What is "open data"?

Examples

Issues with Open Data

Open Data for Spatial Analysis

D G Rossiter

Cornell University, Soil & Crop Sciences Section Nanjing Normal University, Geographic Sciences Department

April 7, 2023

What is "open data"?

Example

Issues with Open Data

1 What is "open data"?

2 Examples



What is "open data"?

Example

Issues with Open Data

1 What is "open data"?

2 Examples

What is "open data"?

What is "open data"?

Open Data for

Spatial Analysis

Examples

- datasets that are **freely available** (almost always by direct download) by anyone . . .
- ... with **no restrictions** on their use or redistribution, other than **attribution** of the original source
 - Explanation at the Open Data Institute¹
- various license types protect the data but allow its use
 - e.g., CreativeCommons licenses²
- Linked to the FAIR concept (see next)

¹https://theodi.org

²https://creativecommons.org/about/cclicenses/

What is "open data"?

Examples

- Findabile, Accessibile, Interoperabile, and Reusabile
- Explicit FAIR principles³
- Open Data refers to **R**eusable, but FAIR does not require Open, just "clear and accessible data usage license"

³https://www.go-fair.org/fair-principles/

What is "open data"?

Examples

Issues with Open Data

• Improve the reliability of science

- many eyes on the same dataset
- errors will be more readily spotted
- data can be reused in comparative/larger studies
- Contribute to better problem-solving
 - Open Data Institute: "innovative use of open data to affect positive change across the globe"⁴

Why open data?

- Democratize science
 - especially for scientists with less access
- Internationalise science
 - workers in any country can get at any dataset
 - easy comparison of similar datasets from different countries
 - cross-border projects, data from different sources

⁴https://theodi.org/about-the-odi/

What is "open data"?

Examples

Issues with Open Data

- official data providers, as part of their mission
 - mapping agencies, census, weather bureaus, traffic ...
- cooperative platforms ("data sharing")
 - e.g, e-Bird⁵
- Individual researchers
 - data repositories, e.g. 4TU.ResearchData⁶
 - Many journals now suggest/require that the data on which the articles are based be deposited as Open Data

⁵https://ebird.org/home
⁶https://data.4tu.nl

What is "open data"?

Examples

Issues with Open Data

1 What is "open data"?

2 Examples

What is "open data"?

Examples

Issues with Open Data



| d ⊳ c | data.4tu.nl/search?q=SOIL G | | | | | | | |
|------------------|--|--|---|---|---|--|--|--|
| 4TU.ResearchData | BROWSE | Search on 4TU.ResearchDataQ. | | | | | | |
| | | | | | USE - | | | |
| | SOIL | | | × | Q | | | |
| | need help? | | | + FOLLOW THIS SEA | ясн | | | |
| | include figshare content | 73 results found | | | sort by: Relevance V | | | |
| | Content Type | DUARET | DATASET | DATASET | DATABLE | | | |
| | Select date V | Soil Hydraulic and | A 10-vear (2009-2019) | Data from: | Spatio-temporal | | | |
| | License CC0 (36) Ut General Terms of Use (21) | Thermal Properties for Land Surface Dataset posted on 24.04.2018 | surface soil moisture dataset produced Dataset posted on 08.01.2021 P. (Pell Zhang ~ | Quantitative visual soil observation for Dataset posted on 23.06.2019 M.W.J. (Marickel van Leeuwen ~ | subsurface soil moisture (40 cm Dataset posted on 05.02.2021 Coleen Carranza ~ | | | |
| | CC BY-54.0 (1) CC BY-54.4.0 (2) GPL-3.0+ (1) | | | | | | | |
| | Group 4TU.ResearchData (73) | DATASET | DATASET | DATASET | DATASET | | | |
| | Chter Institutions (30) Wageringen University and Research (18) University of Teente (14) Cold University of Teente (14) | Data for the article: "Root zone soil moisture estimation | Data from: Quantitative visual soil examination to | Data underlying the research of the Effects of Time-Dependent | Data underlying the research of Geochemical | | | |
| | ushow more | Lataset posted on 05.02.2021 Coleen Carranza V | M.W.J. (Maricke) van Leeuwen V | Dataset posted on 22.07.2020 Z. (Zhonglu) Wang ∨ | Dataset posted on 05.02.2020 S. (Susan) Klinkert V | | | |

Note license types

Spatio-temporal dataset from Dr. Colleen Carranza (Wageningen University)

Examples

Example - World Bank Open Data

| THE WORLD BANK Data Catalog | | | | | | 0. |
|--|---|--|--|--|---|--|
| Search | | Q | | ٨ | Data Collections | Getting Started |
| Global - Dams Database | | | | | | |
| DATA ACCESS AND LICENSING | | THEFT & CHART | in swape 🔟 | | | 🔇 Back |
| The dataset is clearified at Polici Unself-self-cess before the self-self-self- consider the Barks (an access the access the mid-self-self-self-self-self-self-self- adaptive the self-self-self-self-self-self- self-self-self-self-self-self-self-self- | Published (Ver.2 This database cor photovoltaic (PV) and India gathere Association (IHA) the FAO's public gathered from the Putpose of the re interest and obtain database will con continent of Africa | tatins parameters gatherr orgietts on existing hydre d from World Bank hydre Aquastar database for A souther for the seven in solaratiss.info/1 where the solaratiss.info/1 where the solarations Solar Adi in information about the parameter intue to expand to othe and India only, with upco | d after an assessment of the i power dams and other reserv nergystas infordiaset/indiaset/india- nergystas infordiaset/india- nd Dam Database (GRand) for y will be directly indied to this is is to enable potential device y will be directly indied to the regions until a global datase ming updates to include other | technological oirs, starting been expandi ydropower-d statven/datat r Africa.The d e solar resou lopers or stal floating solar it has been o continents. L | potential for develo with a plot in PTIS d to include the Int atabaset also includ ses/damSThis data atabase will be inclu- seholders to quickly PV project. Followin, completed.The curre fish- | pment of floating sola focused on West Afric emailonial Hydropowe set also contains dat used in the Global Sola solar module. Th anigate to the site c g the FY19 updates, th nt coverage spans th |
| The information on this page (the | Overview | Data & Resources | Additional Information | Citation | 5 | |
| these formats. | Type: | Other | | | | |
| | Languages Suppor | rted: English | | | | |
| | External Contact E | mail: rfox@worl | ibank.org | | | |
| | Topics: | Energy and | Extractives | | | |
| | Tags: | energydata | info | | | |
| | GP & CCSAs: | Energy & E | xtractives | | | |
| | Granularity: | Granularity | not specified | | | |
| | Geographical Covi | arage: India, Zam | oia, Angola, Tunisia, Tanzania, N | Aadagascar, E | gypt, Arab Rep. More | e |

"Free and open access to global development data"7

⁷https://data.worldbank.org/

Examples - Governments

What is "oper data"?

Examples

Issues with Open Data

- Many levels of government have published their maps and imagery
 - New York State⁸
 - Tompkins County⁹
 - United States of México, Instituto Nacional de Estadística, Geografía e Informática (INEGI)¹⁰

⁸https://data.gis.ny.gov/ ⁹https://www2.tompkinscountyny.gov/gis ¹⁰https://www.inegi.org.mx/

INEGI - Open Data themes

What is "open data"?

Examples

| d ⊳ C | □ | | | © 🦻 |
|-------|--|---|--|--|
| | NEGI | | | |
| | Iniaio Datos Convisios Transportegia | - Invertional én | | English Otros idiomas Contacto A |
| | Aviso COVID-19 Las medidas derivadas de la emergencia s programas de información de delCE: En ca o de otra indole en ellos, se darán a conoce publicación. | miterargadoon Initaria han afectado en distinta manera los so de exatil: impactos en calidad, cobertura r en las Notas Técnicas que acompañan su | Perspectiva en cifras COVID-19 | Visualizador analítico para el COVID-19 |
| | Producto Interno Bruto (PIB) Trimestral 3.3 Procentaje, 2020 47 | Desocupación 4.5 Porcentaje de la Poclatión Económican Activa, 2021 Ene | Defunciones generales 747,784 Defunciones 2019 | Total de establecimientos, Universo total 6,373,169 Establecimientos, 2019 |
| | . | Ter | nas | |
| | Agricultura, ganaderia y pesca | Catastro y Gestión Ter | ritorial Mag Comer | cio |
| | Comercio extenor | Construction | Educar | non |
| | Empleo y ocupacion | Empresas y establecir | mentos III Gobier | no |
| | Hogares y vivienda | Marse Contérior | or Manuta | Constantiation |
| | Mapas | Marco Geodesico | Marco | Geoestadistico |
| | Pablación | | C course | Central Nacionales |
| | Segurided públice u lucticie | Sociales on financier | Y Salud y | o de la información y comunicaciones |
| | Transporte | Turismo | | syras de la mormación y comunicaciones |

INEGI - Topographic map themes

What is "open data"?

HUNCCI

Examples

| iiii Incui | | English Otros idion | nas Contacto 🗚 |
|--------------------------------------|---|----------------------------|----------------------|
| Inicio Datos Servicios Transparencia | Investigación | Buscar | ٩.) |
| nicio / Temas / Topografia | | | |
| Mapas 🔳 | Topografía | | |
| Topografía | Información que tiene como objetivo representar la infraestructura, orografía, hidro; | grafía y las poblaciones (| del país. En ella se |
| Relieve Continental | registran fielmente todos estos elementos y las relaciones que guardan entre si. Co | ntiene información sobr | e el relieve, rasgos |
| Relieve Submarino | de los rasgos (toponimia) y las localidades. | 100 0011011000001, 0001 | |
| Territorio Insular | Nasa Deserves Desumentación Dublicaciones Hamericatas | | |
| Vias de Comunicación | | | |
| Hidrografia | El."Colima" o "Popocaté Q Selecciona la escala para la descarga de mapas | 1:50 000 | 1:20 000 |
| Climatología | | | |
| Edafología | | | _ |
| Fisiografia | | | |
| Geología | | | |
| Hidrología | | 1 | |
| Humedales Potenciales | | 3m | 5 |
| Uso de Suelo y Vegetación | 5 | 2 1 | |
| Uso Potencial de Suelo | Nogates | | \mathbf{Y} |

What is "open data"?

Examples

Issues with Open Data

INEGI - Downlod topographic map shapefiles



Notice only ESRI Shapefiles meet the "standard for [georeferenced] open data"

Topographic maps imported to QGIS

What is "open data"?

Examples

Issues with Open Data

| p | - | - | | | | | | | | |
|-----------------|---|--------------|----------|-----|--------|--------|-----|------|---|--|
| 0 | * | A | 3 | • | T | 80. | | * | 1 | |
| | | | | ac | ued | ucto5 | 0 | J | | |
| - | | 2 | | ar | ea_c | ultiv | 50 | a | | |
| 8 | | | ٠ | ba | n_m | at50 | p | | | |
| | | < | | bc | rdo | 50_I | | | | |
| 0 | | \checkmark | | ca | lle5 | 0_1 | | | | |
| 2 | | < | | ca | min | o50_l | | | | |
| • | | < | | ca | nalt | 0_1 | | | | |
| 8 | | | | ca | rret | era50 | 7 | | | |
| | | 2 | | ce | mer | terio | 50 | a | | |
| V _{O*} | | 2 | ٠ | ce | mer | terio | 50 | _p | | |
| 1 | | | | co | rrie | nte_a | g5 | 0_I | | |
| 0~ | | 2 | • | co | rrie | nte_d | e5 | 0_p | | |
| 3 | | | | cu | erp | o_agu | a5 | 0_a | | |
| ~ | | | | cu | rva_ | nivel | 50 | υ. | | |
| Va- | | 2 | • | de | pos | ito_de | 181 | 50_p | • | |
| a. | | | | ed | lifica | ciont | 50 | .a | | |
| C- | | | • | ed | lifica | cion | 50 | p | | |
| | | | | in | s_co | munio | cal | 50_1 | 2 | |
| | | | | in: | s_de | porti | v5 | 0_a | | |
| | | | <u>.</u> | in | s_de | portiv | vБ | 0_p | | |
| | | | | in | s_div | rersat | 50 | a | | |
| | | | | in: | s_ine | dustre | 50 | a | | |
| | | | | lin | nitet | 50_I | | | | |
| | | 2 | | lin | der | 50_I | | | | |
| | | | | lin | ea_e | om5 | 0_1 | | | |
| | | | _ | lin | ea_t | ranst | 50 | J | | |
| | | | | lo | calid | lad50 | La | | | |
| | | ~ | • | lo | calid | lad50 | LP | | | |
| | | | • | mi | ina5 | 0_p | | | | |
| | | | 1 | m | ojon | era50 |)_t | • | | |
| | | | | pι | ente | =50_I | | | | |
| | | | 1 | re | fere | ncia5 | 0_ | р | | |
| | | | | su | bes | ta_ele | :50 | 0_a | | |
| | | 2 | • | ta | nqu | s_agu | a5 | 0_p | | |
| | | | ٠ | ta | nqu | ≥50_p | 2 | | | |
| | | | _ | via | s_fer | rea5 | 0_1 | _ | | |
| | | | | | | | | | | |

00



Coordinate 507411,3463667 🕷 Scale 1:13524

What is "oper data"?

Examples

Issues with Open Data

| 9.1 | | Information f | rom provide | | | | |
|----------|-----------------------|-------------------------|---|------------------------------------|-----------------------|---|---------------|
| ⑦ | Information Source | Name Path Storage | curva_niveIS0_I /Useratrossiter/Down ESRI Shapefile | loads/702825278090 | sloonjunto | de datos/curva | nivel50_1.shp |
| * | Symbology | Encoding | windows-1252 | | | | |
| | -, | Geometry | Line (MultiLineStrin | g) | | | |
| e | Labels | CRS Extent | EPSC:102059 - NA 500000.00000000 | D_1983_2011_UTM 00000000,345730 | Zone_12M 8.7539679 | I – Projected 734036326 : 734036326 : | |
| ٩. | Diagrams | Unit | 531/38./23/63/9 meters | 89422834,346629 | 9.9173360 | 173590481 | |
| 9 | 3D View | Feature count | 1,016 | | | | |
| | Fields | Identification | | | | | |
| 0 | Attributes Form | Identifier | | | | | |
| | Joins | Title | r | | | | |
| a, | Auxiliary Storage | Type Language | dataset | | | | |
| ۲ | Actions | Categories | | | | | |
| 9 | Display | acy words | | | | | |
| | Rendering | Extent | | | | | |
| | Variables | CRS Spatial Extent | | | | | |
| 2 | Metadata | Temporal Exten | t | | | | |
| 4 | Dependencie | Access | | | | | |
| | Legend | Fees | | | | | |
| | QGIS Server | Licenses Rights | | | | | |
| | Digitizing | Constraints | | | | | |
| | | Fields | | | | | |
| | | Count 7 | | | | | |
| | | Field | Type | Length | | Precision | Comment |
| | | GEOGRAFICO | String | 14 | 0 | | |
| | | ALTUKA | String | 40 | 0 | | |
| | | CODICO | Soring | 13 | 0 | | |
| | | CALL REPR | String | 12 | 0 | | |
| | | CARTA | String | 15 | 0 | | |
| | | IDENTIFICA | String | 40 | 0 | | |
| | | DOM: NOR | soing | | U | | |
| | | Help Style | Apply | | | | Cancel |

Metadata for "50 m interval contour lines" coverage (shown in QGIS)

INEGI – Soil type maps

What is "open data"?

Examples

| INEGI | | | | | Eng | lish Otros idioma | as Contacto | ^0 | |
|--------------------------------------|---------------------------------------|---|-------------------|-----------------|-------------------|---|------------------|-------|--|
| inicio Datos Servicios Transparencia | Investigación | | | | Busc | ar | | | |
| icio / Temas / Edafología | | | | | | | | | |
| Mapas 🗧 | Edafolog | ía | | | | | | Ð | |
| Topografia | | , mación geoespacial que muestra la di | stribución de los | principales tip | os de suelo en el | territorio nacional | . así como los | | |
| Relieve Continental | atrib | utos físicos, químicos y limitantes físicos | as y químicas pr | esentes, de ac | cuerdo con la bas | e referencial munc | lial del recurso | suelo | |
| Relieve Submarino | | inio (inio per cao cigao cirrigico). | ▼ Ver | más | | | | | |
| Territorio Insular | Mapa De | Mapa Descargas Documentación Herramientas | | | | | | | |
| Vias de Comunicación | Filtros | | | | | | | | |
| Hidrografía | Entidad: | Escala: | | Edición: | | | | | |
| Climatología | Sonora • 1:250 000 • 2007 • Consultar | | | | | | | | |
| Edafología | nogales 🗶 Q | | | | | | | | |
| Fisiografia | Cualquier palabra O Frase completa | | | | | | | - 1 | |
| Geología | C Estándar de datas ablartas | | | | | | | | |
| Hidrología | Columbar of | | | | | | | | |
| Humedales Potenciales | Clave | Título | Escala | Serie | Edición | Tipo de archiv | 0 | | |
| Uso de Suelo y Vegetación | H12-2 | Conjunto de Datos Vectorial | 1:250 000 | 1 | 2007 | PDF + GIF | | | |
| Uso Potencial de Suelo | | Edafológico. Escala 1:250 000 Serie II Continuo Nacional Nogales | | | | 205.58 MB 10.36 TIFF SHF 3.74 MB 451.4 | MB 8 KB | | |

Soil types overlaid on topographic map themes

What is "open data"?

Examples

Issues with Open Data



Heroica Nogales, Sonora Note different design scales

What is "open data"?

Examples

Issues with Open Data

• What is "open data"?





What is "oper data"?

Examples

- Metadata: "data about the data"
- Separate lecture on metadata as part of "data quality" assessment
- Open Data can be freely accessed without interaction with the data provider, so the user must be able to know the data's origin, lineage, limitations ...

Minimal metadata example

What is "open data"?

Examples

Issues with Open Data CSV FTLES: The .csv files contain the soil moisture at 40 cm depth in an agricultural field in Haps, N. Brabant. The columns are the date and measurements per station installed in the field. Contents: Date ^^I^^I^^I - in YYYY-MM-DD HH:MM:SS format, in summer time GMT +1 Lora ## ^^I^^I_ Volumetric soil moisture per station Missing values - NA SHP FILES: Shapefiles for the location of the soil moisture sensors and the boundary covered by the sensors. The coordinate system is Amersfoort / RD New. Contents: Lora_bdy_A $\wedge I \wedge I \wedge I - polygon$ shapefile covering the stations for summer Lora_bdy_B ^^I^^I^^I_ polygon shapefile covering the station for winter 2 and summer 2018. Lora 2017summer RD ^^I^^I point locations of the stations for summer 201 Lora_2017winter_RD^^I^^I point locations of the sations for winter 2017 and summer 2018

Note: Location of 12 points were moved because a part of the field was turned to grass after 2017 harvest. Installation in grass was not desired because of grazing cattle.

What is "open data"?

Examples

Issues with Open Data These problems apply to any spatial data, but if it is Open there is no mechanism to control its use

1 It may be **difficult to interpret** by non-specialists.

• This may benefit those who can understand the information, i.e., already established actors, or actors who can hire specialists.

It may be misused

- e.g., presented at a larger scale than warranted by the design; being used for official purposes when the data source was unofficial
- It may allow "undesireable" actors to participate in areas where they were previously excluded
 - example: "land grabbing", i.e., purchase/leasing of the most productive lands by non-locals, guided by open soils data

| Open Data for Spatial Analysis | End |
|-----------------------------------|-----|
| What is "open data"? | |

Examples