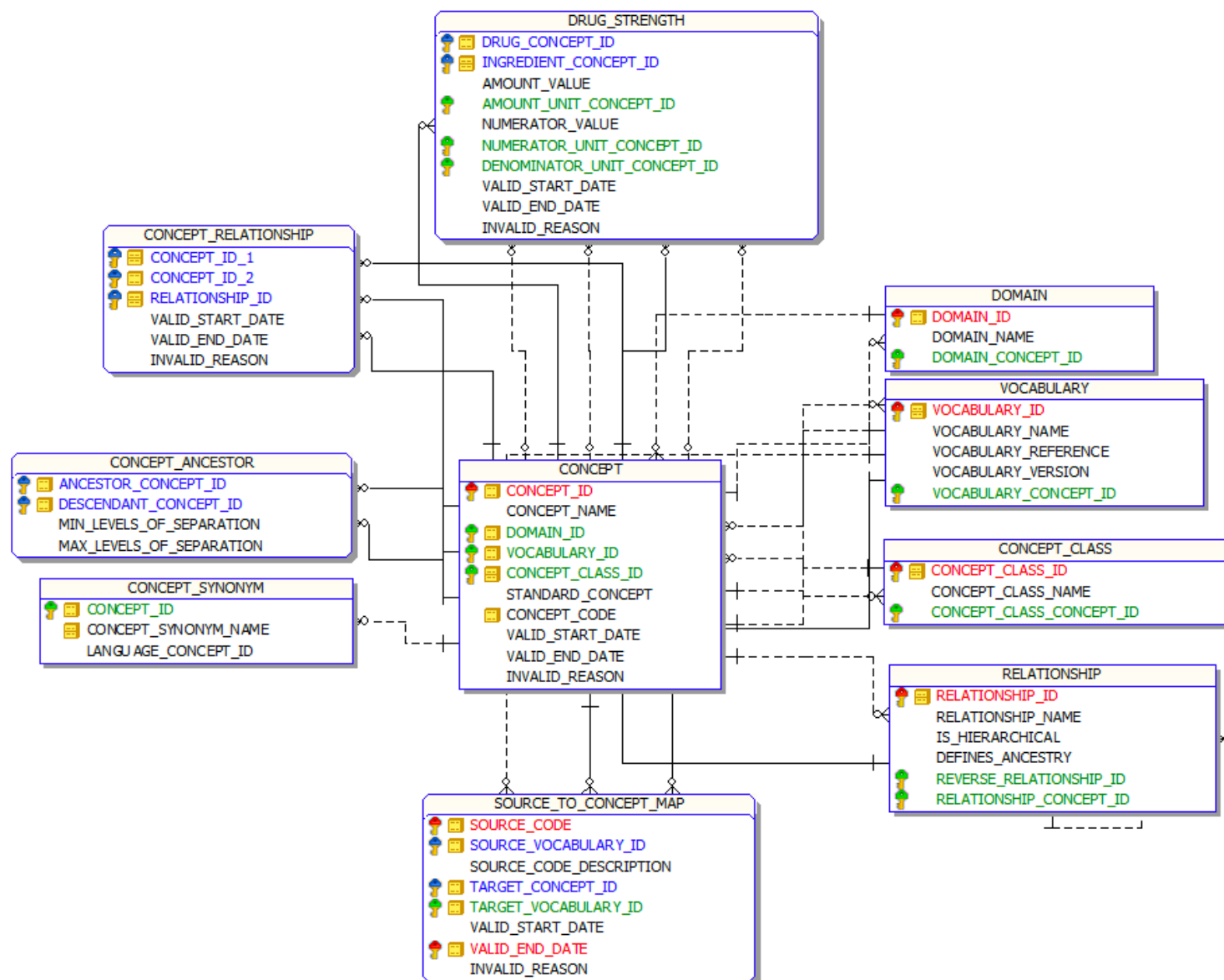
Convenciones

Examinaremos un
MAC que incluye
datos clínicos y de
economía de salud

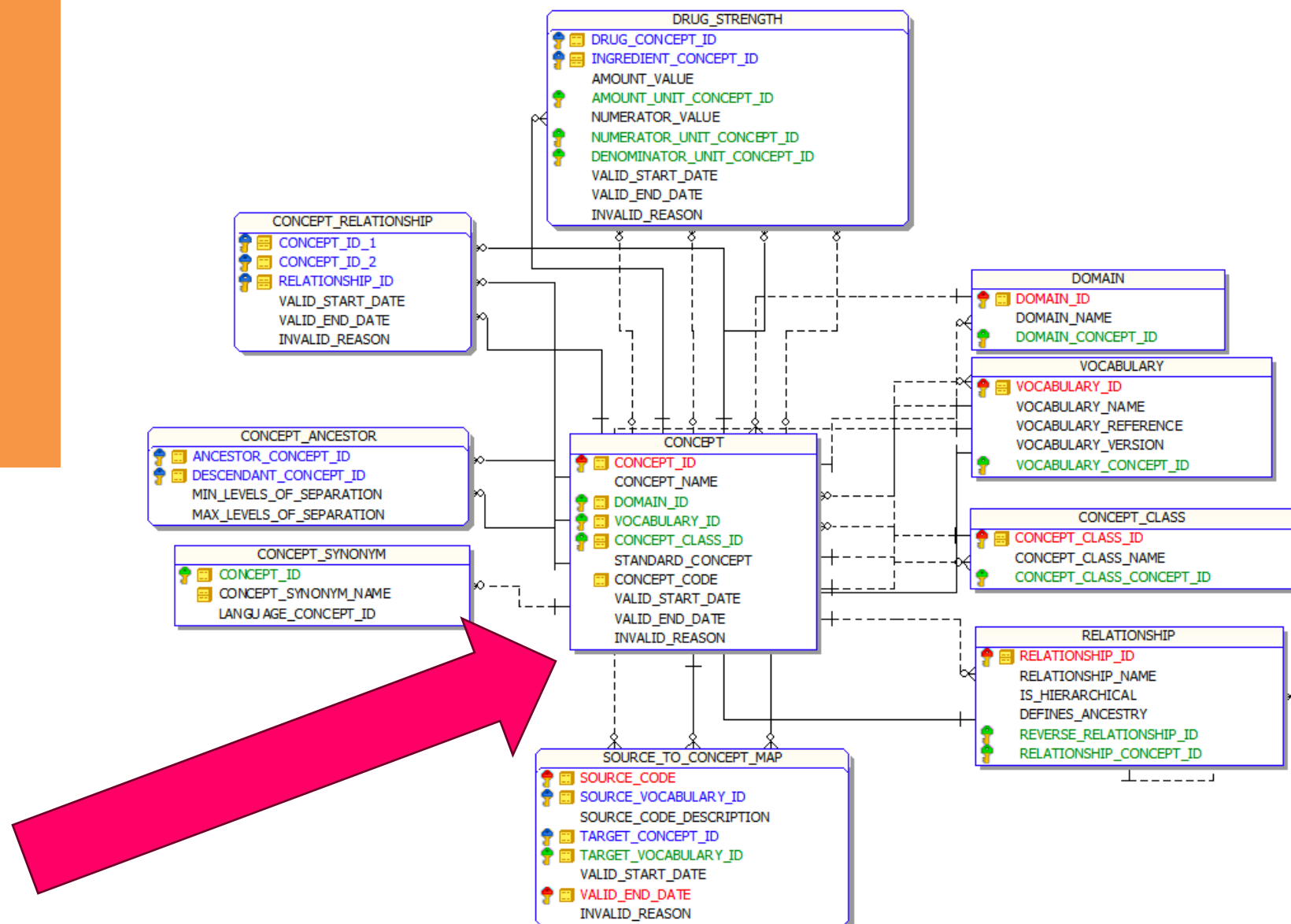




Vocabularios



Vocabularios



Vocabularios

- ¿Cuál son los conceptos?
- Es la forma más abstracta o más básica de describir los datos

Vocabularios

- ¿Cuál son los conceptos?
- Es la forma más abstracta o más básica de describir los datos
- Ejemplo:
 - Paracetamol



Vocabularios

- ¿Cuál son los conceptos?
- Es la forma más abstracta o más básica de describir los datos
- Ejemplo:
 - Paracetamol
 - Medicamento
 - Ingrediente



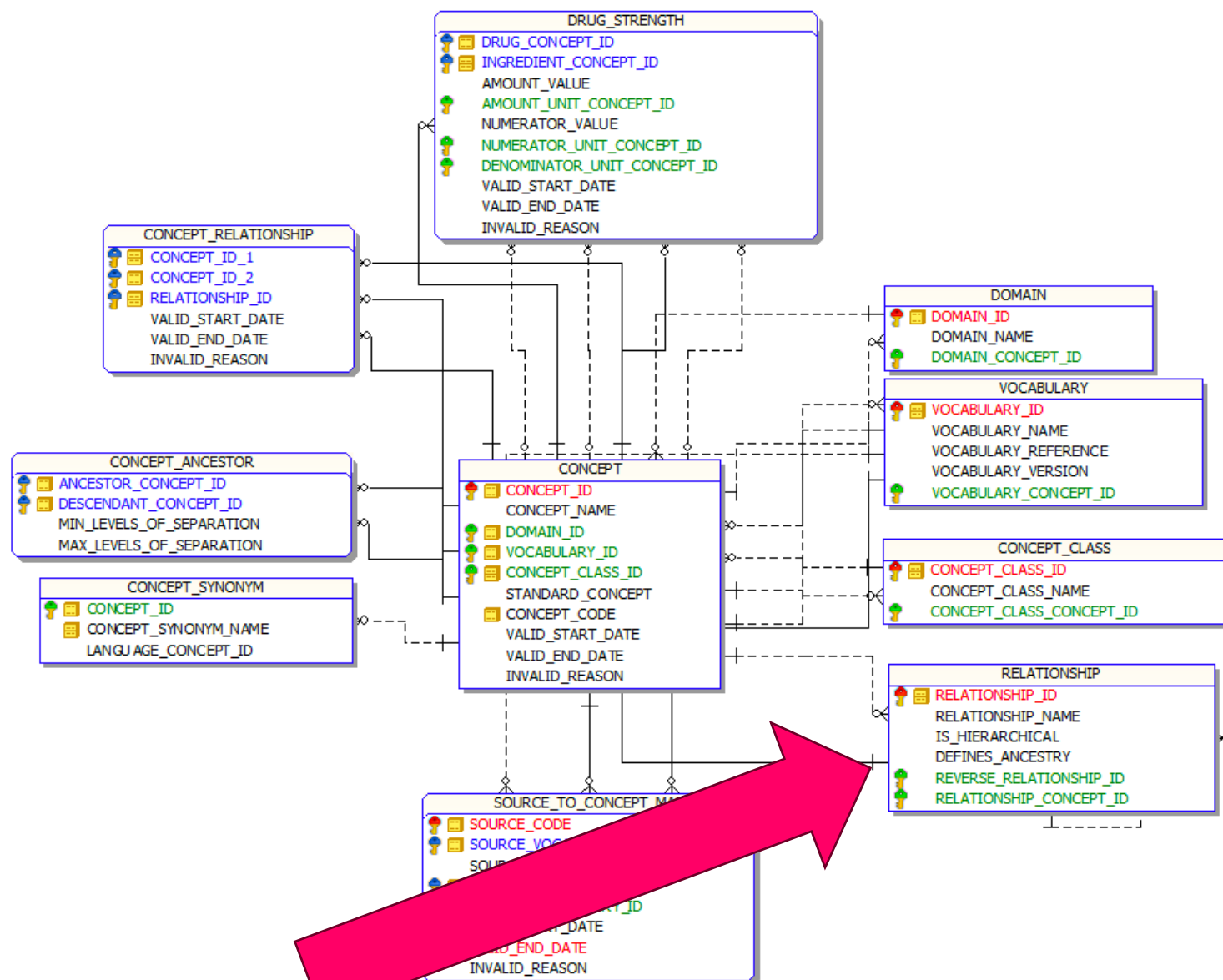
Vocabularios

- ¿Cuál son los conceptos?
- Es la forma más abstracta o más básica de describir los datos
- Ejemplo:
 - Paracetamol
 - Medicamento
 - Ingrediente

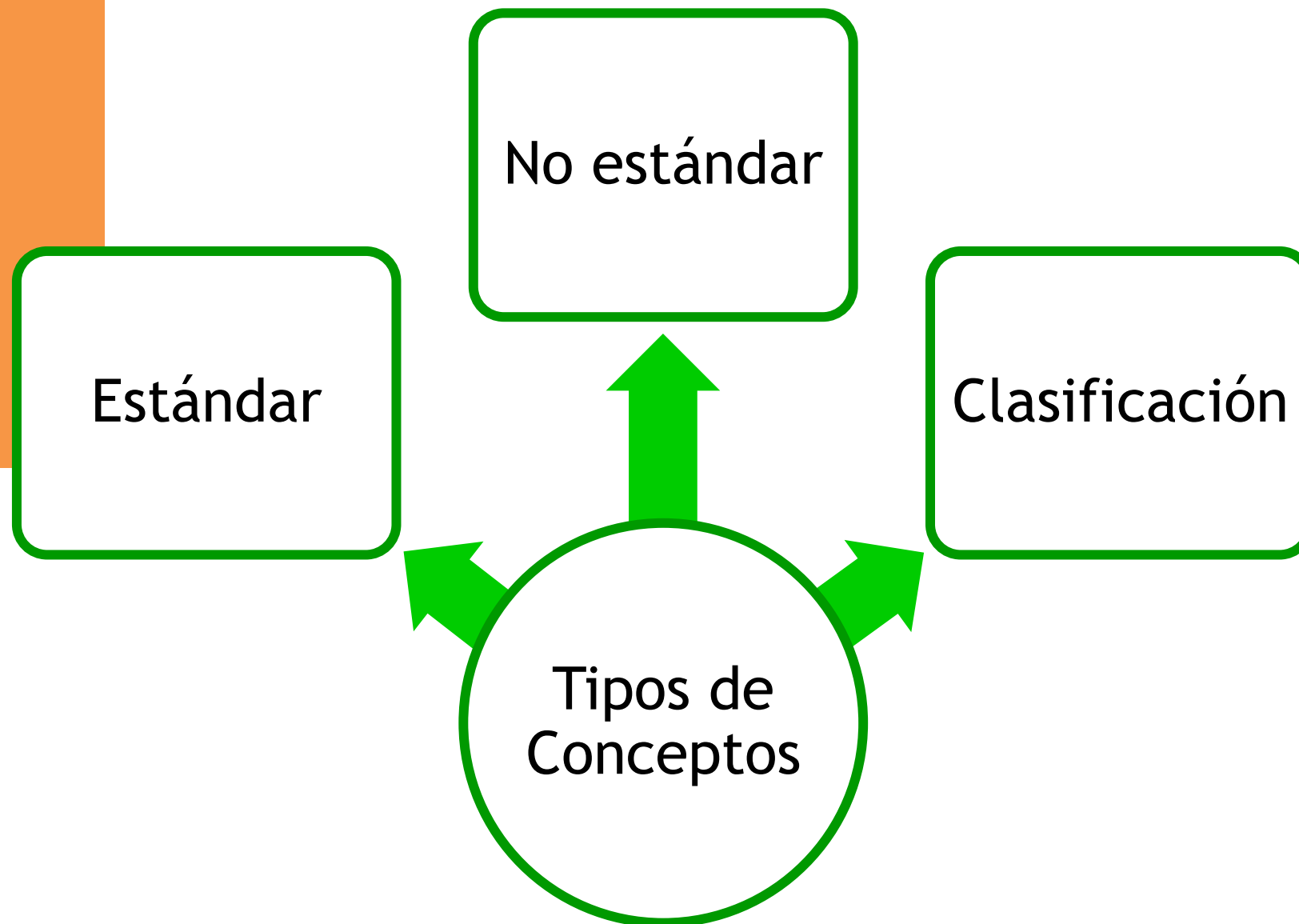
El vocabulario es un repositorio de **conceptos** que contiene **todos los conceptos** utilizados por todos los usuarios de OMOP



Vocabularios



Vocabularios



Vocabularios

Estándar

No estándar

Clasificación

Tipos de
Conceptos

Queremos asignar
nuestros datos a
estas normas



Vocabularios

Estándar

No estándar

Podemos conceptualizar los datos de muchas maneras, por eso cada sistema tiene sus propias variables

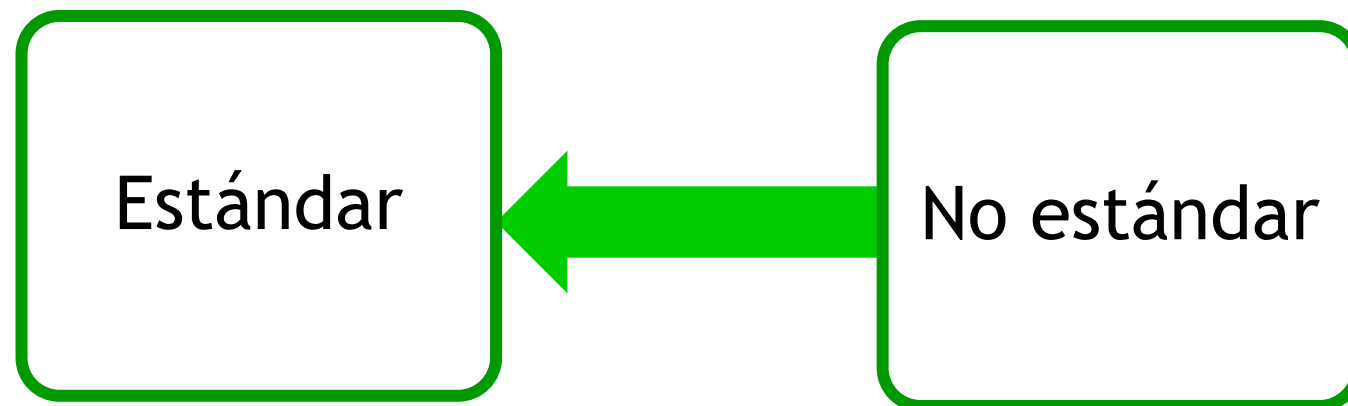
Clasificación

Tipos de
Conceptos

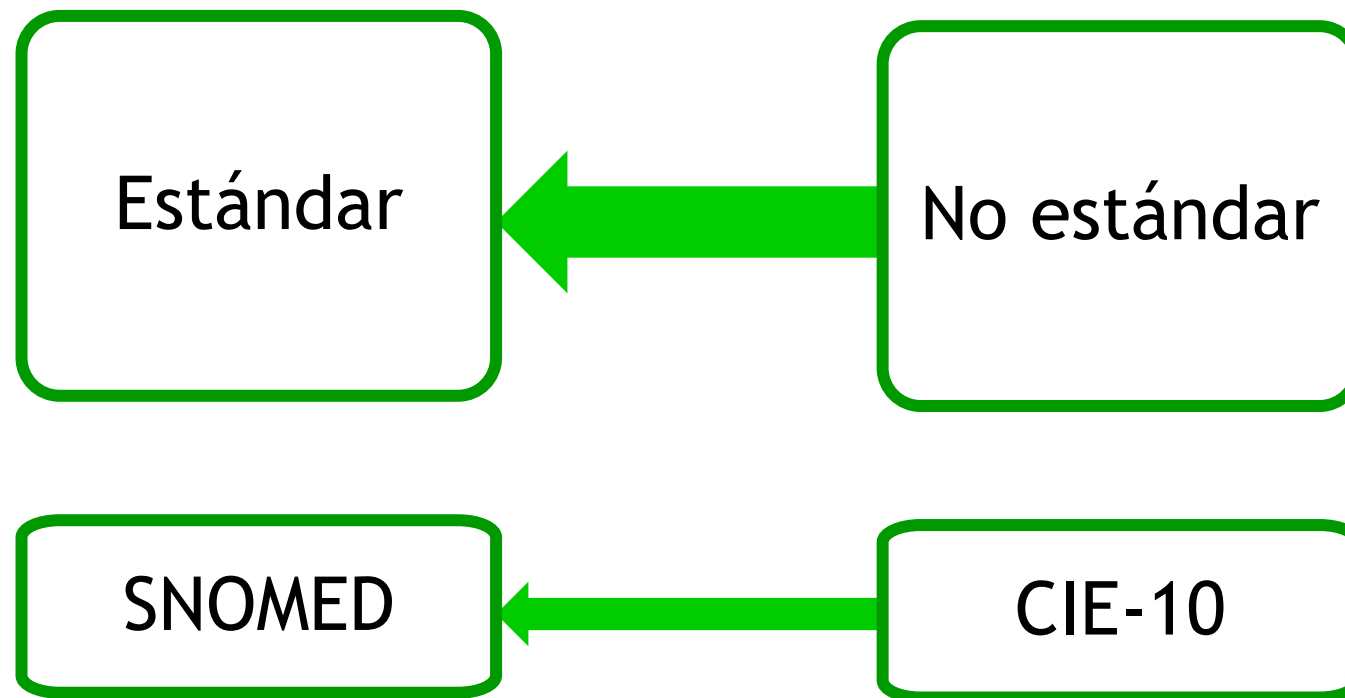
```
graph TD; A((Tipos de Conceptos)) --> B[Estándar]; A --> C[No estándar]; A --> D[Clasificación];
```

The diagram illustrates the classification of concepts into three types. At the center is a green circle labeled 'Tipos de Conceptos'. Three green arrows point outwards from this circle to three separate rounded rectangular boxes. The left box is labeled 'Estándar', the top box is labeled 'No estándar', and the right box is labeled 'Clasificación'. The 'No estándar' box is highlighted with a thick green border, and a text block to its right explains that data can be conceptualized in many ways, leading to different variables in different systems. An orange rectangle with the word 'Vocabularios' is positioned in the top-left corner of the slide.

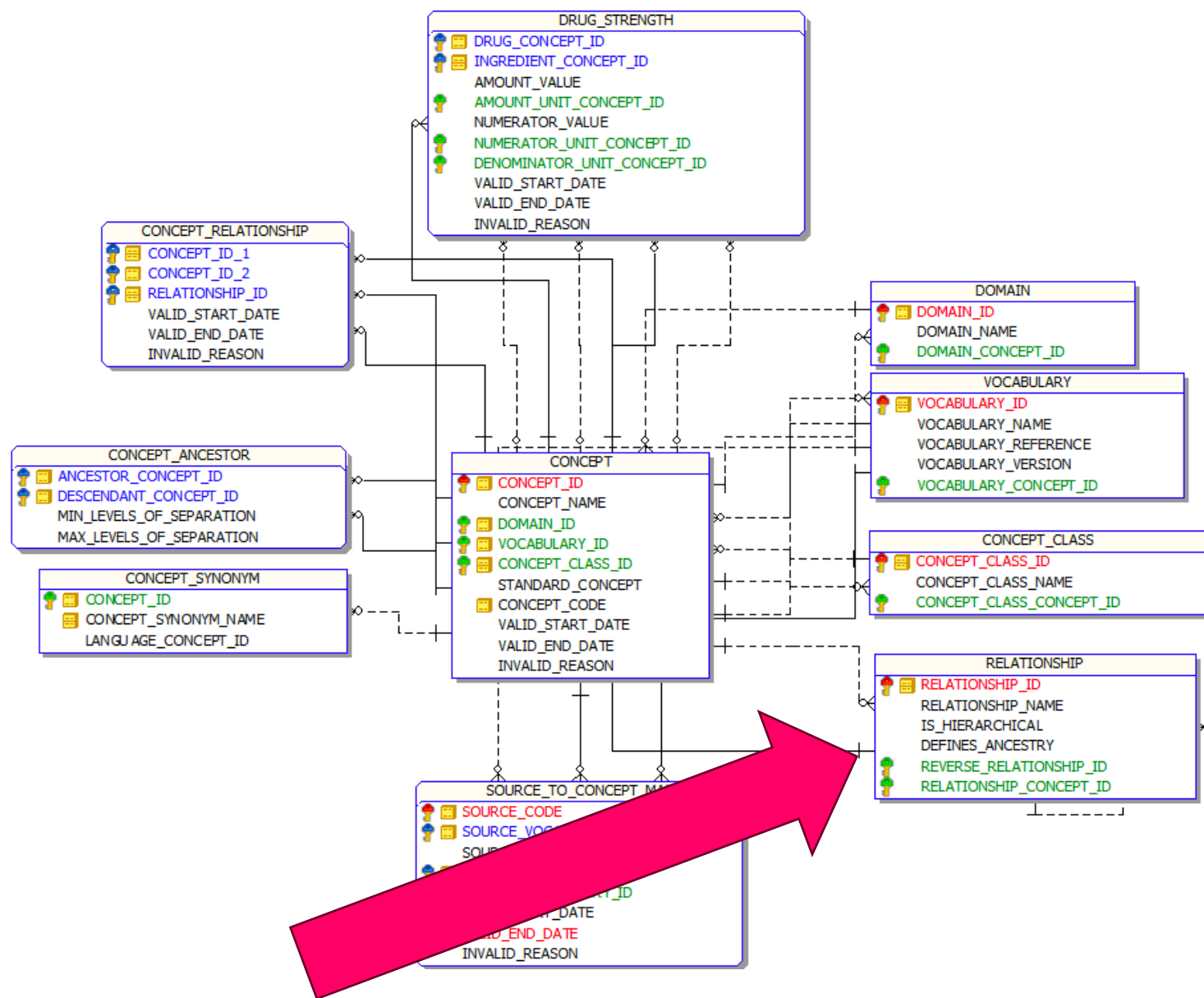
Vocabularios



Vocabularios



Vocabularios



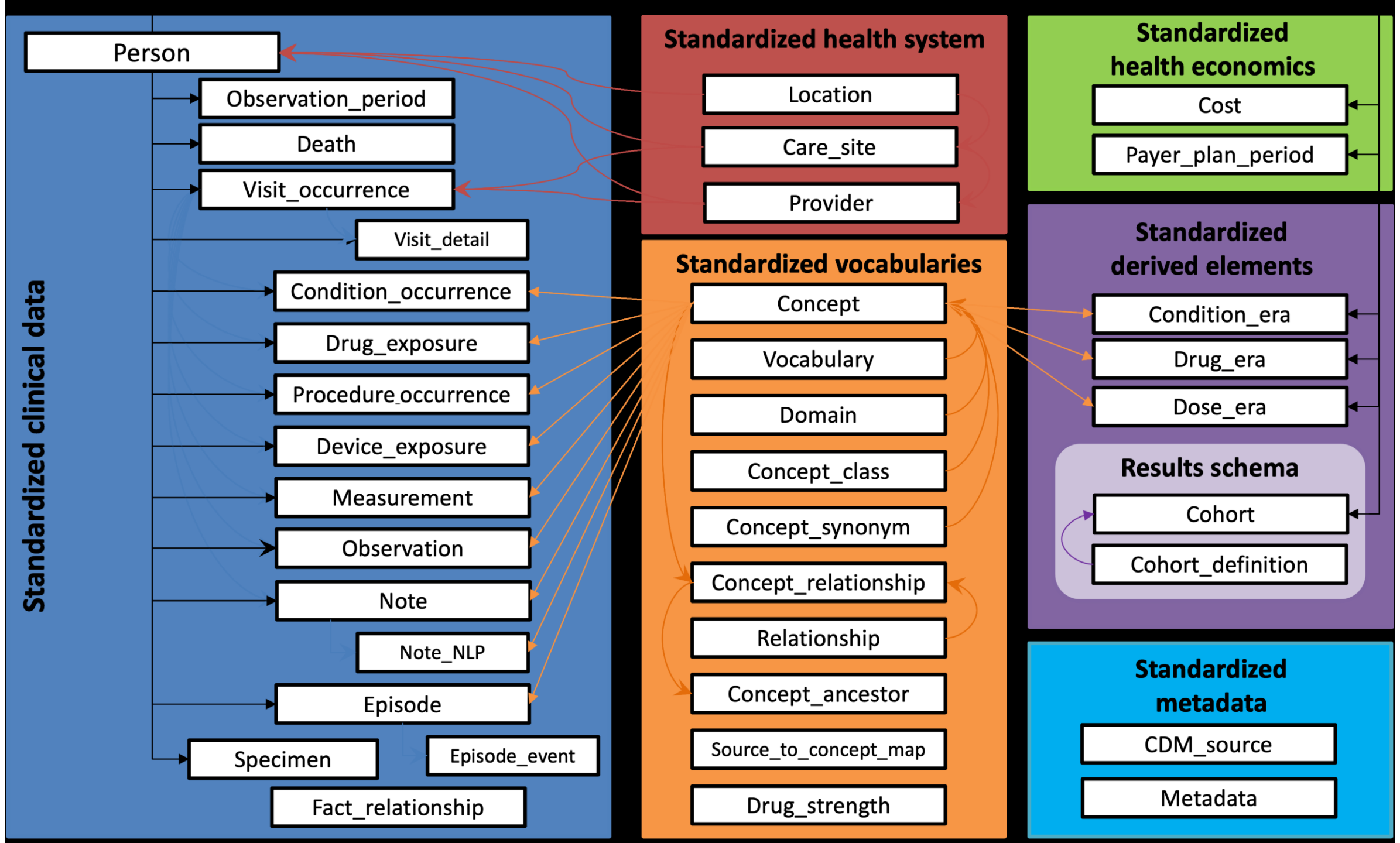
¿Cómo funciona el OMOP?

Componentes

Matrices

Vocabularios

Convenciones



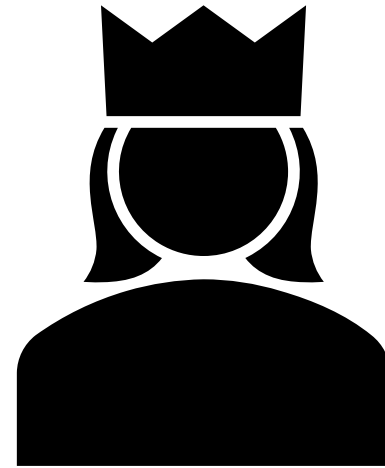
Colección Principal: Datos de Clínica

- OMOP es un modelo centrado en la persona



Colección Principal: Datos de Clínica

- OMOP es un modelo centrado en la persona
 - intentamos asociar hechos sanitarios a una persona concreta



¿Cómo funciona el OMOP?

Componentes

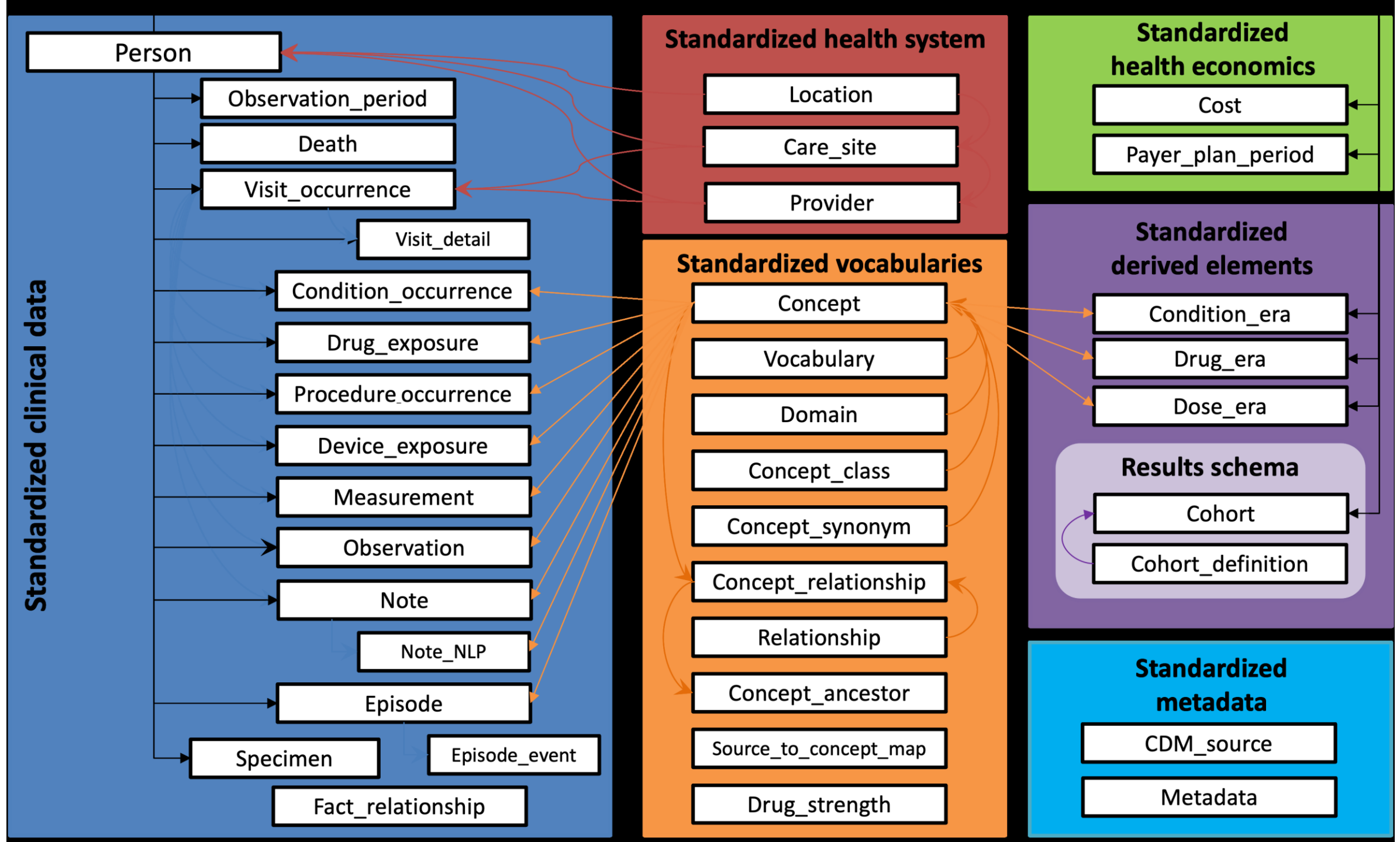
Matrices

Vocabularios

Convenciones

Convenciones

- Proporciona estructura del MAC
- ¿Qué esperamos que haya en cada colección?



Convenciones

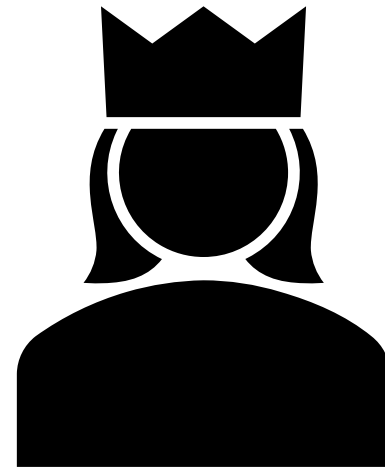
- Proporciona estructura del MAC
- ¿Qué esperamos que haya en cada colección?

Convenciones

- Proporciona estructura del MAC
- ¿Qué esperamos que haya en cada colección?
 - por convención, ciertas matrices deben estar en ciertas colecciones

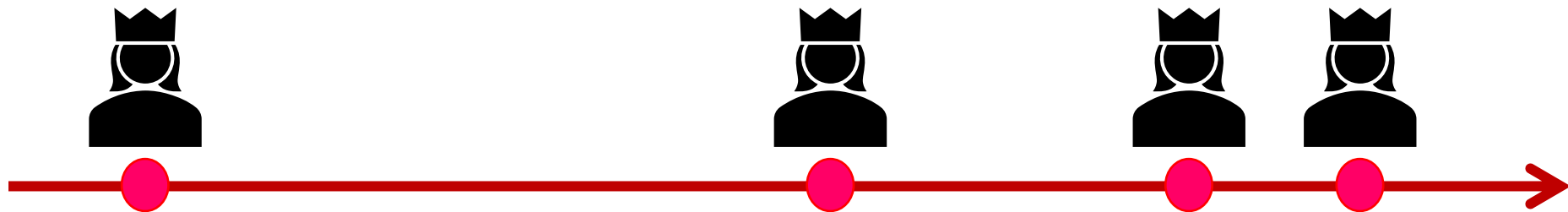
Convenciones de la Colección Principal

- Siempre
 - Matriz de persona
 - Matriz de periódico de observación



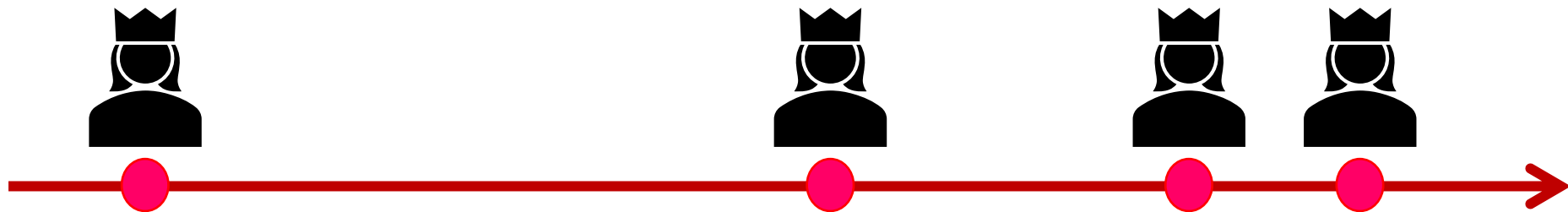
Convenciones de la Colección Principal

- Siempre
 - Matriz de persona
 - Matriz de periódico de observación



Convenciones de la Colección Principal

- Siempre
 - Matriz de persona
 - Matriz de periódico de observación

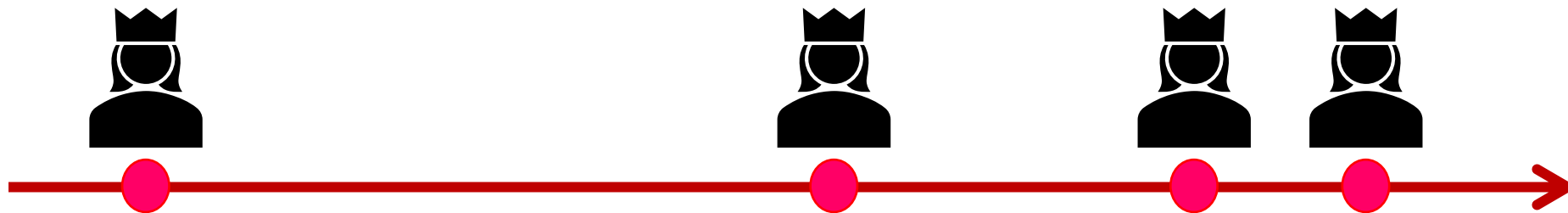


Convenciones

- Proporciona estructura del MAC
- ¿Qué esperamos que haya en cada colección?
 - por convención, ciertas matrices deben estar en ciertas colecciones
- ¿Qué esperamos que haya en cada matriz?
 - por convención, ciertos campos deben estar en ciertas matrices

Convenciones de la Colección Principal

- Siempre
 - Matriz de persona
 - Matriz de periódico de observación
- Matriz de persona
 - _ID_PERSONA
 - _VALOR_FUENTE
 - ... i otras



¿Cómo funciona el OMOP?

Componentes

Matrices

Vocabularios

Convenciones

Ámbitos en el Vocabulario



OMOP: Recurso Abierto

- 11.027.290 conceptos
- 142 vocabularios
- 44 ámbitos de vocabulario
- 82.142.038 relaciones de conceptos
- 87.967.689 ancestros de conceptos

OMOP: Recurso Abierto

- Se ha probado con otros MACs y ha dado los mejores resultados
 - Exhaustividad
 - Integridad
 - Flexibilidad
 - Sencillez
 - Integración
 - Implementabilidad
- Organizaciones que lo utilizan
 - Hospital
 - Universidad
 - Direcciones del ministerio



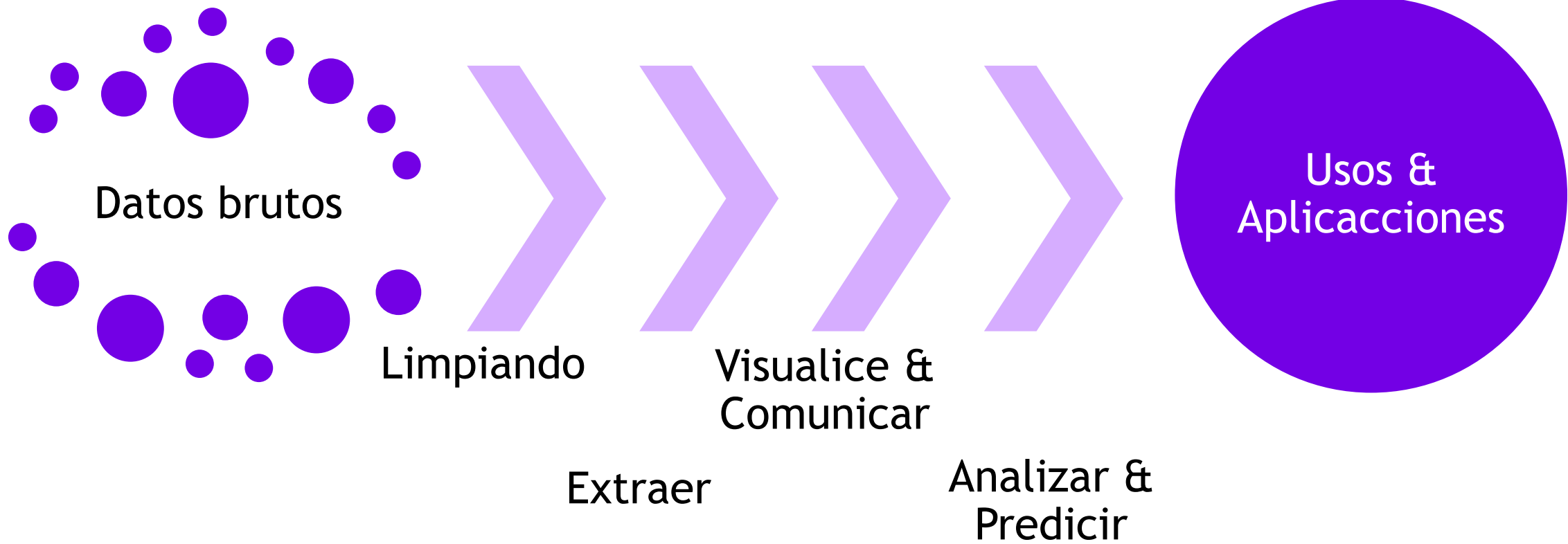
Descanso

Introducción a la Ciencia de Datos



¿Cuál es la ciencia de datos?

- Conjunto de habilidades para gestionar, procesar, utilizar y analizar datos
- Proceso de utilización de datos para resolver problemas y responder preguntas
 - Canalización de datos (“pipelines”)



Definición de los Datos

- Componente fundamental de la información
- Información traducida a una forma fácil de procesar y transferir
- Recopilación de hechos

¿Que son los Datos?

- Números
- Palabras
- Medidas
- Observaciones
- Descripciones

Datos Cuantitativos

Numérico

- Cualquier valor
- Valores continuos
- Cantidades medidas

Discreto

- Sólo determinados valores
- Números enteros
- Cantidades contadas

Datos Cualitativos

- Descripción o representación
- Texto
 - Forma libre
 - Clases: ordinal, categorías
- Gráfico

Forma de Datos

Datos Estáticos

- Registros
- Números
- Valores
- Gráficos

Datos Dinámicos

- Audio
- Video

Datos Grandotes: “Big Data”

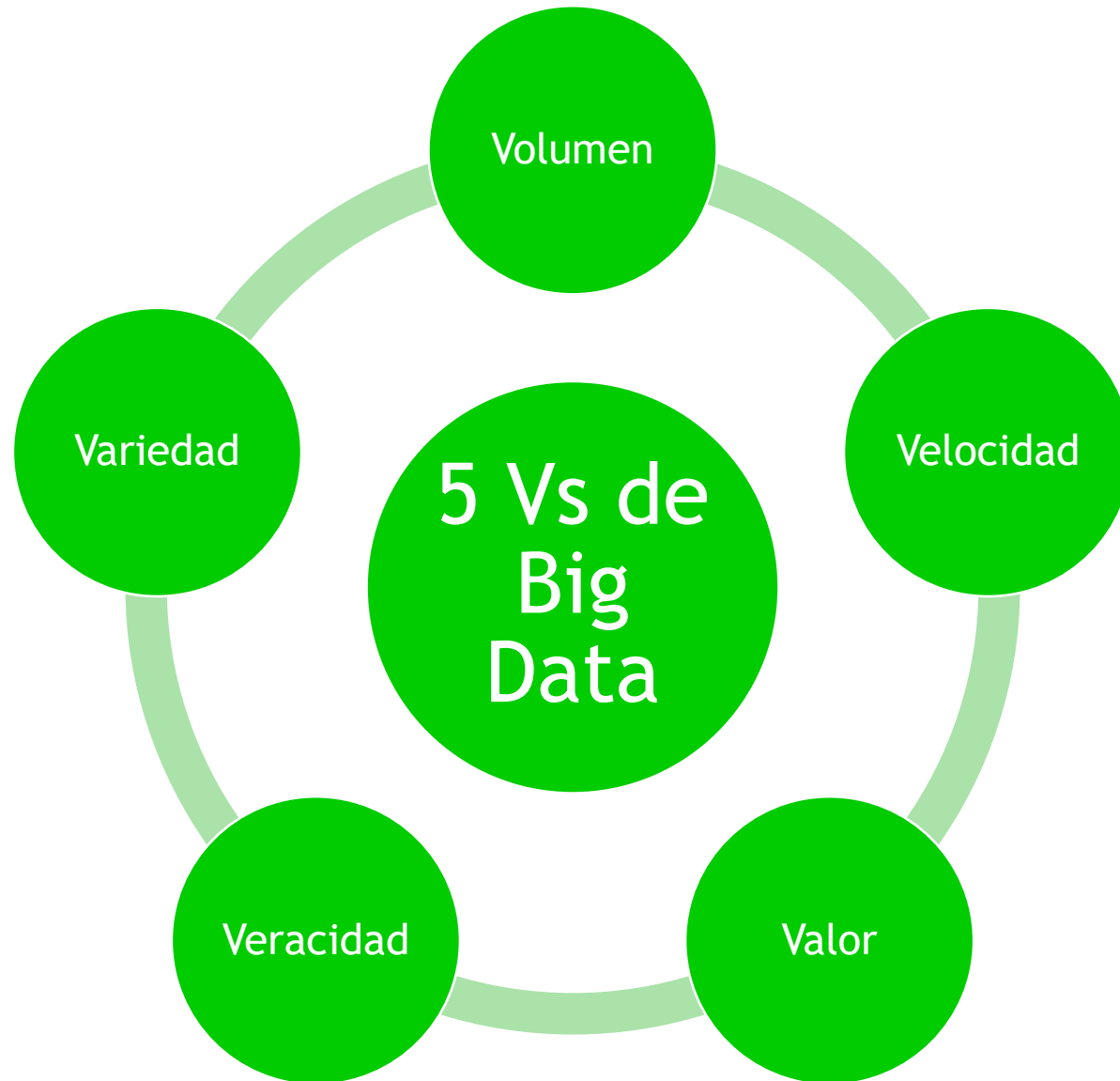
- Todo el mundo genera enormes cantidades de datos
 - Compras: en línea, tarjetas de débito, tarjetas de crédito
 - Comunicación: teléfonos móviles, fijos, correo electrónico, videoconferencia
 - Redes sociales: Facebook, Instagram
 - Entretenimiento: Netflix, YouTube
 - Uso de Internet: búsquedas, interacción con empresas
- Todos los sistemas de la vida cotidiana moderna generan datos

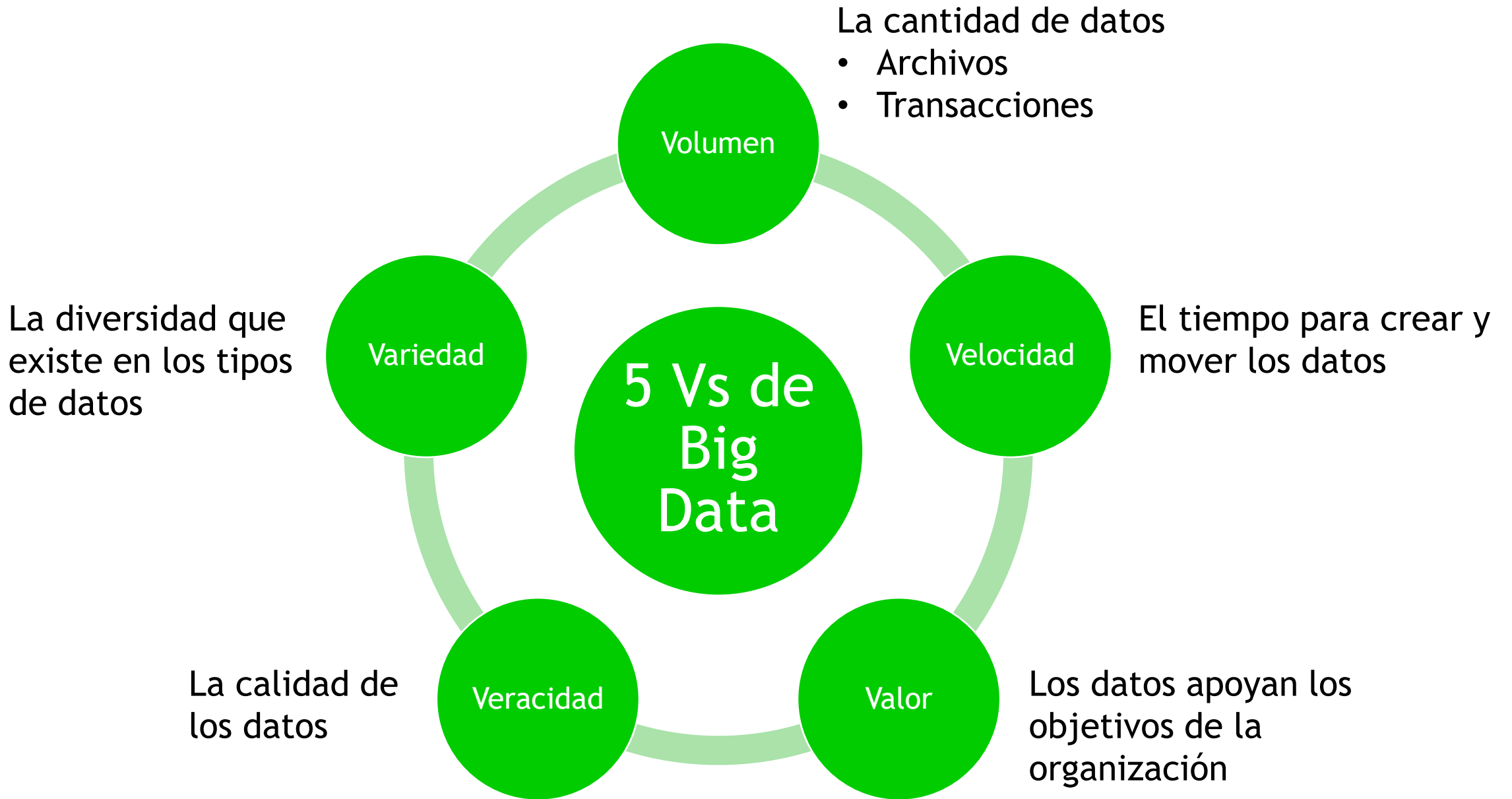
Datos Grandotes de Salud

- ¿Qué tipos de datos se generan durante la asistencia sanitaria?
- ¿Cómo podemos acceder, manipular, analizar y utilizar estos datos generados continuamente?

Datos Grandotes: “Big Data”

- Gran volumen de datos
 - Estructurados
 - No estructurados
- Los macrodatos son útiles porque hay suficientes datos para crear conocimiento y comprensión.
- Se pueden analizar para obtener información que ayude a tomar mejores decisiones.





Datos Enlazados

- Uso de Internet u otras redes para conectar datos relacionados
- Mejores prácticas recomendadas para exponer, compartir y conectar datos

Datos Abiertos (“Open Data”)

- Datos a los que cualquiera puede acceder, utilizar, modificar y compartir libremente con cualquier fin
- Disponibles y accesibles
 - Disponibles a un coste razonable
 - Preferiblemente descargables por Internet
 - Cómodos y modificables

Datos Abiertos

- Reutilización y redistribución
 - Se permite la reutilización
 - Se permite añadir a otros conjuntos de datos
- Participación universal
 - Todo el mundo debe poder utilizarla
 - No se discrimina por campos de actividad, personas o grupos.

Datos Abiertos

- Datos comunicados

Big Data

- Datos privados utilizados en negocios y seguridad nacional

- Grandes conjuntos de datos procedentes de la investigación científica y las redes sociales

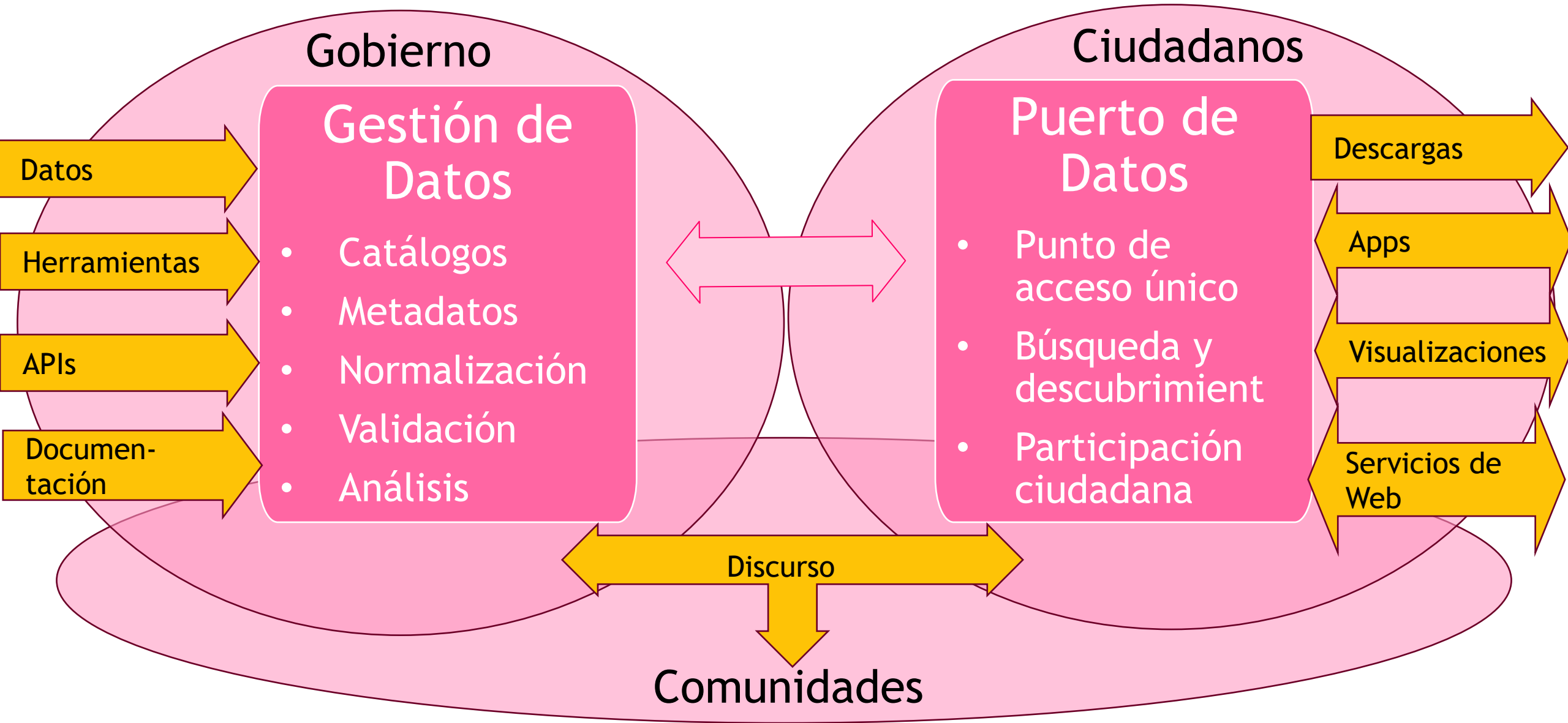
Grandes conjuntos de datos públicos gubernamentales

- Datos públicos de la administración regional

Gobierno Abierto

- Compromiso con los ciudadanos

Plataforma de Gobierno Abierto



Datos Grandes y Abiertos de Salud

Datos Grandes y Abiertos de Salud

[https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(23\)00157-7/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(23)00157-7/fulltext)

Health Policy



Mapping and evaluating national data flows: transparency, privacy, and guiding infrastructural transformation



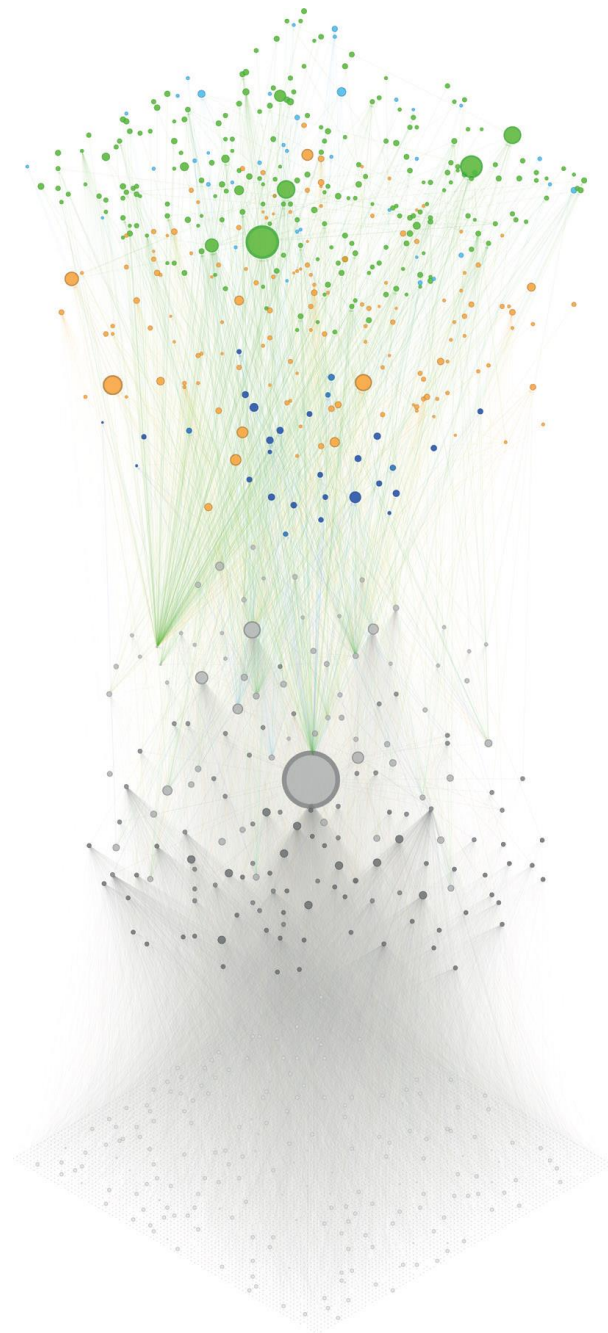
Joe Zhang, Jess Morley, Jack Gallifant, Chris Oddy, James T Teo, Hutan Ashrafian, Brendan Delaney, Ara Darzi



The importance of big health data is recognised worldwide. Most UK National Health Service (NHS) care interactions are recorded in electronic health records, resulting in an unmatched potential for population-level datasets. However, policy reviews have highlighted challenges from a complex data-sharing landscape relating to transparency, privacy, and analysis capabilities. In response, we used public information sources to map all electronic patient data flows across England, from providers to more than 460 subsequent academic, commercial, and public data consumers. Although NHS data support a global research ecosystem, we found that multistage data flow chains limit transparency and risk public trust, most data interactions do not fulfil recommended best practices for safe data access, and existing infrastructure produces aggregation of duplicate data assets, thus limiting diversity of data and added value to end users. We provide recommendations to support data infrastructure transformation and have produced a website (<https://DataInsights.uk>) to promote transparency and showcase NHS data assets.

Lancet Digit Health 2023; 5: e737–48

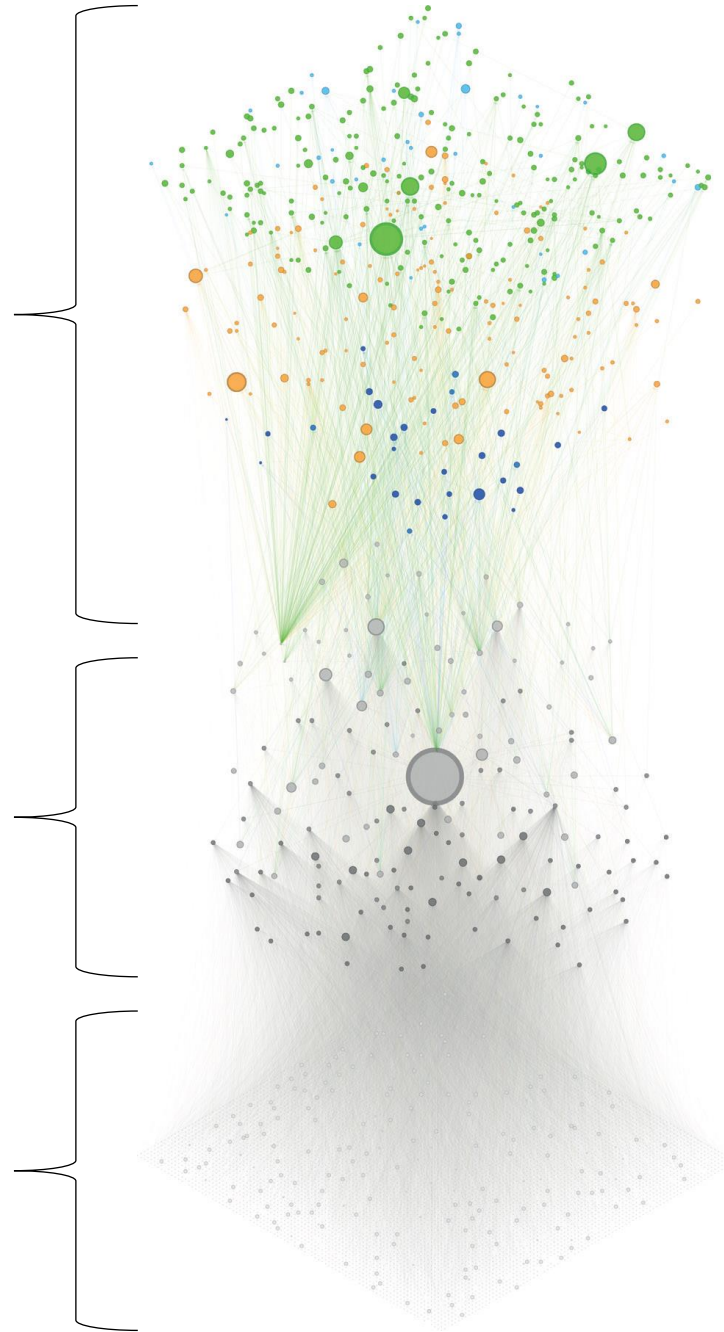
Institute of Global Health Innovation, Imperial College London, London, UK (J Zhang BMBCh, Prof H Ashrafian PhD, Prof B Delaney PhD, Prof A Darzi PhD); Department of Critical Care (J Zhang) and London Medical Imaging and AI Centre (Prof J T Teo PhD), Guy's and St Thomas' NHS



Consumadores de Datos

Extractores de Datos

Fuentes de Datos



Consumadores de Datos

Universidades
Centros de investigación

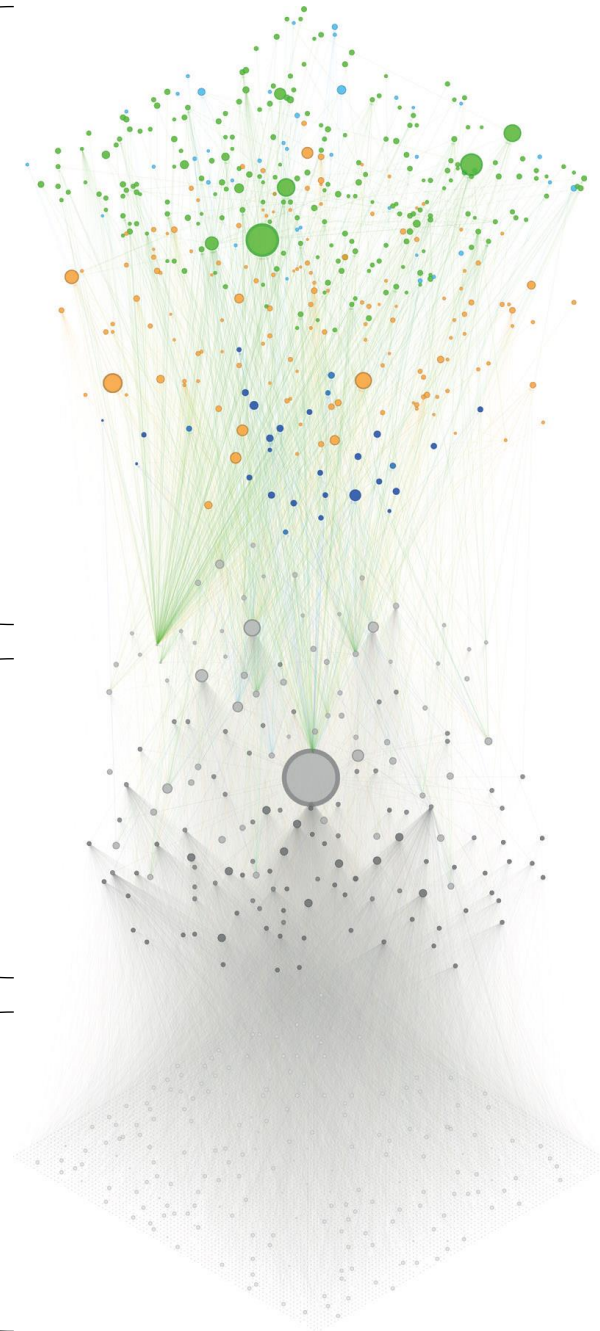
Negocios privados

Extractores de Datos

NHS
RU
Sin ánimo de lucro

Fuentes de Datos

Fuentes



Consumadores de Datos

Universidades
Centros de investigación

Negocios privados

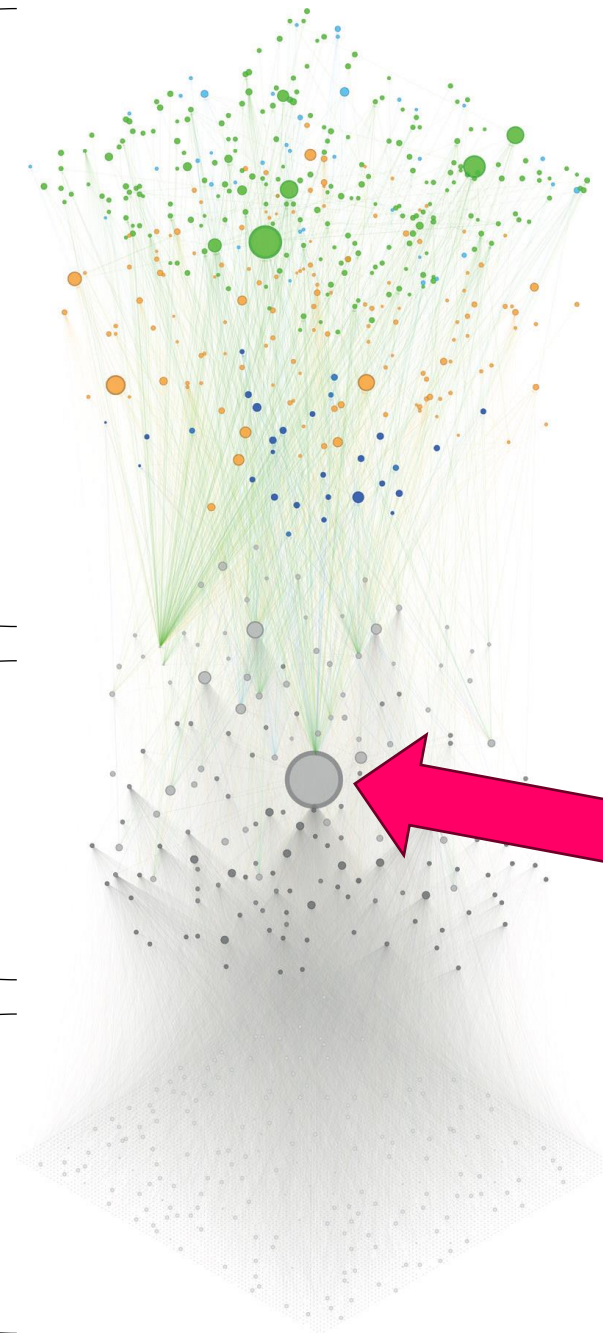
NHS
RU
Sin ánimo de lucro

Extractores de Datos

NHS Digital

Fuentes de Datos

Fuentes



Resumen de las Recomendaciones de la Estrategia de Datos de Salud

- **Confianza pública en el uso de los datos de salud**
 - Mejorar la transparencia y fomentar la participación de pacientes y ciudadanos
 - Pasar a la analítica dentro de entornos de datos controlados y seguros
 - Reconsiderar los modelos de gobernanza y el enfoque de los datos desidentificados

Resumen de las Recomendaciones de la Estrategia de Datos de Salud

- **Transformación de las infraestructuras**
 - Los datos deben ser una capacidad centralizada del Servicio Nacional de Salud, y los flujos de datos deben ser descubiertos, mapeados y racionalizados
 - Nuevas soluciones infraestructurales para investigar y reducir las desigualdades digitales y de datos, y evitar la exclusión digital

Resumen de las Recomendaciones de la Estrategia de Datos de Salud

- **Futuras capacidades basadas en datos**
 - Desarrollar datos multimodales, incluida la genómica, para potenciar a los investigadores y la medicina personalizada
 - Nuevas orientaciones e infraestructuras para apoyar una colaboración comercial segura con los sectores de ciencias de la vida, tecnología sanitaria y farmacéutico
 - Apoyar a los responsables de la toma de decisiones clínicas a todos los niveles y aprovechar las tecnologías de inteligencia artificial



Descanso

La Visualización de Datos



Los Datos son Abundantes

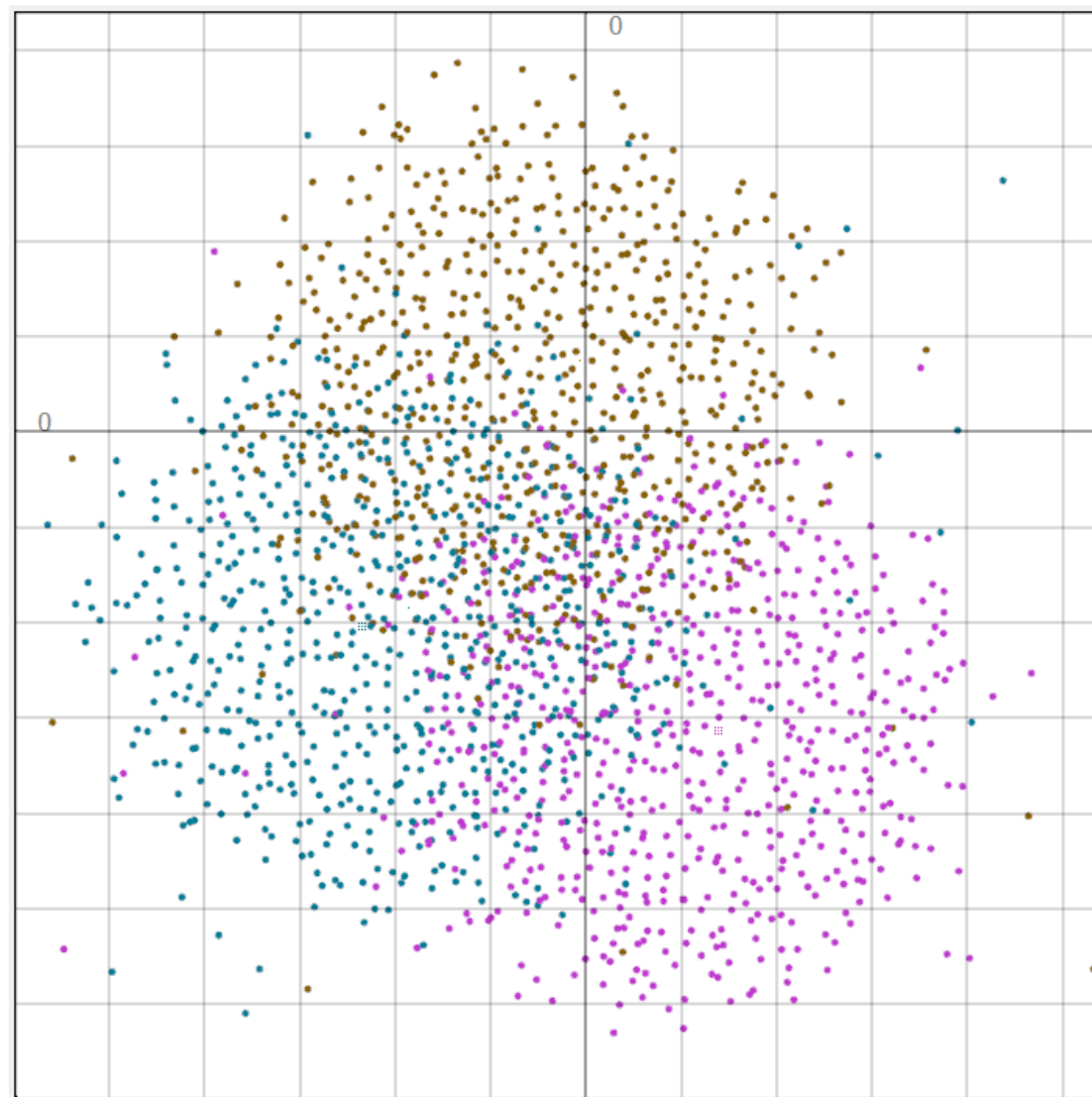
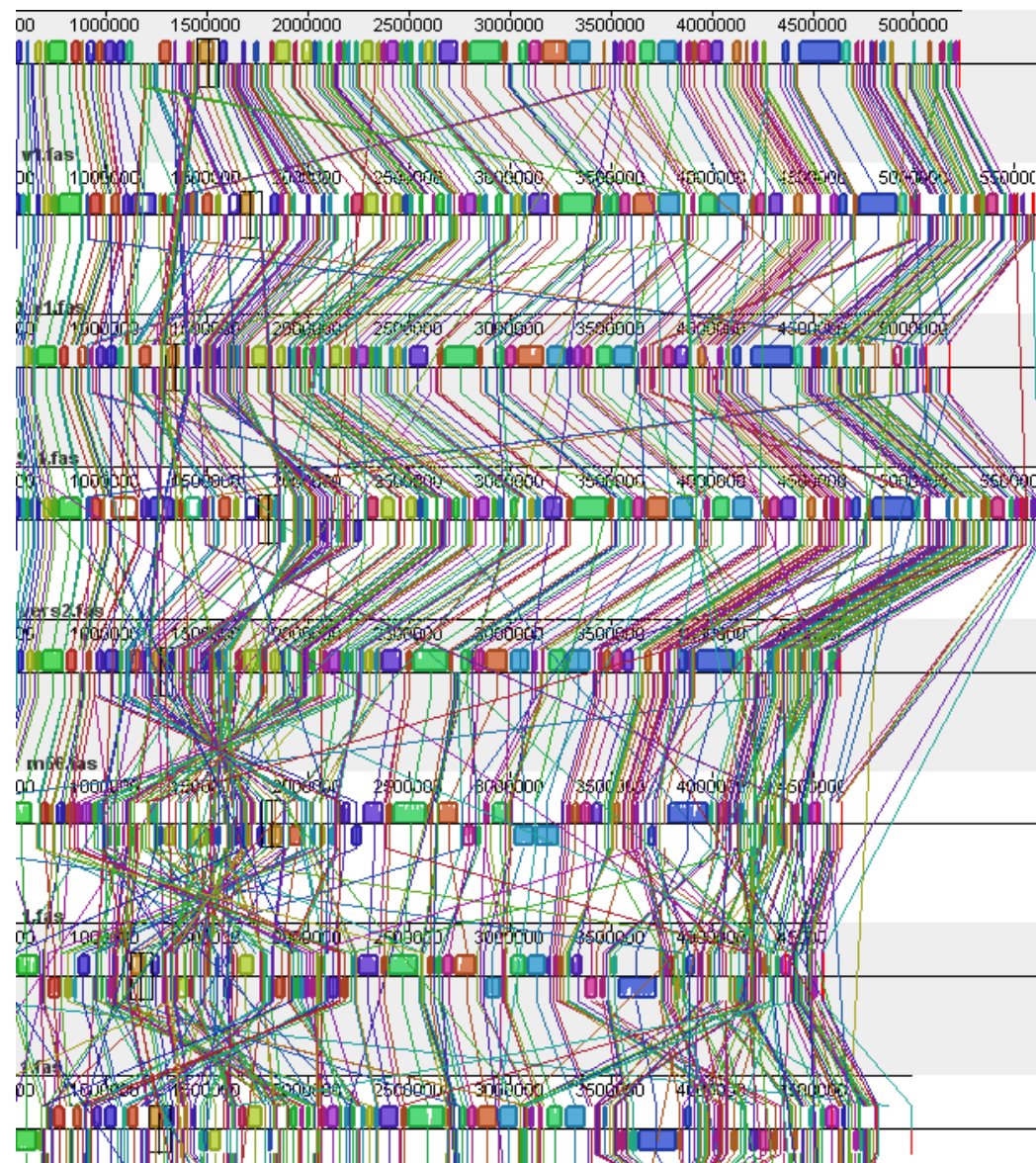
- Esta abundancia es abrumadora
- ¿Qué podemos hacer con ella?

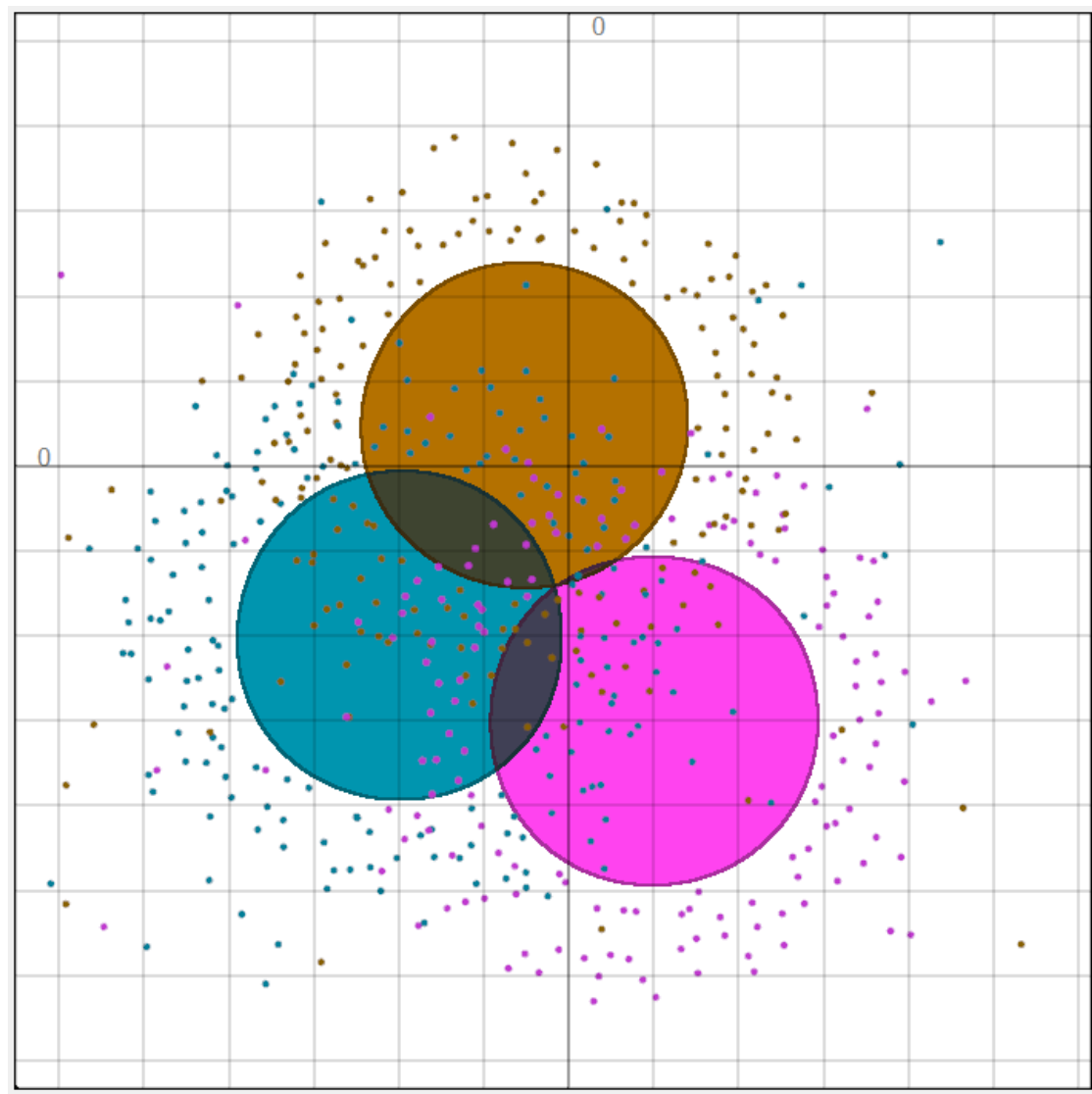
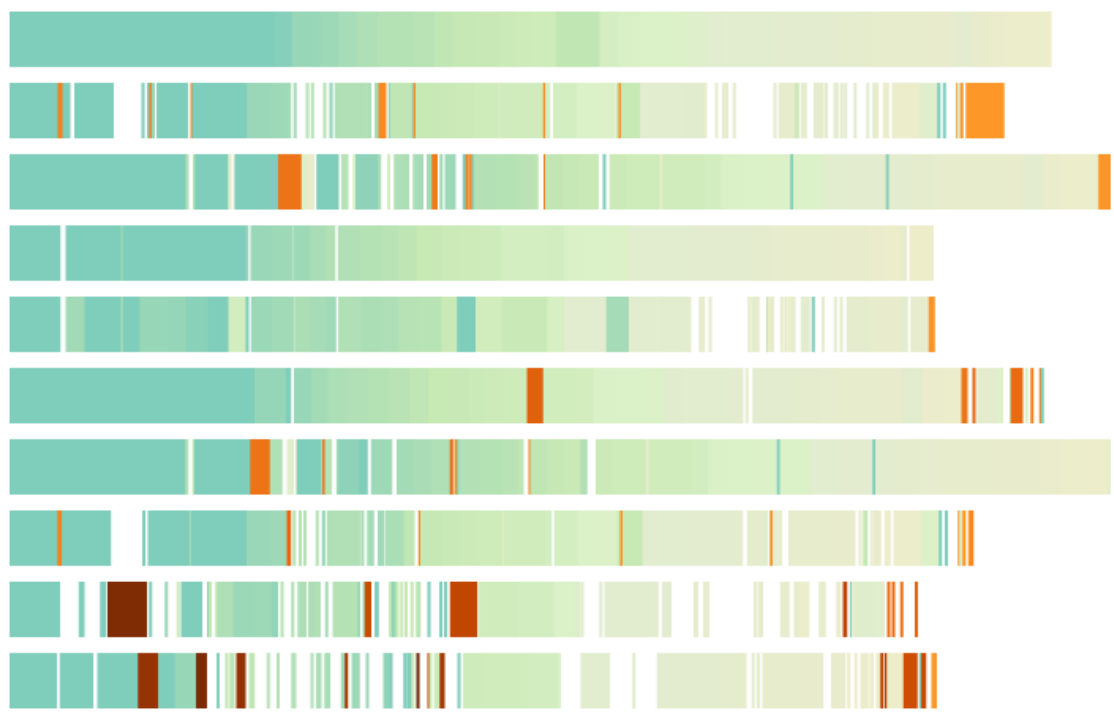
Visualizaciones

- Imágenes o gráficos para representar información
- Las visualizaciones tienen una finalidad
 - Reflexionar sobre los datos
 - Explorar los datos
 - Comunicar información
 - Permiten comprender los datos
- El objetivo es crear una representación eficaz y precisa de los datos

Visualizaciones

- No importa cómo se cree la visualización
- Una buena visualización no tiene por qué ser elaborada.
 - Pero sí es necesario tomar decisiones intencionadas
- No sirve de nada resolver muy bien el problema incorrecto





¿Qué hace que una visualización sea buena?

Buena

- Cumple eficazmente su **propósito**
- Facilita que el espectador vea lo que necesita ver

Mala

- Puede que no haga que las cosas sean fáciles de ver
- Puede hacer que se vea lo incorrecto
- Puede hacer que el espectador vea algo que no existe

4 Pasos para Crear Visualizaciones

1. Tarea
2. Datos
3. Deseño
4. Detalles

4 Pasos para Crear Visualizaciones

1. **Tarea**
2. Datos
3. Deseño
4. Detalles

1. Tarea – ¿Qué intentamos hacer?

- Interpretar datos realmente grandes y complejos
- Aprender de colecciones de archivos y tablas
 - ¡Hay demasiados para que los leamos nosotros solos!
- Demostrar la presencia de los patrones que encontramos
- Son una combinación de cálculo y percepción

1. Tarea: Comparaciones

- A menudo, en salud pública, queremos hacer **comparaciones**
- ¿Existen diferencias?
- ¿Significan algo estas diferencias?
- ¿Están relacionadas?
- ¿Es una mejor que la otra?

1. Tarea: Comparaciones

- A menudo, en salud pública, queremos hacer comparaciones
- Las comparaciones son una tarea importante
- Implica múltiples cosas
- Merece especial atención
- Casi todos los análisis de datos pueden concebirse como una comparación

Preguntas para Visualizar una Comparación

Preguntas para Visualizar una Comparación

¿Cuál es la comparación?

- Objetivos
- Acciones

¿Por qué es difícil?

- Número de objetivos
- Objetivos grandes o complejos
- Relaciones complejas

Preguntas para Visualizar una Comparación

¿Cuál es la comparación?

- Objetivos
- Acciones

¿Por qué es difícil?

- Número de objetivos
- Objetivos grandes o complejos
- Relaciones complejas

¿Cuál es la comparación?

- ¿Cuáles son los elementos de la comparación?
- Comparación: **examinar** las semejanzas y diferencias de **dos o más cosas**
 - **Objetivos**: conjunto de cosas que se comparan
 - **Acción**: qué hacer con la relación entre ellas

Los Objetivos de la Comparación

- ¿Sabemos lo que queremos comparar?
- Explícito: tenemos un conjunto conocido de objetivos
- Implícito: es posible que no lo sepamos
 - Comparación con algún punto de referencia
 - Comparar con nuestras suposiciones/conocimientos
 - Comparar con cosas que conocemos

Las Acciones de la Comparación

¿Coincide con mis expectativas?

¿Qué grupos son similares?

¿Qué grupos son diferentes?

¿En qué se diferencian?

¿Sólo son diferentes en determinadas situaciones?

Las Acciones de la Comparación

¿Coincide con mis expectativas?

- Medir o cuantificar una relación

¿Qué grupos son similares?

- Identificar cosas similares

¿Qué grupos son diferentes?

- Medir o cuantificar una relación

¿En qué se diferencian?

- Identificar diferencias significativas

¿Sólo son diferentes en determinadas situaciones?

- Contextualizar las relaciones

Preguntas para Visualizar una Comparación

¿Cuál es la comparación?

- Objetivos
- Acciones

¿Por qué es difícil?

- Número de objetivos
- Objetivos grandes o complejos
- Relaciones complejas

¿Por qué es la comparación difícil?

- (no todas las comparaciones son difíciles)
- Demasiados objetivos para comparar
- Objetivos grandes o complejos
- Relaciones complejas
- Son retos de la escala!
- (Hay otros retos posibles)

4 Pasos para Crear Visualizaciones

1. Tarea
- 2. Datos**
3. Deseño
4. Detalles

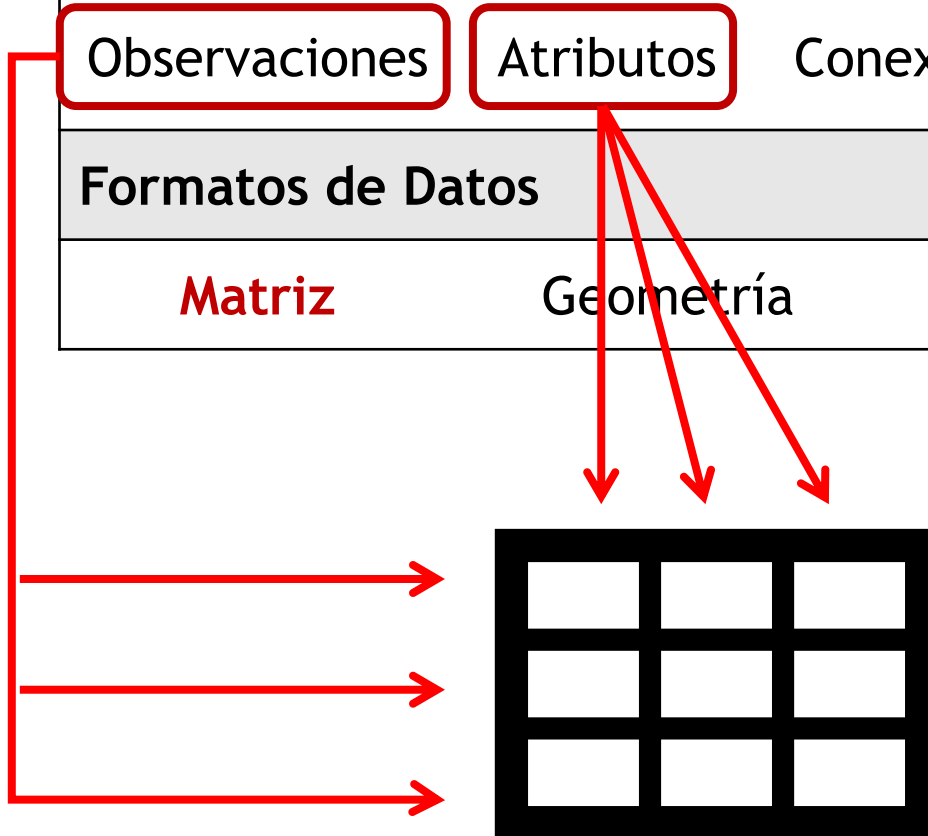
2. Datos

- A veces empezamos con datos, y vemos si hay algo interesante (elegimos una tarea)
- ¿De qué recursos se dispone para realizar la tarea?

Datos			
Componentes de Datos			
Observaciones	Atributos	Conexiones	Posiciones
Formatos de Datos			
Matriz	Geometría	Red	

Tipos de Atributos
• Categorías
• Pedido
• <i>Tipo de Orden</i>
• ordinal
• cuantitativo
• <i>Dirección de Orden</i>
• secuencial
• divergente
• cíclico

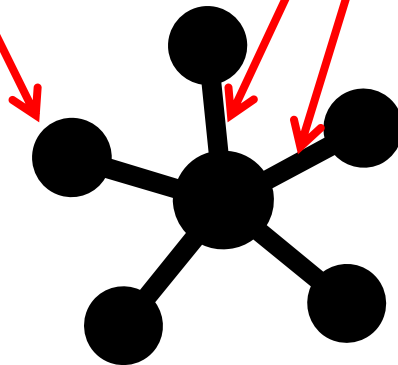
Datos			
Componentes de Datos			
Observaciones	Atributos	Conexiones	Posiciones
Formatos de Datos			
Matriz	Geometría	Red	



Datos			
Componentes de Datos			
Observaciones	Atributos	Conexiones	Posiciones
Formatos de Datos			
Matriz	Geometría	Red	



Datos			
Componentes de Datos			
Observaciones	Atributos	Conexiones	Posiciones
Formatos de Datos			
Matriz	Geometría	Red	



4 Pasos para Crear Visualizaciones

1. Tarea
2. Datos
3. Diseño
4. Detalles

3. Deseño

- ¿Cuál es la estrategia para convertir los datos en algo visual?
- Deseño: acto de tomar decisiones intencionadas en la creación de algo
- Dependiendo del tipo de datos y la tarea, podemos elegir distintos tipos de gráficos

3. Deseño

<https://r-graph-gallery.com/>



3. Diseño

Distribución



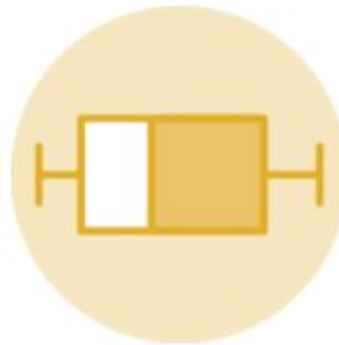
Violin



Density



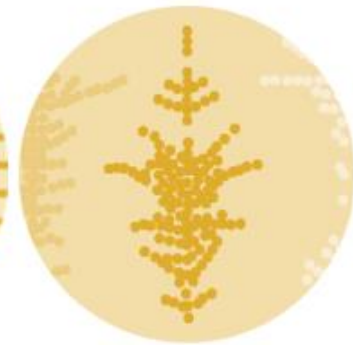
Histogram



Boxplot



Ridgeline



Beeswarm

3. Diseño

Correlación



Scatter



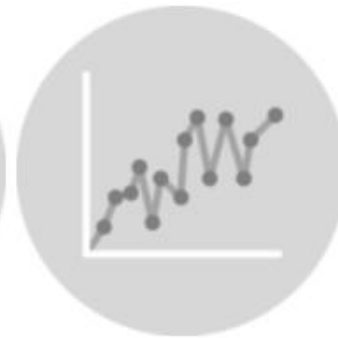
Heatmap



Correlogram



Bubble



Connected scatter



Density 2d

3. Diseño

Clasificación



Barplot



Spider / Radar



Wordcloud



Parallel



Lollipop



Circular Barplot



Table

3. Diseño

Partes de un todo



Grouped and
Stacked barplot



Treemap



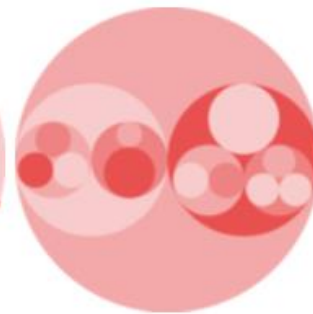
Doughnut



Pie chart



Dendrogram



Circular packing

3. Diseño

Evolución



Line plot



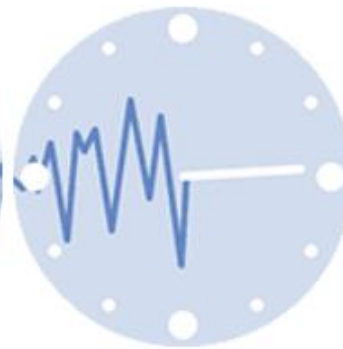
Area



Stacked area



Streamchart



Time Series

3. Deseño

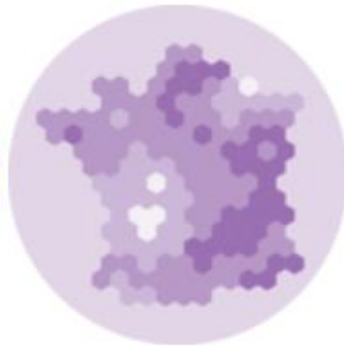
Mapa



Map



Choropleth



Hexbin map



Cartogram



Connection



Bubble map

3. Deseño

Flujo



Chord diagram



Network



Sankey



Arc diagram

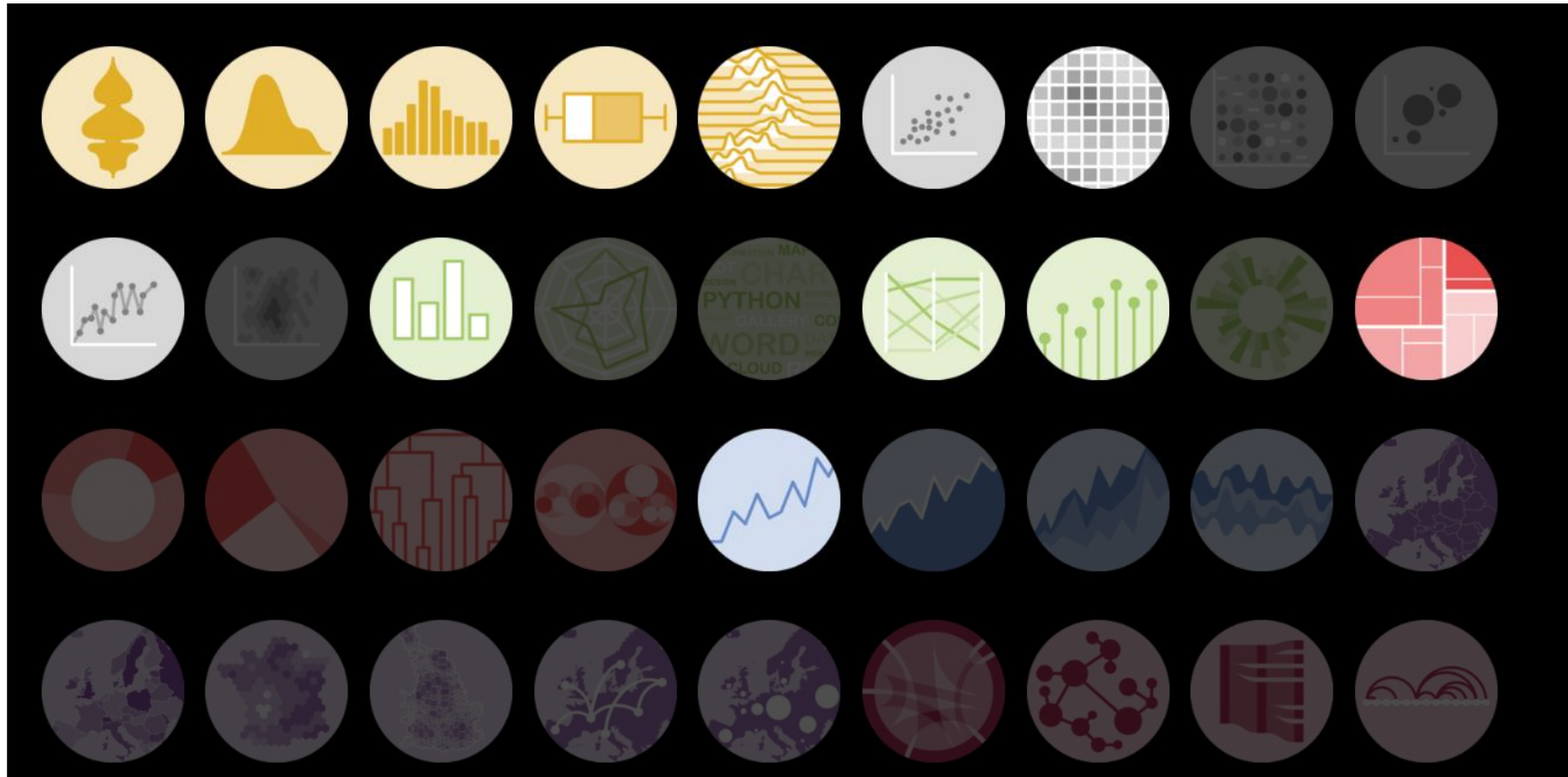


Edge bundling

3. Diseño



3. Diseño



3. Diseño

- ¿Cuál es la estrategia para convertir los datos en algo visual?
- Diseño: acto de tomar decisiones intencionadas en la creación de algo
- Dependiendo del tipo de datos y la tarea, podemos elegir distintos tipos de gráficos
- Es importante tener en cuenta cómo se representarán los datos en el gráfico
 - Forma
 - Color
 - Tamaño
 - Resaltar
 - Etiquetas

4 Pasos para Crear Visualizaciones

1. Tarea
2. Datos
3. Deseño
4. Detalles

4. Detalles

- ¿Cuál es la imagen o el gráfico concreto?
- ¿Cuáles son las opciones?
 - Colores específicas

- [ColorBrewer: Color Advice for Maps \(colorbrewer2.org\)](http://colorbrewer2.org)

ARAB SPRING



ARAB SPRING



Práctica

Aplicación de Insecticidas en El Oro

- Durante 3 años, 3 ciudades de El Oro utilizaron insecticidas para combatir los mosquitos
- Había 4 tipos de insecticidas
- Hubo 4 métodos de aplicación
- La fumigación se realizó en casas o barrios

Aplicación de Insecticidas en El Oro

- Durante 3 años, 3 ciudades de El Oro utilizaron insecticidas para combatir los mosquitos
- Había 4 tipos de insecticidas
- Hubo 4 métodos de aplicación
- La fumigación se realizó en casas o barrios

Práctica

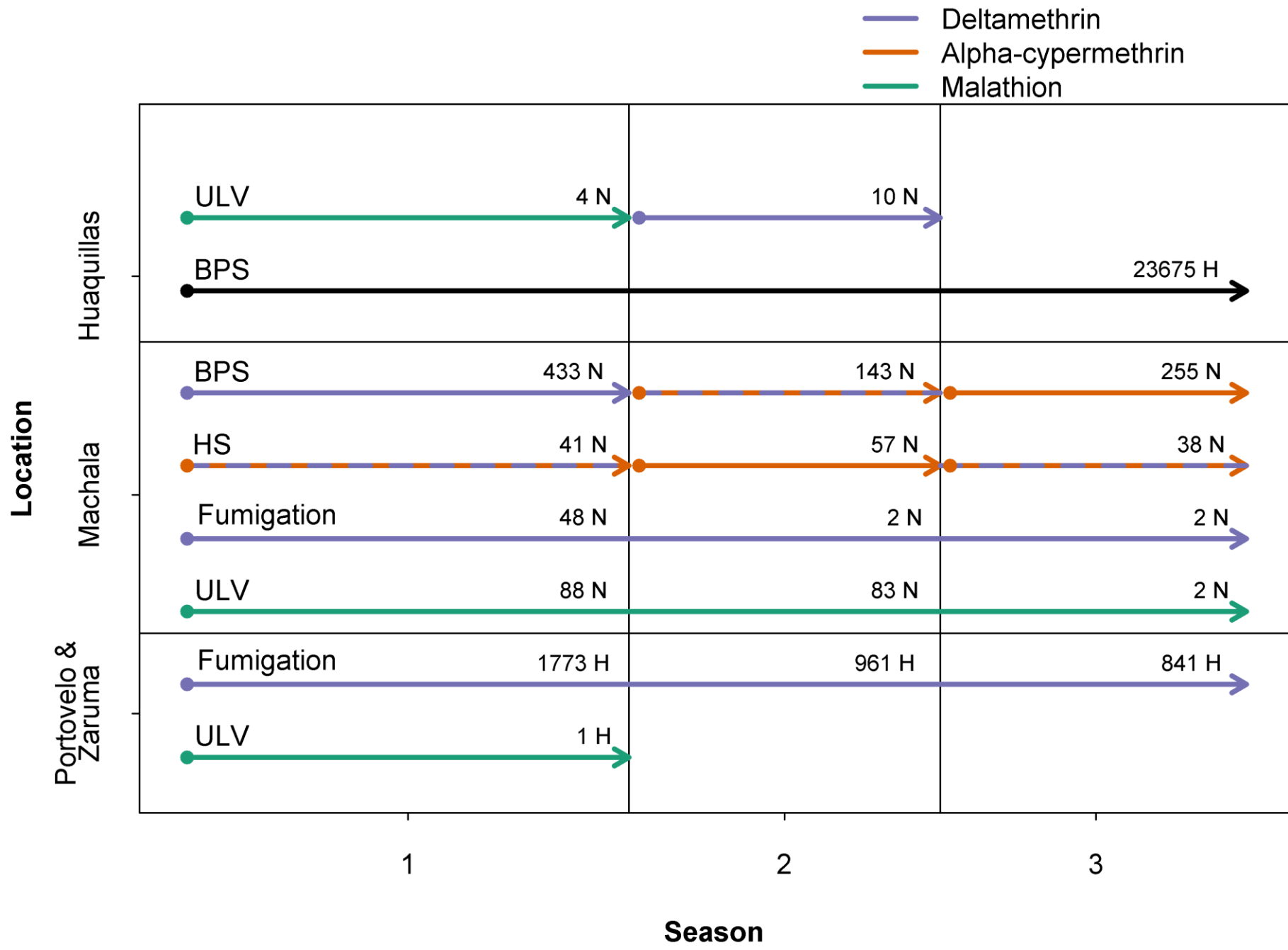
- Definir una tarea de comparación
- Cree un diseño para completar su tarea
- Refinar los detalles

Datos

- **Años:** 2016, 2017, 2018
- **Ciudades:** Huaquillas, Machala, Portovelo
- **Insecticidas:** deltamethin, cypermethrin, malathion, desconocido
- **Métodos de aplicación:** fumigación, por mano, por mochila, volumen bajo
- **Ubicaciones:** casas o barrios

Práctica

- Definir una tarea de comparación
- Cree un diseño para completar su tarea
- Refinar los detalles





Descanso

Discurso Final

