

USING SECONDARY DATA

Data collection methods

Presenter : WAFA ABDULSAKHI

Training delivery: (September 2022)

IDEAMAP
SUDAN
(IDeAMapSudan)

INTEGRATED DEPRIVATION AREA MAPPING SYSTEM FOR DISPLACEMENT DURABLE SOLUTIONS AND SOCIOECONOMIC RECONSTRUCTION IN KHARTOUM, SUDAN



(1)

ما هي البيانات الثانوية؟

**WHAT IS
SECONDARY
DATA?**

(2)

البيانات المؤسسية

**INSTITUTIONAL
DATA**

(3)

حشد البيانات

**CROWDSOURCE
D DATA**

(4)

استخدام البيانات الثانوية

**USING
SECONDARY
DATA**



(1)

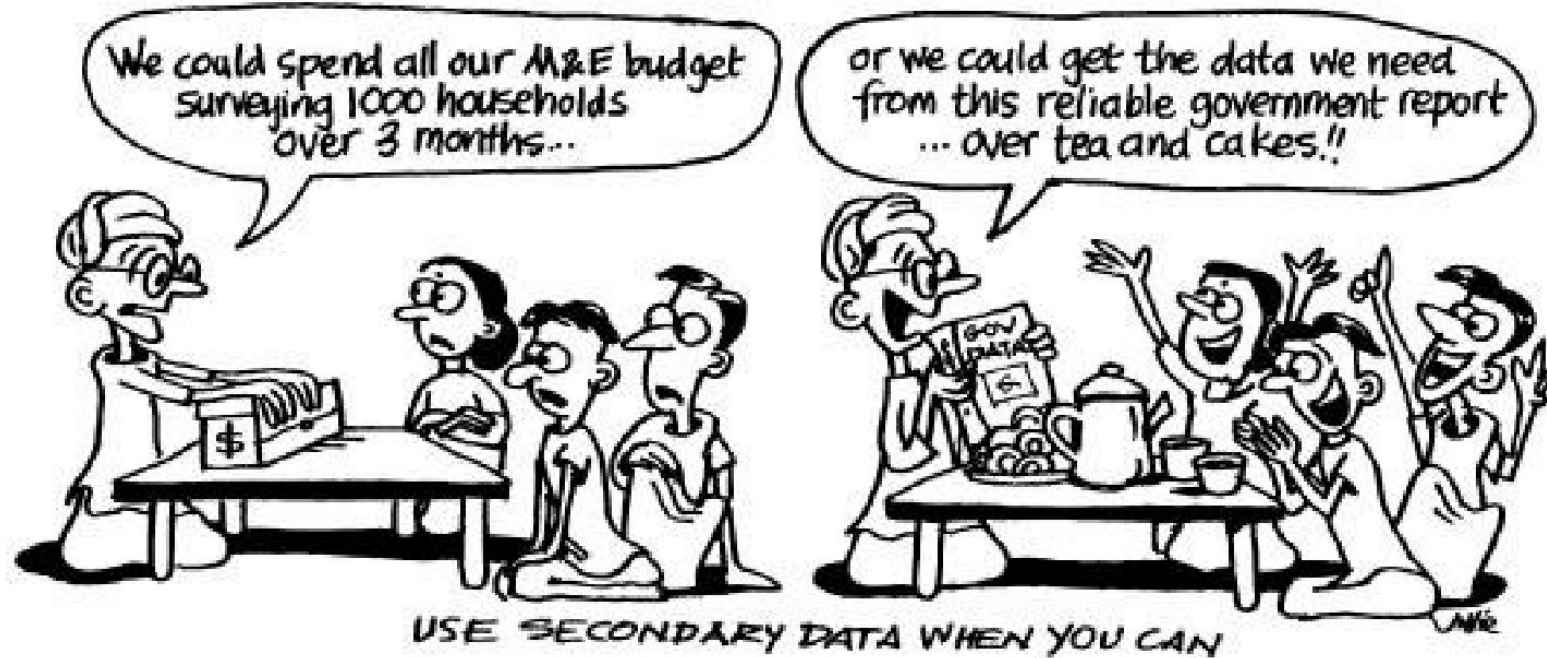
WHAT IS SECONDARY DATA?

- Using Secondary data
- Data Assessment
- Metadata

- استخدام البيانات الثانوية
- تقييم البيانات
- البيانات الوصفية

WHAT IS SECONDARY DATA

ما هي البيانات الثانوية



ببساطة ، البيانات الثانوية هي البيانات التي لم يتم جمعها من قبل المستخدم

لا يُقصد بالبيانات الثانوية أن تحل محل البيانات الأولية. ومع ذلك ، يمكن أن يكون لها دور مهم في وضع البيانات الأولية في سياقها

Simply put, Secondary Data is data that **was not collected** by the user

- ✓ It can save you money, time and effort.

Secondary **data** is **NOT** meant to replace Primary Data

- ✓ It can, however, have an important role in contextualizing Primary Data

Prior to being used, Secondary data has to be assessed with respect to:

1. **Availability** (can I have it?)
2. **Relevancy** (Is this what I need?)
3. **Accuracy** (is the detail enough?)
4. **Scope** (does it concern my research topic?)

قبل استخدام البيانات ، يجب تقييم البيانات الثانوية فيما يتعلق بالآتي:

1. التوفر (هل يمكنني الحصول عليها؟)
2. الصلة (هل هذا ما أحتاجه؟)
3. الدقة (هل التفاصيل كافية؟)
4. النطاق (هل يتعلق بموضوع البحث الخاص بي؟)

Prior to being used, Secondary data has to be assessed with respect to:

قبل استخدام البيانات ، يجب تقييم البيانات الثانوية فيما يتعلق بالآتي:

1. **Availability** (can I have it?)

1. التوفر (هل يمكنني الحصول عليها؟)

2. **Relevancy** (Is this what I need?)

2. الصلة (هل هذا ما أحتاجه؟)

3. **Accuracy** (is the detail enough?)

3. الدقة (هل التفاصيل كافية؟)

4. **Scope** (does it concern my research topic?)

4. النطاق (هل يتعلق بموضوع البحث الخاص بي؟)

Sufficiency (to what extent does it meet my needs?)

الكفاية (إلى أي مدى يلبي احتياجاتي؟)

METADATA

البيانات الوصفية

Metadata is **'data about data'**, i.e. additional information about the dataset

البيانات الوصفية هي "بيانات تصف البيانات" ، أي معلومات إضافية حول مجموعة البيانات

METADATA



Descriptive

Information for identification and discovery



Structural

How the data is organized



Administrative

Technical source, data type, creation process



Reference

Content and quality of statistical data

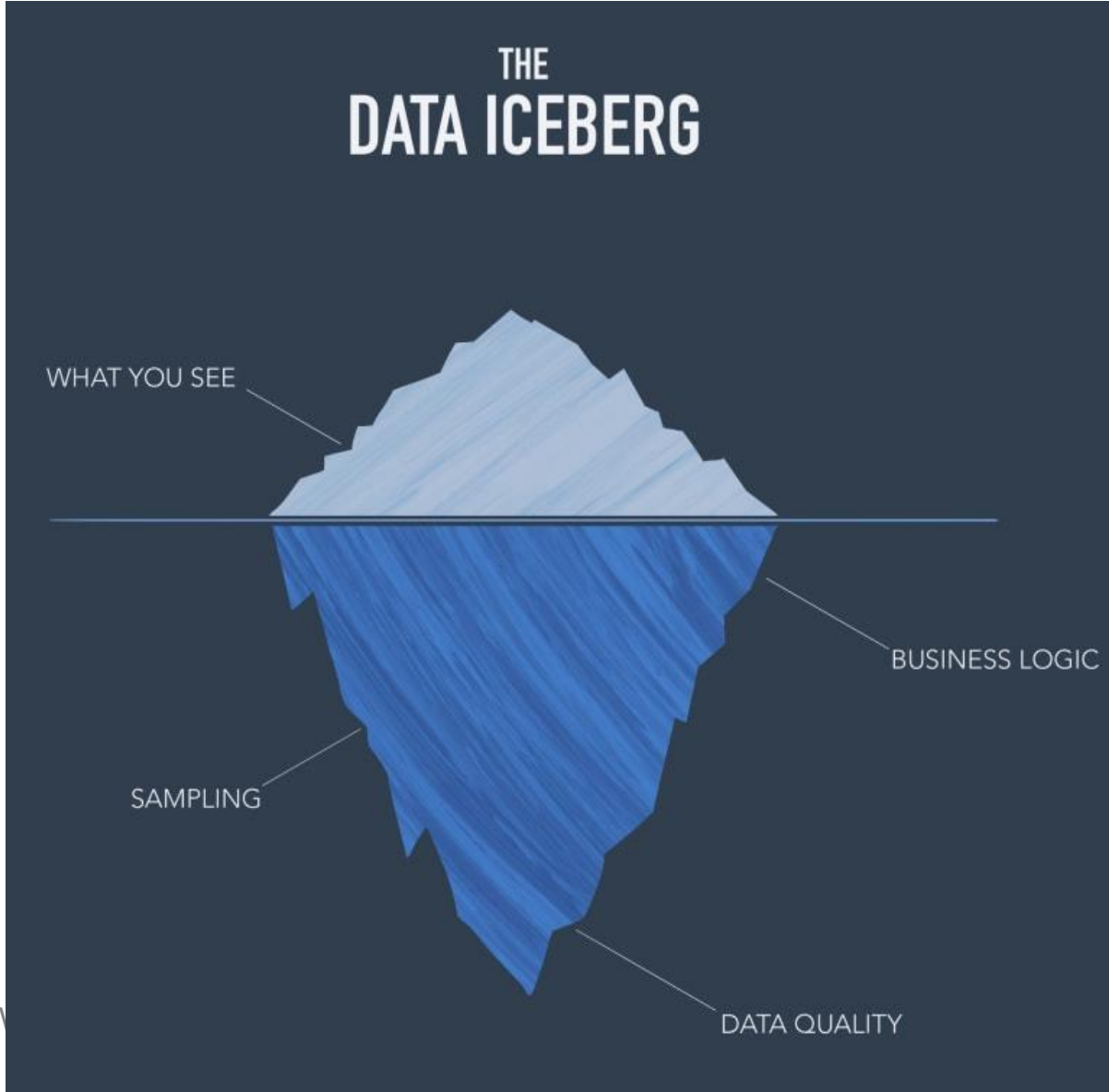


Statistical

Statistical information of the data: distribution, outliers, etc.

METADATA

البيانات الوصفية



تختلف طريقة تقديم البيانات الوصفية والمعلومات التي تقدمها. يمكن أن تشمل:

- معلومات عن مخرج البيانات
- نسب البيانات
- معلومات منهجية
- تعليمات الاستخدام

The way metadata is offered and the information it provides varies. It can include:

- Information about the producer
- Lineage of the data
- Methodological information
- Terms of use



(2)

INSTITUTIONAL DATA

- Web Services

• خدمات الويب

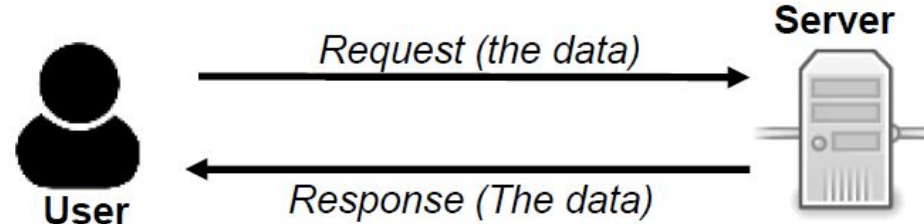
WEB SERVICES

OGC (Open Geospatial Consortium) web services are **open standards** aimed at ensuring **data interoperability**

OGC (Open Geospatial Consortium) web خدمات الويب هي معايير مفتوحة تهدف إلى ضمان إمكانية التشغيل البيني للبيانات

The most common are the **Web Map Service (WMS)** and the **Web Feature Service (WFS)**

- From a **practical point of view**, these are the essential differences;



الأكثر شيوعًا هي خدمة خريطة الويب (WMS))
وخدمة ميزات الويب (WFS)

- من وجهة نظر عملية ، هذه هي الاختلافات الأساسية:

	WMS	WFS
Data distributed as:	Static Image	Vector
Is the data Downloadable?	No	Yes
Typical use	Base map	Retrieving the data





(3)

CROWDSOURCED DATA

- Advantages
- Limitations
- Use Case: Open Street Map (OSM)
- Data of Unknown Origins

- مزایا
- سلبيات
- حالة الاستخدام: Open Street Map (OSM)
- بيانات من مصادر غير معروفة

CROWDSOURCED DATA

بيانات التعهيد الجماعي هي بيانات لا تعتمد على سلطة مركزية. فهي مصنوعة من مشاركة المتطوعين عادةً.

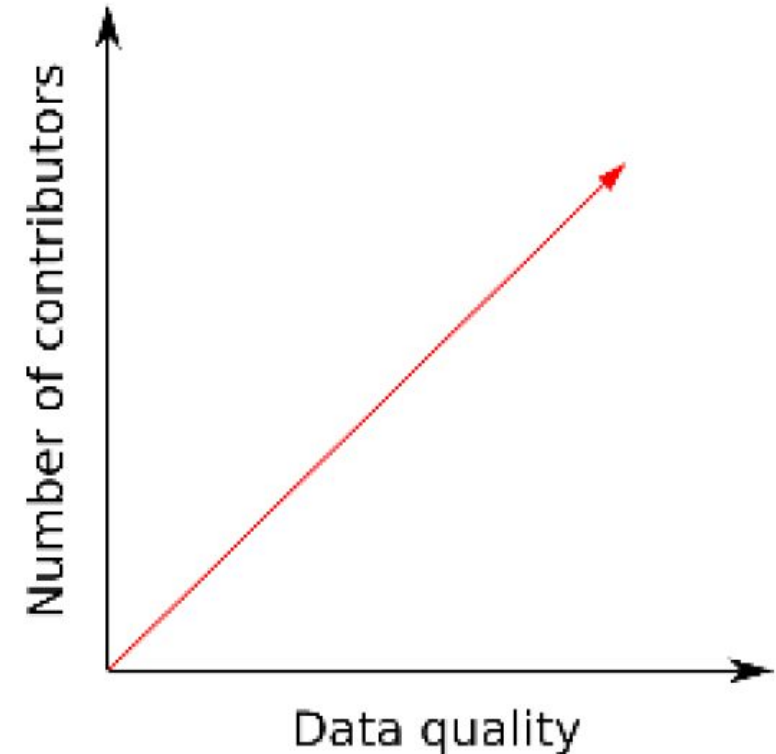
Crowdsourced data is data that does not depend on a central authority. Instead it is made from the contributions of, usually, volunteers.

بشكل عام ، تعتمد جودة هذه البيانات و جودة و مشاركة المجتمع المستخدم لها.

أفضل مثال في المجال الجغرافي هو مشروع بيانات Open Street Map (OSM).

Generally, crowdsourced data is as good as the community behind it is.

The best example in the geo domain is the **Open Street Map (OSM) data project**.



ADVANTAGES

المزايا



- سهل الاستخدام ، و قيود أقل.
- قد يكون أكثر
- اكتمالاً من المصادر التقليدية في المناطق ذات الأهمية الاقتصادية المنخفضة
- يمكن أن تكون أيضاً أكثر ثراءً من المصادر التقليدية حتى في المناطق "البارزة"
- تمكين المستخدم (أي شيء مفقود؟ إصلاحه أو إضافته بنفسك)
- يوفر طبقة معلومات تكميلية لمعرفة الأشياء / المعالم جيداً
- الفعالية من حيث التكلفة
- منظور جديد (Fresh perspective)

ADVANTAGES

Advantages

- Easy to use, few license restrictions
- It may be more complete than traditional sources in areas of low economic interest
- Empowers the user (i.e. something is missing? fix it or add it yourself)
- Provides a complementary information layer to well know objects/features
 - Cost-effectiveness





السلبيات

- من الصعب تقييم مستويات جودة البيانات
- المستخدم هو المسؤول الوحيد عن أي عواقب قد تنسب إلى استخدام البيانات
- بيانات وصفية ضعيفة
- الميل إلى عدم التجانس في التفاصيل والأسلوب

LIMITATION

Disadvantages

- It is hard to assess the levels of data quality
- The user is the sole responsible for any consequences that might be imputable to data usage
- Poor metadata
- Tendency for heterogeneity in detail and style



OPEN STREET MAP (OSM)

- Open Street Map is possibly the largest repository of geo-data in vector format.
- Founded in 2006, it is currently the largest crowdsourcing project in the geo-world
- (more than 5 million registered users as of November 2018)
- **Don't be a passive user! Contribute!**



- Open Street Map هو أكبر مستودع للبيانات الجغرافية بصيغة المتجهات vector format
- تأسس في عام 2006 ، وهو حاليًا أكبر مشروع **crowdsourcing project** في العالم الجغرافي
- (أكثر من 9 ملايين مستخدم مسجل اعتبارًا من نوفمبر 2022)
- لا تكن مستخدمًا غير فعال! ساهم!

DATA OF UNKNOWN ORIGINS

بيانات من أصول غير معروفة

Sometimes you have access, or you are given access, to data of which you are not sure what the origin is.

في بعض الأحيان قد تتحصل على بعض البيانات ، أو يتم منحك حق الوصول إلى بيانات لست متأكدًا من أصلها.

When this happens, all the procedures described so far apply plus a few extras:

- Try to validate the dataset using a second dataset whose lineage you know;
- Restrain from sharing or publishing those datasets because you don't know what license restrictions you may be violating.



في هذه الحالة،
تتطبق جميع الإجراءات التم توضيحها مسبقاً.
بالإضافة إلى بعض الإضافات:

- حاول التحقق من مجموعة البيانات باستخدام مجموعة بيانات ثانية تعرف نسبها ؛
- امتنع عن مشاركة هذه البيانات أو نشرها لأنك قد تتعدى على بعض قيود الترخيص التي لا تعرفها.



(4)

USING SECONDARY DATA

- Aggregation
- Spatializing data
- Filtering
- Acknowledging gaps

- تجميع
- البيانات المكانية
- الفلتره
- الاعتراف بالثغرات

USING SECONDARY DATA

استخدام البيانات الثانوية

The first rule when using secondary dataset is to keep in mind **the data assessment** and **metadata**

The second rule is to **use common sense** and have a critical look at the data



KEEP
CALM

AND

FOLLOW
THE RULES

✓ القاعدة الأولى عند استخدام مجموعة البيانات الثانوية هي مراعاة تقييم البيانات والبيانات الوصفية

✓ القاعدة الثانية هي استخدام الفطرة السليمة وإلقاء نظرة نقدية على البيانات

USING SECONDARY DATA

استخدام البيانات الثانوية

- In what units is the information being expressed? (i.g. is it percentage? Mileage? Square meters? Square km?)
 ● ماهي الوحدات الكمية المستخدمة لوصف المعلومات؟ (على سبيل المثال ، هل هي النسبة المئوية؟ الأميال؟ أمتار مربعة؟ كم مربع؟)
- Can you derive new data? (i.g. the dataset does not include aging index, but it does have the total number of inhabitants and the those who are 15 year old or younger)
 ● هل يمكنك استنباط بيانات جديدة؟ (على سبيل المثال ، لا تتضمن مجموعة البيانات مؤشر الشيخوخة ، ولكنها تحتوي على إجمالي عدد السكان وأولئك الذين تبلغ أعمارهم 15 عامًا أو أقل)
- Are there gaps in the data?
 ● هل توجد فجوات في البيانات؟

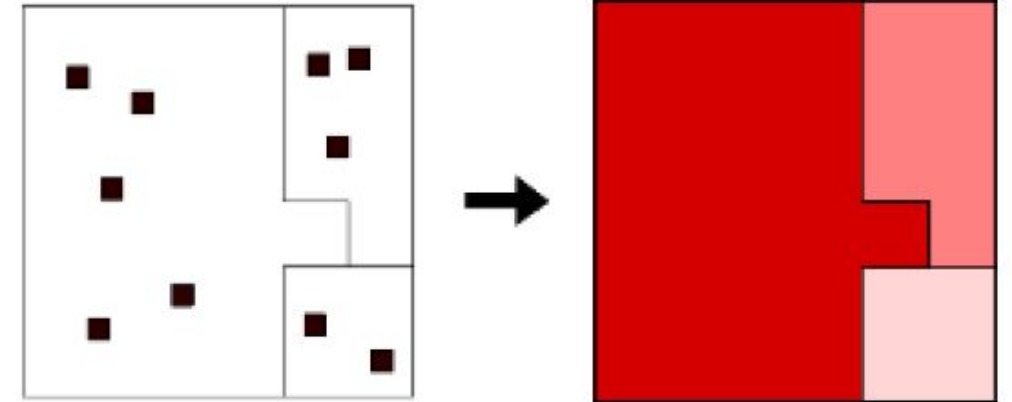


AGGREGATION

تجميع

- (Spatial) Aggregation is done at the expense of detail: the bigger the aggregation unit, the lesser the detail
- Sometimes the aggregation unit of secondary data is not the same as the unit you are using in our analysis. (i.g. you have data aggregated by district but you are working at the household level)

- تجميع البيانات المكانية يكون خصماً على التفاصيل، و الهدف منها تسهيل قراءة المعلومات و تحليلها
- في بعض الأحيان ، تختلف وحدة تجميع البيانات الثانوية عن الوحدة التي تستخدمها في تحليلك
- . (على سبيل المثال ، لديك بيانات مجمعة حسب المنطقة ولكنك تعمل على مستوى الأسرة)



□ Aggregation unit
■ Observation

■ 5 observations
■ 3 observations
■ 2 observations

- على الرغم من هذه المشكلات ، فإن تجميع البيانات يوفر عددًا من المزايا :
 - يقلل من تعقيد مجموعة البيانات - لا تعني التفاصيل الملاءمة للغرض المطروح ؛
 - يمكن أن يجعل اكتشاف الأنماط والاتجاهات أسهل ؛
 - يجعل إدارة البيانات وتحليلها أبسط.
- Despite these problems, aggregating data offers a number of **interesting advantages**:
 - Reduces the complexity of the dataset – detail does not mean fitness for the purpose at hand;
 - It can make the detection of patterns and trends easier;
 - Makes processing and data management simpler.

Completeness is not always there.
Managing data gaps is essential in research.

There is not one single approach to this problem, but some basic principles **should be kept in mind:**

- Be consistent, especially if the goal is to compare.
- Depending on the nature of the data, estimating missing values might be an option
- ALWAYS document what you did and acknowledge the consequences in a critical way.

البيانات الثانوية قد لا تكون مكتملة في أغلب الأوقات و عليه يجب عليك إدارة فجوات البيانات

لا يوجد نهج واحد لهذه المشكلة ، ولكن يجب أن يكون هناك بعض المبادئ الأساسية ضع في اعتبارك:

- كن متسقًا ، خاصة إذا كان الهدف هو المقارنة.
- اعتمادًا على طبيعة البيانات ، قد يكون تقدير القيم المفقودة خيارًا
- وثق دائمًا ما فعلته واعترف بالعواقب بطريقة حاسمة.

THANK YOU FOR YOUR ATTENTION!

تسهيل قوانين مشاركة المعلومات بين الجهات الحكومية يقلل من أعباء جمع البيانات

Facilitating information sharing laws between government agencies reduces the burden of data collection