



可持续发展大数据国际论坛
International Forum on Big Data for
Sustainable Development Goals

Curating Big Data on Population and Settlements to Monitor the Sustainable Development Goals (SDGs)

Robert R. Downs¹, Robert S. Chen¹, and Alexander M. de Sherbinin¹

¹Center for International Earth Science Information Network (CIESIN)

Columbia Climate School, Columbia University

International Forum on Big Data For Sustainable Development Goals (SDGs)

Session: Big Data Curation and Cloud Computing for SDGs

10:30 a.m. CST, 8 September 2021 (10:30 p.m. EDT, Sept. 7, 2021)



SDGs and Data on Population Distribution and Settlement Patterns



- Used to assess progress on the SDGs that focus on people
 - People are one of the five SDG Pillars
 - Critical to efforts to monitor SDG progress in disaggregated way as called for in 2030 Agenda
 - Supports the allocation of resources to help those most in need
 - Helps identify vulnerable and underserved groups that might otherwise be left behind
- Population data are needed for many SDG indicators
 - Indicator development (SDGs 1-5)
 - Establish baselines and measure progress (many SDGs)
 - Determine per capita rates (SDG 11, Sustainable Cities)
 - Identify urban versus rural areas (SDG 11, Sustainable Cities)
 - Assess risks of hazards to human settlements (SDGs 11, Sustainable Cities, and 13, Climate Action)
- Used alone or integrated with other types data
 - Earth observations (remote sensing data)
 - In situ data (e.g. citizen science and crowdsourced data)





UN-GGIM Global Fundamental Geospatial Data Themes



United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM)

- Adopted minimum list of 14 *global fundamental geospatial data themes* to facilitate measurement, monitoring, and management of sustainable development



Global Geodetic Reference Frame



Geographical Names



Addresses



Functional Areas



Buildings and Settlements



Land Parcels



Transport Networks



Elevation and Depth



Population Distribution



Land Cover and Land Use



Geology and Soils



Physical Infrastructure



Water



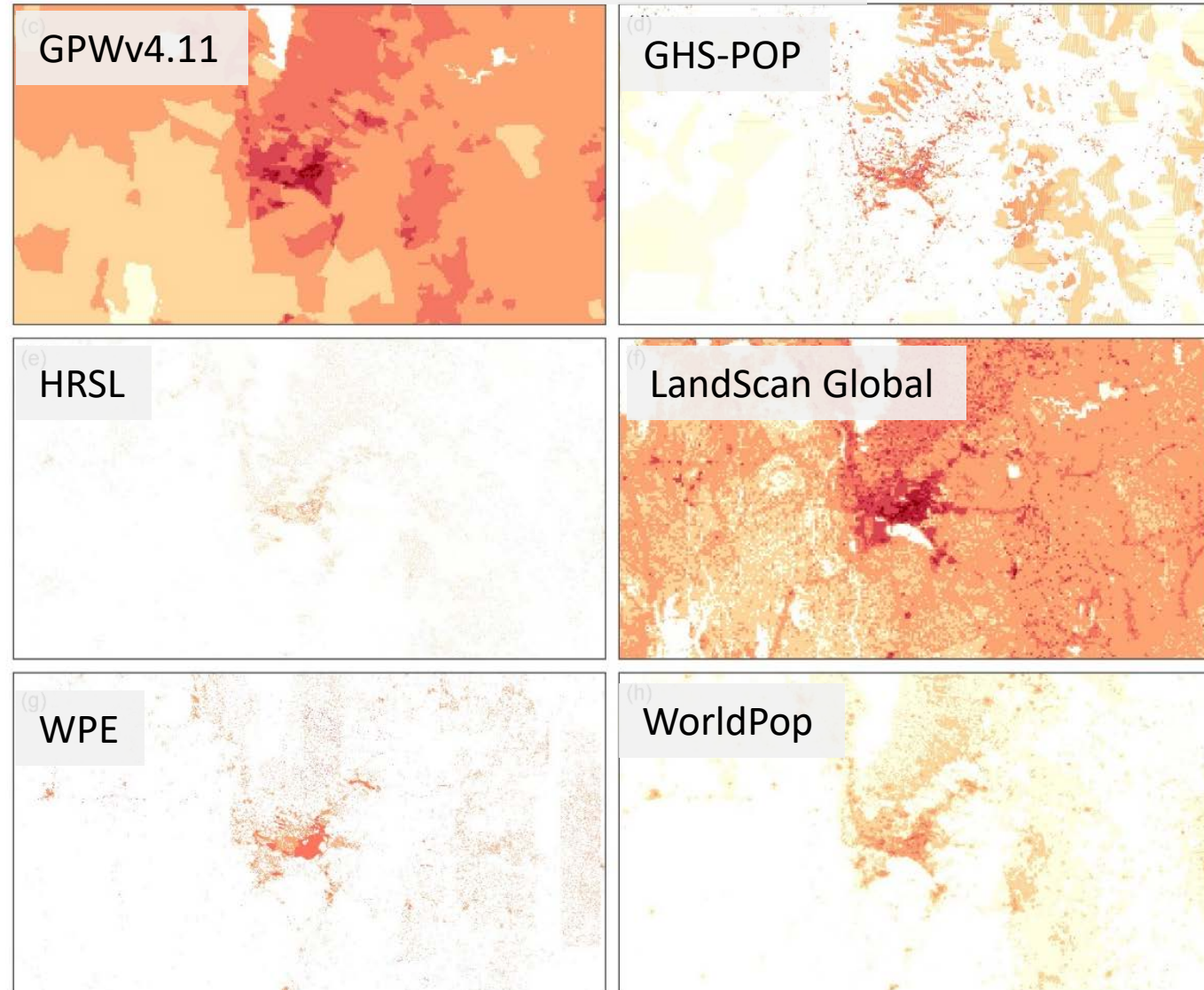
Orthoimagery

Many Different Gridded Population Data Sets Are Now Available



Greater Nairobi, Kenya

1. Gridded Population of the World (GPWv4.11)
2. Global Rural Urban Mapping Project (GRUMPv1)
3. LandScan Global Population Database (LandScan Global)
4. WorldPop
5. Global Human Settlement Layer - Population (GHS-POP)
6. World Population Estimate (WPE)
7. History Database of the Global Environment (HYDE) Population Grids v3.2
8. High Resolution Settlement Layer (HRSL)
9. European GHS Population Grid (GHS-POP-EUROSTAT)
10. Gridded Population Mapping (Demobase)



Number of Persons: (a) per unit; (c-h) per grid cell

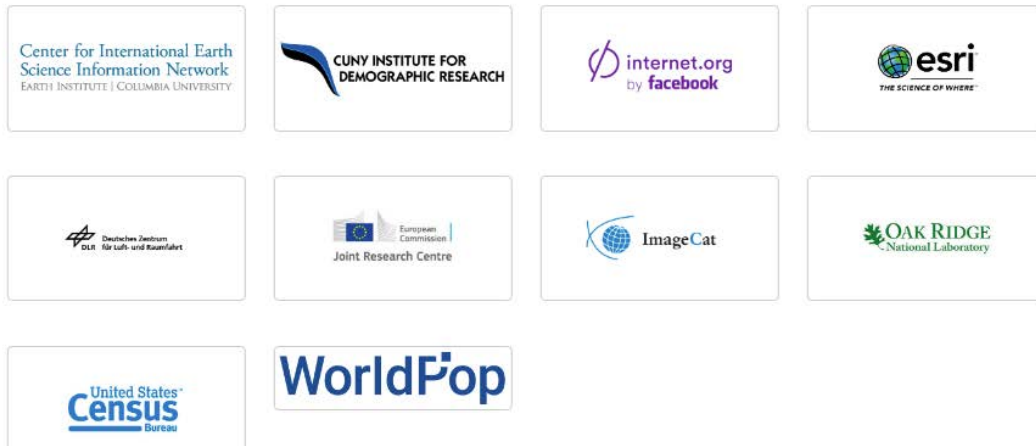
0-1 1-5 5-25 25-250 250-1000 1000-5000 5000-25000 >25000

POPGRID Data Collaborative and Web Site



Data Providers

At the core of the POPGRID Data Collaborative is a group of gridded population and settlement data product developers. By clicking on the logos of each of the groups below you will find links to their institutional home pages and lists of the data products they develop. In some cases there are also links to publications describing data and methodologies behind the products.



POPGRID Viewer

Get Population Estimates

Shapetile

Rectangle / Polygon

Population Data

GPWv4.11 UN Adjusted 2020 Count

Opacity

Legend: < 1, 1 - 5, 5 - 25, 25 - 250, 250 - 1,000, 1,000 - 5,000, 5,000 - 25,000, 25,000 +

Results

Population Estimates by Data Set and Year

POPGRID

Population Estimates

Source

Source	Population Estimates
GHSL 2015	1,168,070
GPWv4 2015	1,167,638
LandScan 2015	543,416
WorldPop 2015	1,175,809
HRSL 2015	1,165,745
ESRI WPE 2016	854,999
LandScan 2018	586,270
GPWv4 2020	1,270,014
WorldPop 2020	1,336,866

Population Estimates by Data Set and Year

Source	Population Estimates
GHSL 2015	1168070
GPWv4 2015	1167638
LandScan 2015	543416
WorldPop 2015	1175809
HRSL 2015	1165745
ESRI WPE 2016	854999
LandScan 2018	586270
GPWv4 2020	1270014
WorldPop 2020	1336866

Data Quality Message(s)

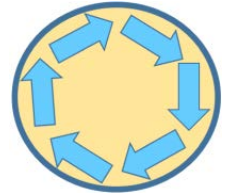
- The average national WPE reliability ranking is 2
- HRSL has coverage
- The average size of national input units in GPW is 1588 square kilometers

<https://www.popgrid.org/>

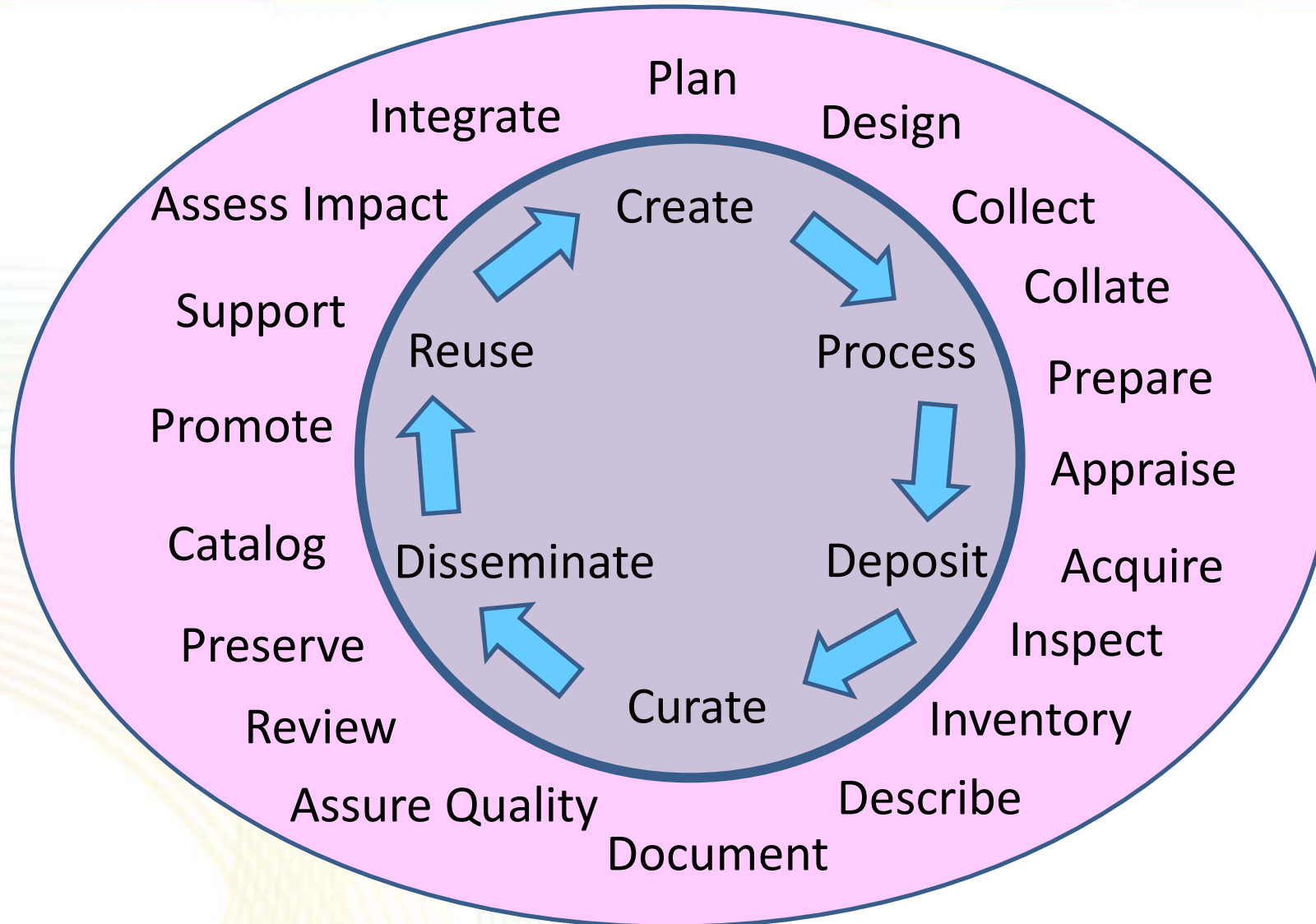
Need for Lifecycle Management and Curation of Open Data



- Using longitudinal data to monitor, observe trends, and make decisions
 - Need access to past, present, and future population and settlement data
 - Open data curated for interoperability facilitates integration with other data
 - Continuing access to preserved data through a persistently resolvable global identifier
- Diverse communities need to use data to assess progress on SDGs
 - Data must be discoverable across different disciplines, sectors, and countries
 - Documentation must foster understanding across various levels of expertise
 - Need to support today's users and future users
 - Help data producers to enable use of data beyond their own discipline



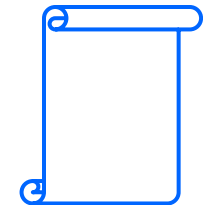
Data Lifecycle Management



Enabling Data Reuse Across Disciplines, Sectors and Countries



- Cross-disciplinary and Cross-sector representation on data product teams
 - Data collection
 - Data product development
 - Curation
 - Review
- Document data to facilitate understandability among potential users
 - Scientists
 - Planners and decision makers
 - Instructors and students
 - Journalists



Complementary Principles to Support Long-Term Transdisciplinary Use of Population and Settlement Data



Findable **A**ccessible **I**nteroperable **R**eusable

Collective Benefit **A**uthority to Control **R**esponsibility **E**thics



 **Global Earth Observation System of Systems (GEOSS)**
Group on Earth Observations

Data Sharing Principles

Data Management Principles

FAIR Principles for Scientific Data Management



- Findable



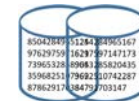
- Easily discoverable with a global persistent identifier, described in metadata that includes the identifier, and indexed in a searchable catalog

- Accessible



- Metadata are persistently accessible and retrievable from the identifier via a standard protocol

- Interoperable



- Metadata is understandable by user community, references other metadata, and is based on vocabularies

- Reusable



- Metadata contain a clear license and provenance information and utilize standards of the user community

CARE Principles for Indigenous Data Governance



- **Collective Benefit**
 - Inclusive indigenous community data development, governance, engagement, reuse, and outcomes
- **Authority to Control**
 - Indigenous community decision-making on data rights, access, governance, and stewardship
- **Responsibility**
 - Indigenous community capacity development, data capabilities, and use within language and culture
- **Ethics**
 - Indigenous community participation and assessment of risks, benefits, and future uses

Based on: Carroll et al. 2020. The CARE Principles for Indigenous Data Governance. Data Science Journal.
<http://doi.org/10.5334/dsj-2020-043>. Also, see: <https://www.gida-global.org/care>

TRUST Principles for Digital Repositories



– Transparency

- Terms of use for repository and holdings, expedient digital preservation; responsible stewardship



– Responsibility

- Meeting community standards, delivering data services, managing rights, sensitive data, and security



– User focus

- Displaying data metrics, populating catalogs for discovery, and meeting evolving community needs



– Sustainability

- Plan for continuity, risk and disaster management, and funding to provide long-term access to data



– Technology

- Implement data curation capabilities and plan for threat prevention, detection, and response



Based on: Lin, et al. The TRUST Principles for digital repositories. 2020. <https://doi.org/10.1038/s41597-020-0486-7>
Endorsed by many organizations: <https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories>



GEOSS Data Sharing Principles

GEOSS Data Sharing Principle 1.

Data, metadata and products will be shared as Open Data by default, by making them available as part of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE) without charge or restrictions on reuse, subject to the conditions of registration and attribution when the data are reused;



GEOSS Data Sharing Principle 2.

Where international instruments, national policies or legislation preclude the sharing of data as Open Data, data should be made available with minimal restrictions on use and at no more than the cost of reproduction and distribution;



GEOSS Data Sharing Principle 3.

All shared data, products and metadata will be made available with minimum time delay.



GEOSS Data Management Principles

Discoverability



DMP-1: Metadata for Discovery

Accessibility



DMP-2: Online Access

Usability



DMP-3: Data Encoding



DMP-4: Data Documentation



DMP-5: Data Traceability



DMP-6: Data Quality-Control

Preservation



DMP-7: Data Preservation



DMP-8: Data and Metadata Verification

Curation



DMP-9: Data Review and Reprocessing



DMP-10: Persistent and Resolvable Identifiers

Curation Issues for the Population and Settlement Data Community



- What data management and access standards should apply?
 - Scientific, development community, government, and/or international organizations?
- How to balance competing or conflicting principles and needs?
 - Open access, indigenous rights, privacy/confidentiality, national security, endangered species, people vulnerable to conflict or oppression, etc.
- Who is responsible for curation?
 - Data producers/collectives? National or institutional archives? International organizations and funders?
 - What involvement should users and other stakeholders have, e.g., indigenous peoples?
- How can accessibility and usability be improved for SDG applications?
 - Documentation, training, tool development, open services, validation/intercomparison, ??
 - Partnerships with key stakeholders and networks?



Thank you!

The Bill and Melinda Gates Foundation provides support to the
POPGRID Data Collaborative

<https://www.popgrid.org/>



BILL & MELINDA
GATES *foundation*

 COLUMBIA CLIMATE SCHOOL
CENTER FOR INTERNATIONAL EARTH SCIENCE
INFORMATION NETWORK

TRENDS