International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Data management for interdisciplinary research: Challenges and opportunities

Elena Rovenskaya

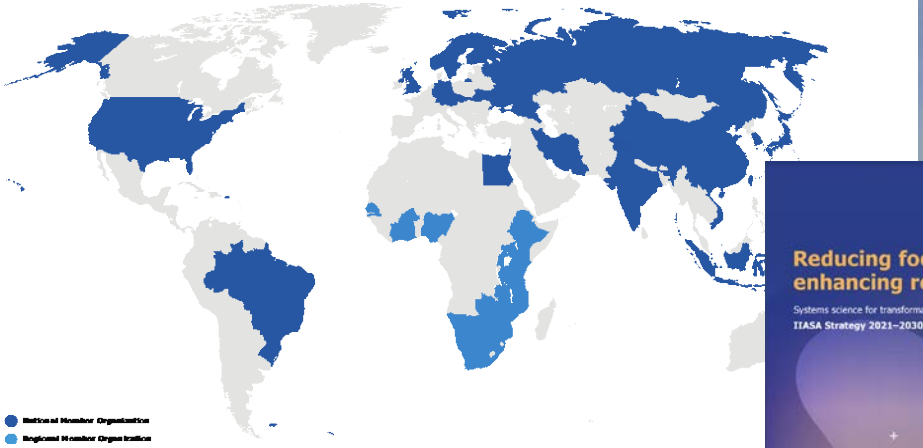
rovenska@iiasa.ac.at

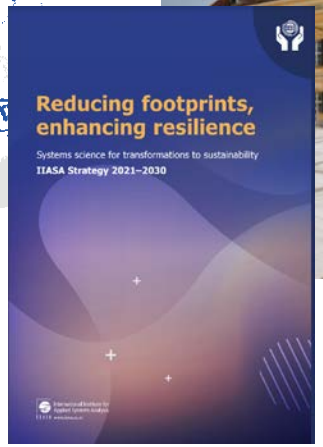
Program Director

Advancing Systems Analysis Program

International Institute for Applied Systems Analysis (IIASA)

Laxenburg, Austria



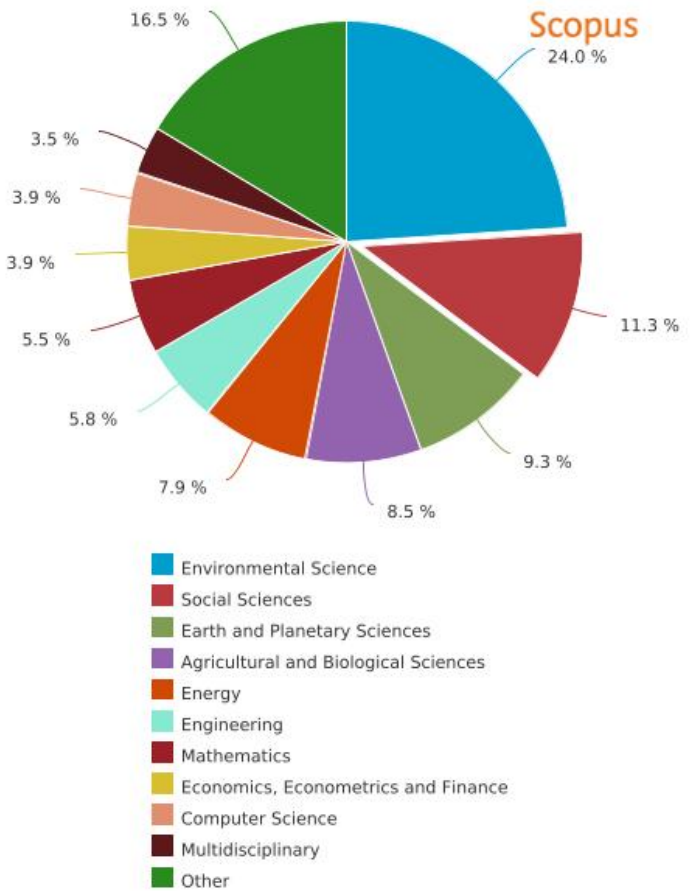


IIASA

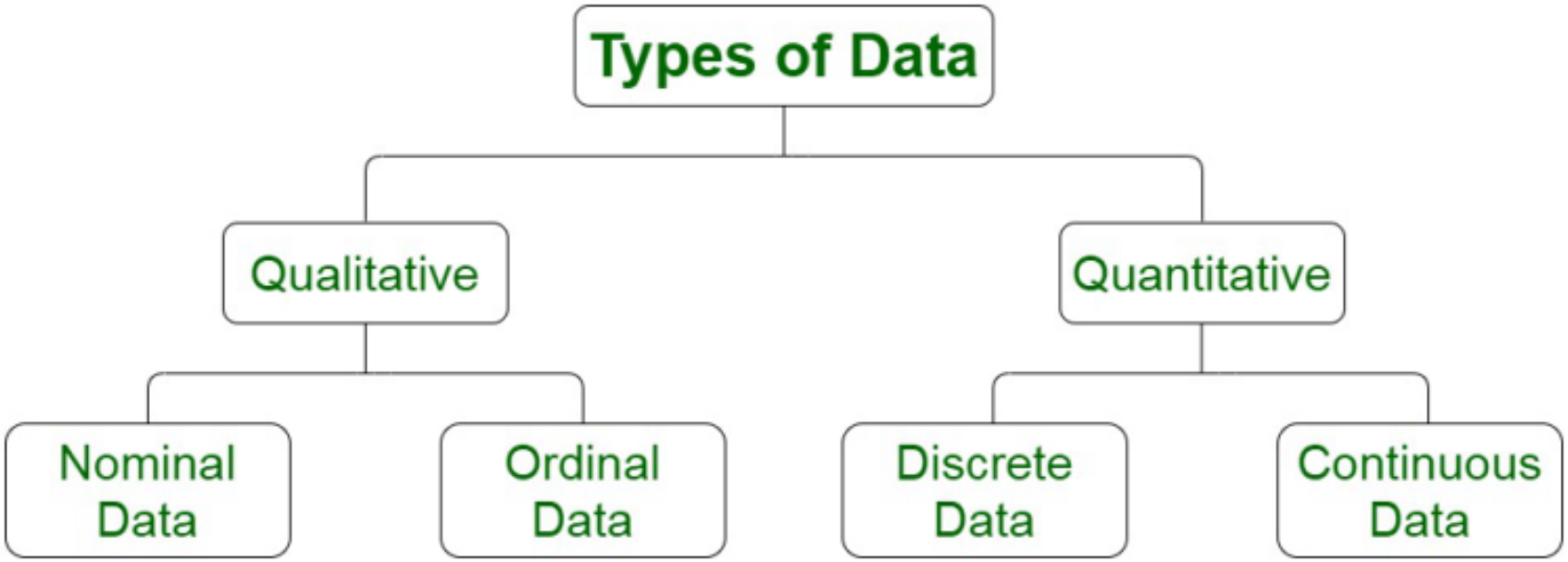
- Conducts policy-oriented high-impact research into global challenges arising from economic and technological development such as climate change or sustainable management of natural resources.
- Supported by a consortium of 21 national and regional member organizations representing diverse countries and regions (over 60% of the global population and almost 70% of the global economy).
- Established in 1972 to help bridge political divides between East and West via science (science diplomacy). Located in Austria, near Vienna.
- IIASA consistently serves as a knowledge partner for various UN agencies and national/regional policymaking bodies.

Diversity of disciplines

and interdisciplinary research areas and programs areas represented at IIASA



Diversity of research questions requires diversity of data



e.g., belongless to a political party

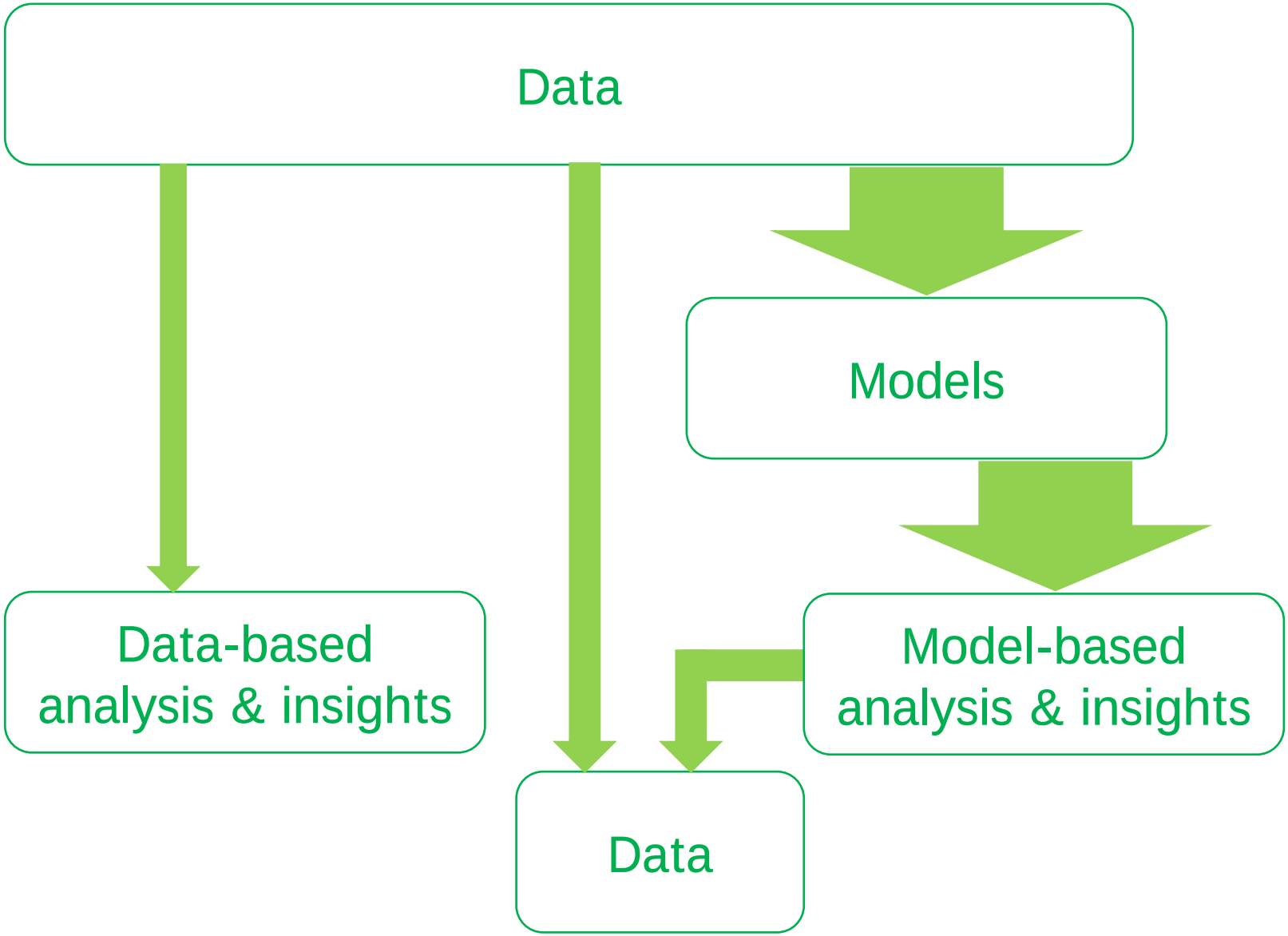
e.g., degree of agreement with a statement

e.g., number of protected areas (national parks etc.)

e.g., temperture, precipitation

Source: https://www.linkedin.com/pulse/kinds-data-vijay-s-snsrcas/

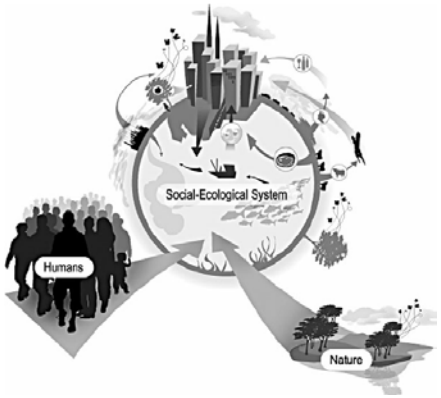
How data is used



Data requirements for socio-environmental modelling

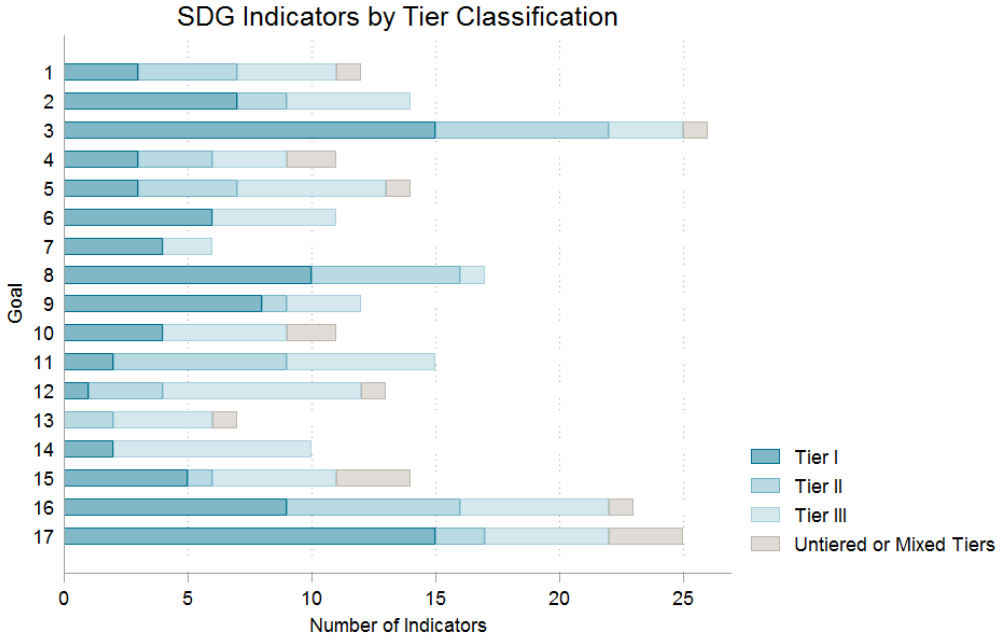
- Both global data and local, spatially detailed data are needed
- Especially pronounced data gaps exist in low-income countries
- Sometimes, data is made available with delays

Data requirements for socio-environmental modelling



- Multi-dimensional multi-disciplinary data are needed
- Soft factors (e.g., behavior) are particularly challenging in terms of data
- Some data is very expensive (e.g., microdata)

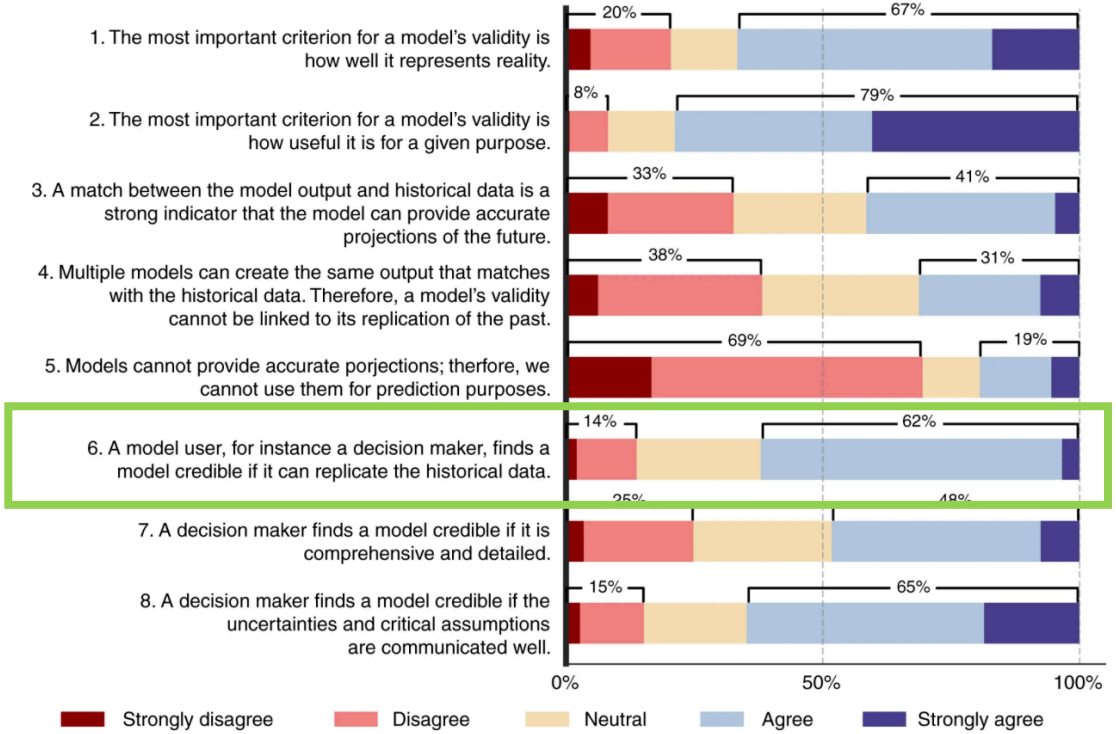
•Only 42% of indicators are Tier I, i.e., have an established methodology and regularly accessible data.
•Only 62% of Tier 1 indicators – or 25% of all indicators – could be found online in a publicly accessible format.



Source: https://www.cgdev.org/blog/sdg-indicators-serious-gaps-abound-data-availability

Data requirements for socio-environmental modelling

- Typically, data are required for model validation
- Often, the ability to reproduce the past is considered a key criterion for validation
- Having reliable data on the past is often a challenge



Survey responses to the key issues in model validation. The figure shows the responses given to the survey questions about the key issues in model validation, such as the validity criterion (Question 1 and 2), the role of historical data (Questions 3–5) and the decision-makers' view (Questions 6–8). The length of the bars refer to the fraction of responses given to each question on the Likert scale from Strongly disagree to Strongly agree. The majority of respondents consider both the representation of reality and usefulness as important validity criteria, support the usability of models for prediction purposes, and acknowledge the decision-makers' demand for transparency. Source data of this figure are provided in Supplementary Data 1

Source: Eker at al (2018)

IIASA's contribution to data provision

Novel Data Ecosystems for Sustainability (NODES) research group at IIASA works towards mobilizing the tools of citizen and data science combined with Earth observations to monitor, analyze, and foster progress towards the UN Sustainable Development Goals (SDGs).

- Generating new and novel datasets: land use, pollution, economic well-being, ...
- Advancing CS methods for data generation: quality, accuracy, ...

Table 1. Legend of the forest management layer

Raster value	Class
10	Naturally regenerating forest without any signs of human activities (e.g., primary forests)
20	Naturally regenerating forest with signs of human activities, e.g., logging, clear cuts, etc.
31	Planted forest
32	Short rotation plantations for timber
40	Oil palm plantations
53	Agroforestry

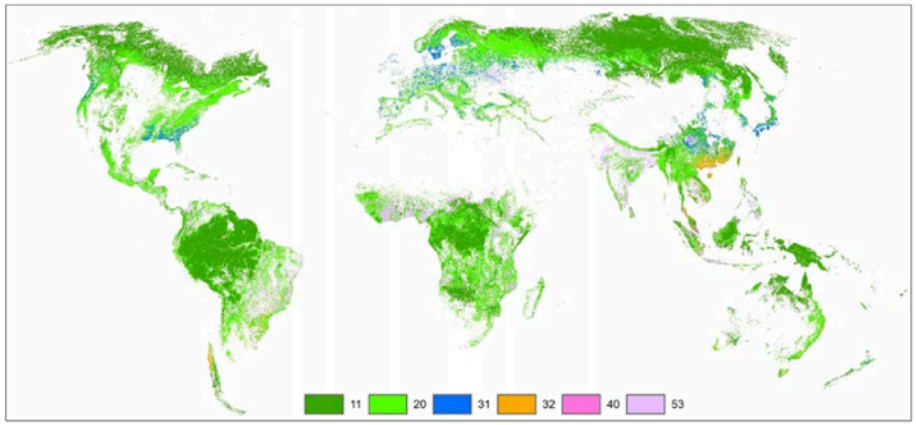
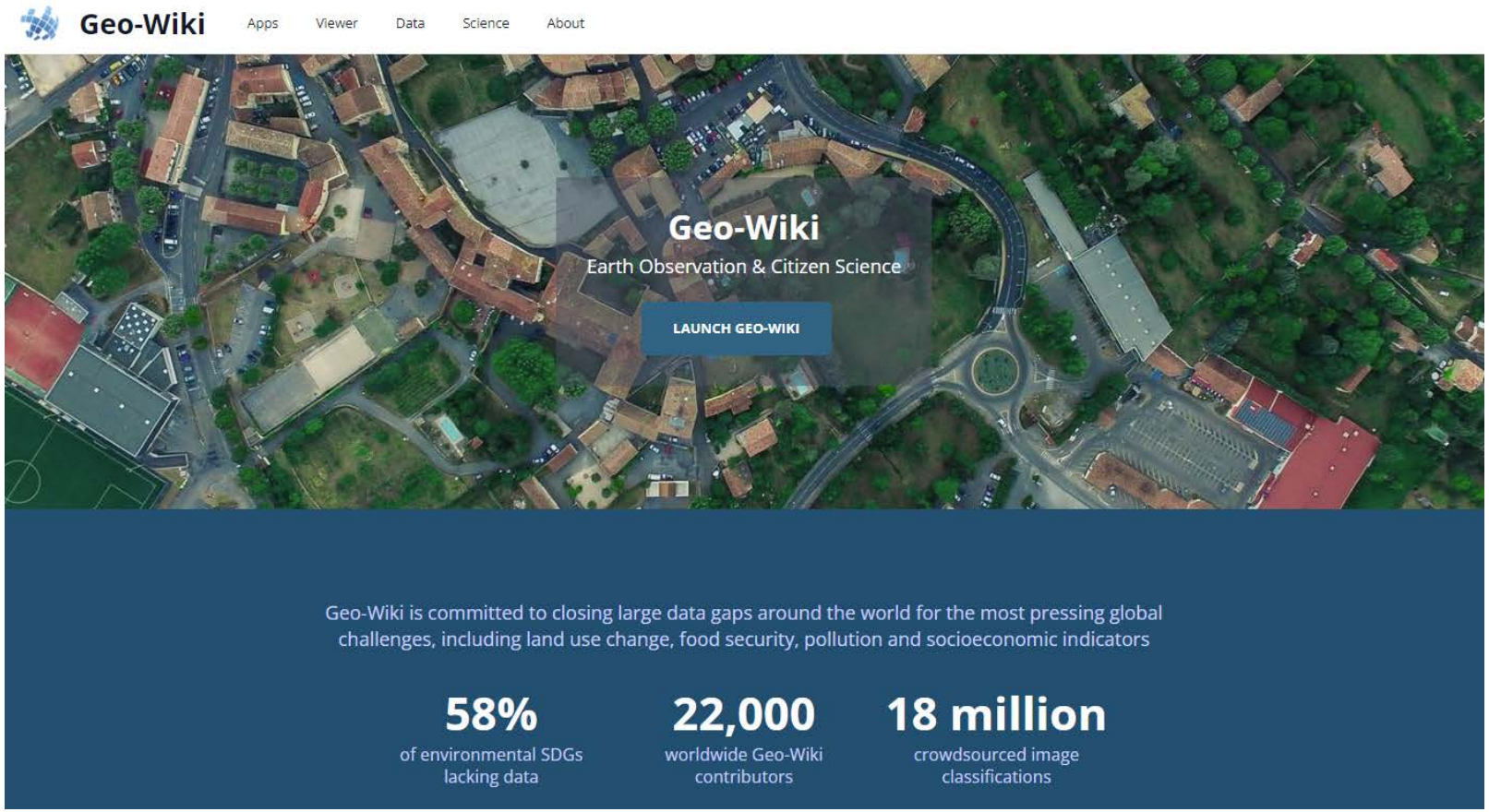


Figure 2. Global map of forest management (see the detailed legend in Table 1).

Source: Lesiv et al. (2020)

IIASA's contribution to data provision

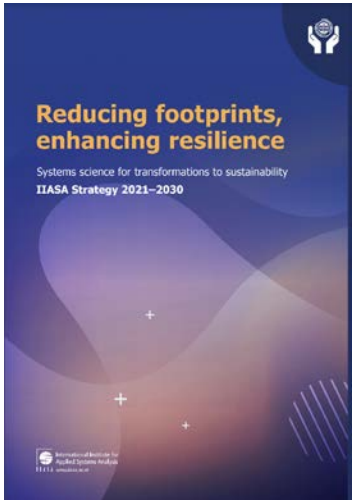


geo-wiki.org

Invitation to contribute!

Research data management at IIASA

Commitment to Open Science



Open science

IIASA produces world-class science that is regularly published in high-impact publications and made freely-accessible via the IIASA publications repository. IIASA scientists will be encouraged and supported to further advance open access by making IIASA research, data, models, and analytical tools openly available and accessible, in a responsible manner, to all.

IIASA's Open Access policy

As a signatory to the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, IIASA is committed to providing unrestricted and cost-free online access to its scientific publications and data for all users and researchers.

Ecosystem of IIASA policies supporting research data management

IIASA Strategy 2021-2030

- Aspires to adhere to open science

IIASA Research Plan 2021-2024

- Outlines major research areas which define the kinds of data to be used

Open Access policy

- Encourages IIASA researchers to make their research outputs, incl. data, open-access

Open Access to Models and Tools policy

- Views data as integral part of models
- Encourages IIASA researchers to follow FAIR principles (Findable, Accessible, Interoperable, Reusable)

Data Management procedure

- Presents key principles of data management at IIASA
- Establishes roles and responsibilities of different internal stakeholders in relation to data management

Data Protection rules

- Presents IIASA rules to comply with the GDPR, DSG, and FOG

Data stewards – Roles and responsibilities ToR

- Outlines key responsibilities of data stewards

Metrics for Performance and Impact procedure

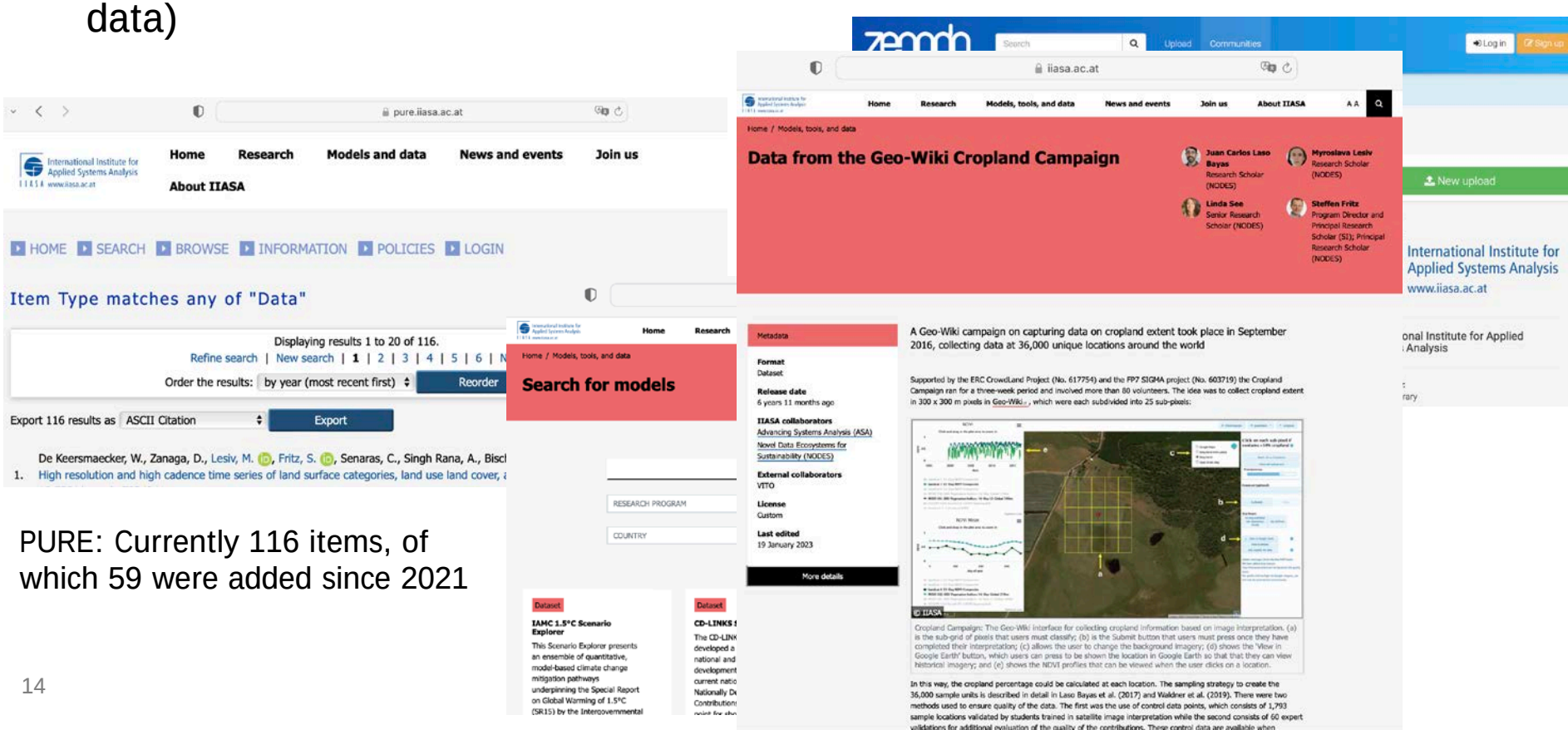
- Provides internal incentives to researchers, research groups and programs to make data open-access
- Includes the number of datasets made available and the number of downloads

Making IIASA datasets open-access

To adhere to the FAIR principles (Findable, Accessible, Interoperable, Reusable), IIASA researchers are encouraged to share data via internal or external platforms, including

- IIASA repository PURE
- IIASA's Zenodo community

- In addition, IIASA datasets are features at the IIASA website (incl. meta data)



PURE: Currently 116 items, of which 59 were added since 2021

Key internal stakeholders

Institute Leadership

- Create an environment that supports researchers in their efforts to make data open-access (incl. internal incentives)

Library and Knowledge Resources

- support the Leadership in development of relevant policies
- support researchers with data management plans
- support IIASA PURE and serve as the curator for meta data on the datasets produced by IIASA

Communications and External Relations

- provide featuring IIASA datasets over the IIASA website

Information and Communication Technology

- provide the necessary IT infrastructure

Researchers and Software Developers

- Produce research data
- Adhere to IIASA policies and relevant funders' requirements
- Innovate in making data serve scientific and other communities

Data Stewards

- Support researchers in practicing data management and sharing & facilitate making data FAIR
- One data steward in each IIASA program

Institute Data Coordinator

- Coordinates and supports the work of data stewards

Data stewards: Opportunities & Challenges

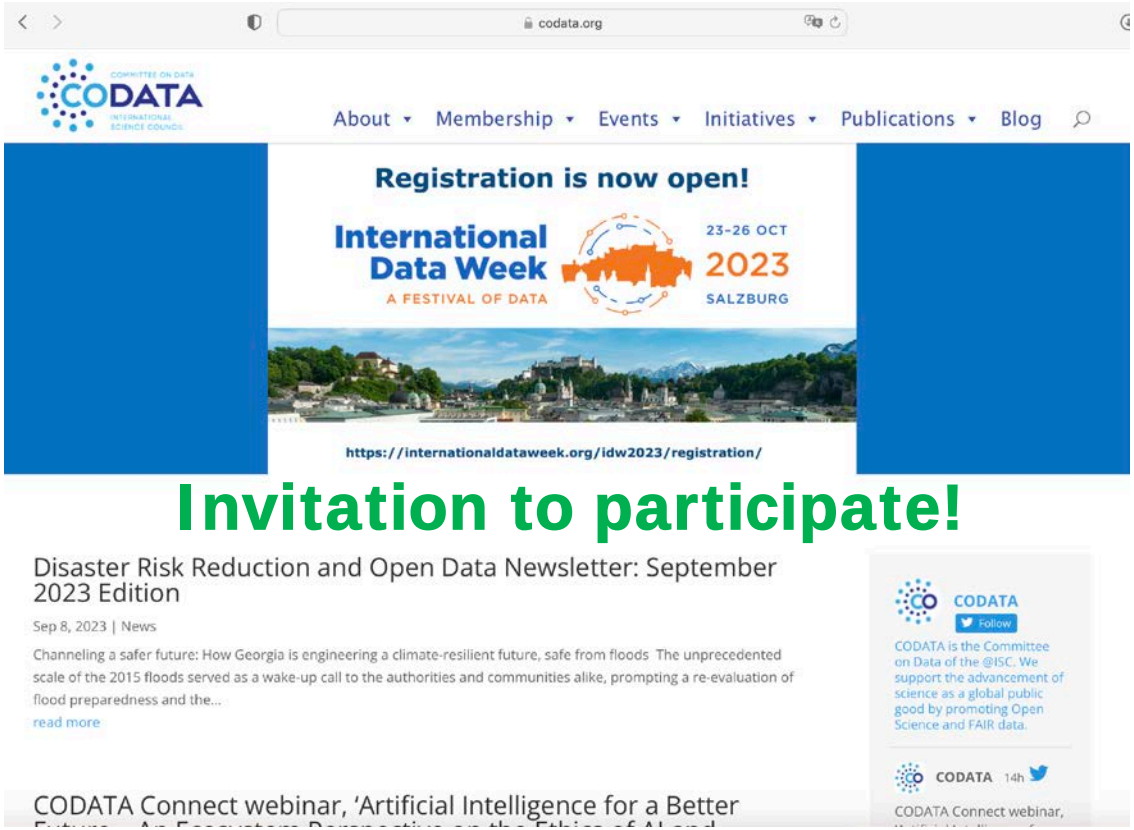
- Pressure/nudge to make data open (need to justify investment)
- Accumulation of expertise, a go-to person on data management and sharing
- Ensured continuity (i.e., reduced dependence on specific individuals on maintenance of datasets)
- Transition to professional data stewardship is both a technical and a cultural shift

- Different cultures regarding data in different disciplines
- Diversity of data types across domains
- Need to possess a unique set of expertise: technical, legal, strong interpersonal skills, etc. Need training. Career track for data stewards?
- Different levels of data literacy among researchers
- Complexity of incentives for researchers

CODATA

Committee on Data of the International Science Council (ISC)

- Exists to promote global collaboration to improve the availability and usability of data for all areas of research
- Supports the principle that data produced by research and susceptible to be used for research should be as open as possible and as closed as necessary
- Works to advance the interoperability and the usability of such data



Conclusions and final remarks

- Making our research outputs including data open can greatly accelerate scientific progress
- Research funded by public funding should, in general, be open to the public (as open as possible and as closed as necessary)
- Data is a key enabler for sustainability science and decision making
- Interdisciplinary research requires reliable data that are multi-dimensional across domains, long time-scale, spatially granular, is available without delays
- Responsibility for providing sustainable access to research data should be embraced by research institutions (data stewardship)
- Sharing good practices across research institutions will accelerate the transition to professional data stewardship in research

Questions? Comments? Feedback?