United Nations | Department of Economic and Social Affairs
Statistics • SDG Knowledge

LinkedSDG

Connecting SDGs-related Information Resources with Linked Data

Luis G. González Morales, Aaron Helton,
Amit Joshi and Sean Lovell

The 2030 Agenda data challenge

- The availability and access to high-quality, timely, disaggregated and reliable data is **fundamental** to the successful implementation and monitoring of the 2030 Agenda.
- The challenge is not only to **fill data gaps**—but also to **integrate** and make the vast amounts of existing SDG-related data and information **accessible** to decision makers in a **meaningful** way.

Semantic web

- Semantic web technologies provide the means enhance information seeking, targeting and integration through metadata
- The goal of a semantic web is to make metadata about heterogeneous, distributed information resources machine readable and available through the web, so that third-party applications can find these resources, access them, reason over them and operate on them.
- In order to bring information resources into the semantic web, it is necessary to describe these resources by means of standardized, machine-readable metadata schemas

LinkedSDG

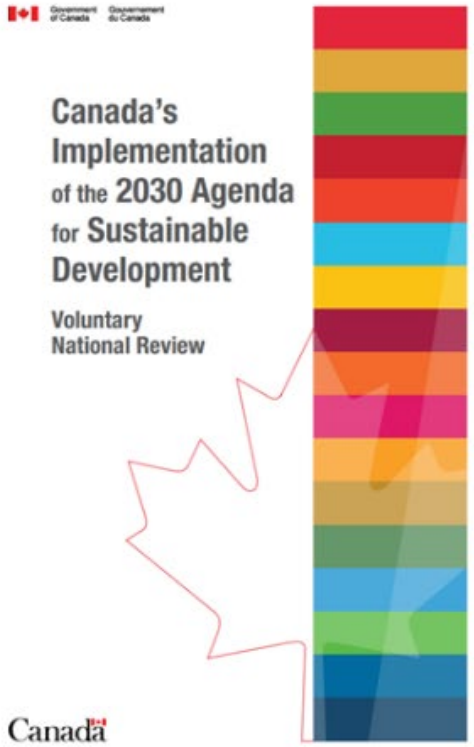
- Developed under the leadership of the Statistics Division and the Division for Sustainable Development Goals' (DSDG) of the United Nations Department of Economic and Social Affairs, with the support of resources from the EU grant entitled "SD2015: delivering on the promise of the SDGs".
- Allows to automatically connect SDGs-related information resources using linked open data techniques.
- Provides a foundation for **standards-based data exchange** and builds a common framework and guidelines to improve the **visibility, interoperability and usability** of citizen-science data and other information resources relevant to SDGs.

http://linkedsdg.apps.officialstatistics.org/

LinkedSDG

- **Key features**
  - Automatic Extraction of Concepts related to SDGs
  - Automatic Extraction of Geographical Areas
  - Interactive SDG wheel chart for most relevant SDG
  - Discoverability of statistical data
  - Support for Multilingualism
  - Developer Friendly

LinkedSDG How it works…



DATA SERIES

Tuberculosis incidence (per 100,000 population)

country	year	value	unit of measurement
Canada	2000	6.4	Per 100,000 population
Canada	2001	6.6	Per 100,000 population
Canada	2002	6.1	Per 100,000 population

Upload a document (PDF, DOC, DOCX, HTML) related to SDGs or paste its URL

Extract concepts

Link to Sustainable Development Goals, targets, indicators, and statistical data series

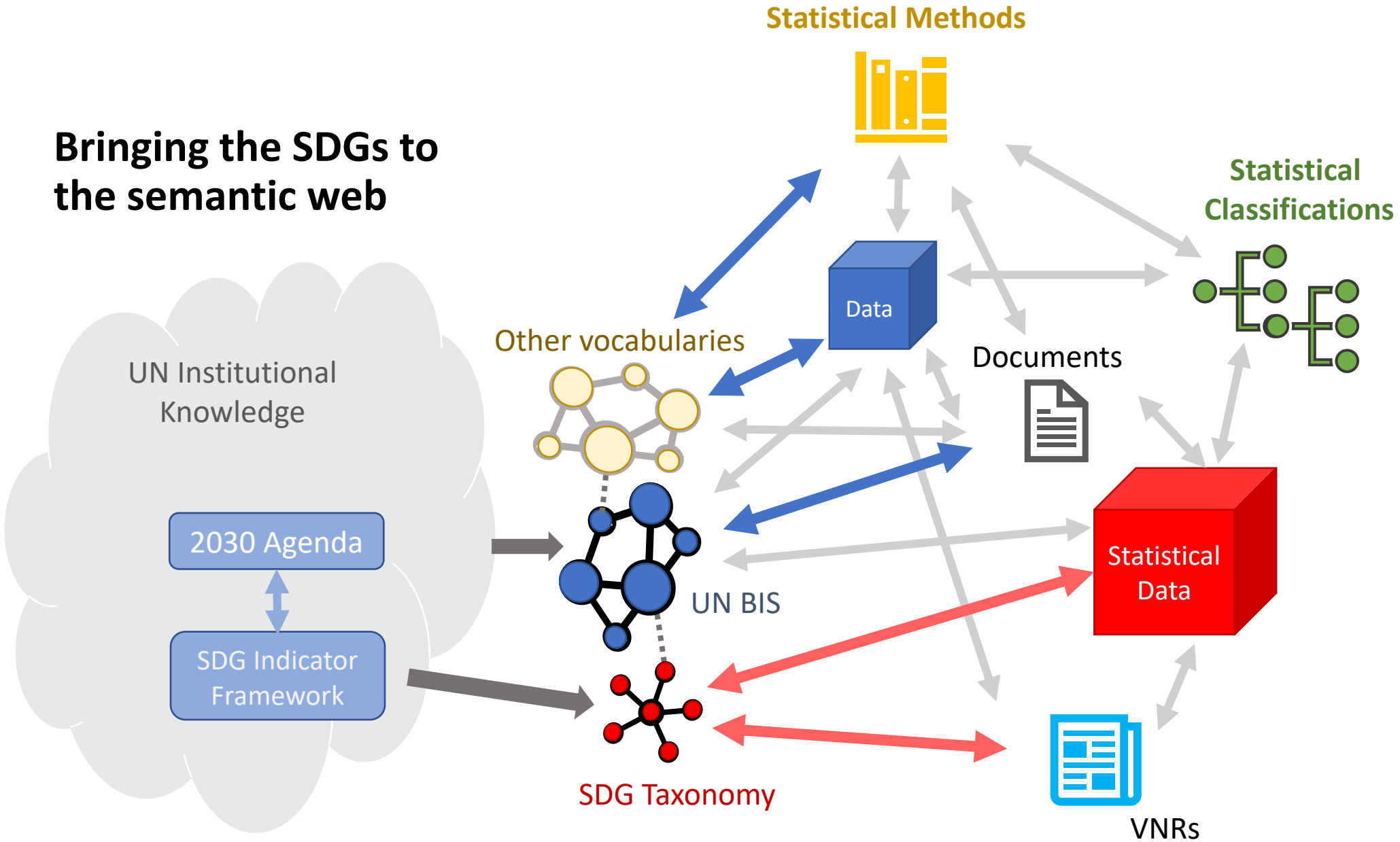
LinkedSDG

http://linkedsdg.apps.officialstatistics.org/

http://linkedsdg.org

statistics@un.org

Bringing the SDGs to the semantic web



Creating semantic content

- The availability of a critical mass of semantic content is crucial for wide adoption of semantic web technologies
- The question of how to create (1) **ontologies** and (2) **resources annotated with semantic metadata** deserves priority attention.
- The process of translating user needs into semantic artifacts involves several actors, including:
  - Domain experts
  - Knowledge engineers
  - Tool developers

Process of ontology creation

1. Conduct domain analysis and identify and exploit knowledge sources
2. Produce a first conceptualization of the domain
    - Proto-concepts
    - Initial relations and axioms
3. Formalize into a prototype ontology
4. Refine and validate

Creating semantic annotations can be done at different levels of sophistication

- Manual annotation
- Information extraction and natural language processing techniques
- Machine learning algorithms based on a small user-generated training set

→ Creation of semantic annotations can be integrated into the content authoring process, or take place afterwards

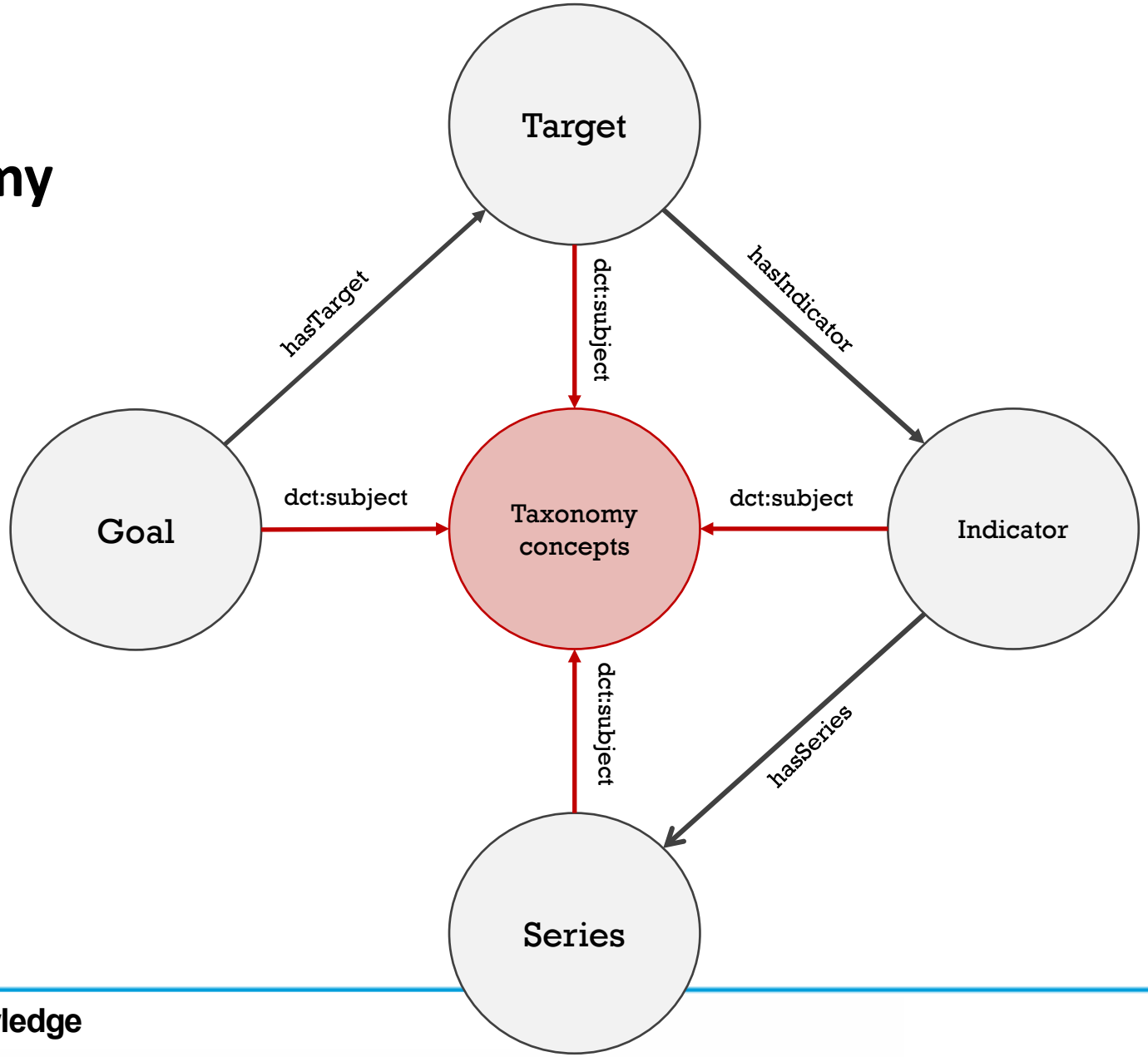
SDG Knowledge Organization System (KOS) :

SDG meets Semantic Web

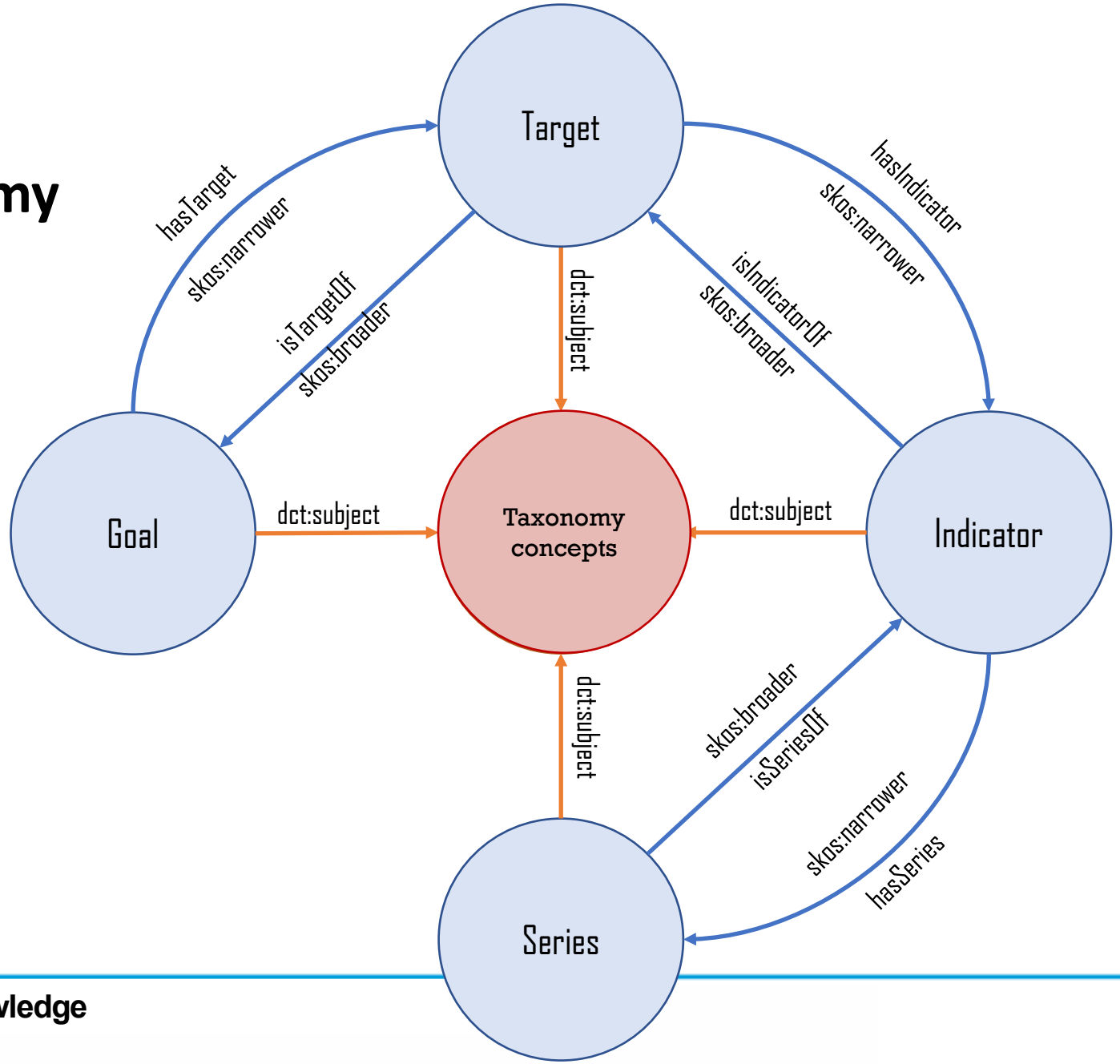
SDG KOS

- System of unique identifiers for the SDGs
  - Linked to the UN Bibliographic Information System and other key vocabularies
  - Web-based mechanism to make SDG-related data and information assets accessible, inter-operable and re-usable
  - Allows users to organically discover relevant and authoritative SDG data and information among the multitude of available sources on the web.
  - Enables the development of applications that integrate authoritative data, concepts and definitions from a large number of formerly unknown sources

SDG taxonomy



SDG taxonomy



United Nations A/RES/70/1

General Assembly

Distr.: General
21 October 2015

Seventieth session
Agenda items 15 and 116

Resolution adopted by the General Assembly on 25 September 2015

[*without reference to a Main Committee (A/70/L.1)*]

70/1. Transforming our world: the 2030 Agenda for Sustainable Development

The General Assembly

Adopts the following outcome document of the United Nations summit for the adoption of the post-2015 development agenda:

Transforming our world: the 2030 Agenda for Sustainable Development

Preamble

This Agenda is a plan of action for people, planet and prosperity. It also seeks to strengthen universal peace in larger freedom. We recognize that eradicating

Sustainable Development Goals

Goal 1. End poverty in all its forms everywhere

Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture

Goal 3. Ensure healthy lives and promote well-being for all at all ages

Goal 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

Goal 5. Achieve gender equality and empower all women and girls

Goal 6. Ensure availability and sustainable management of water and sanitation for all

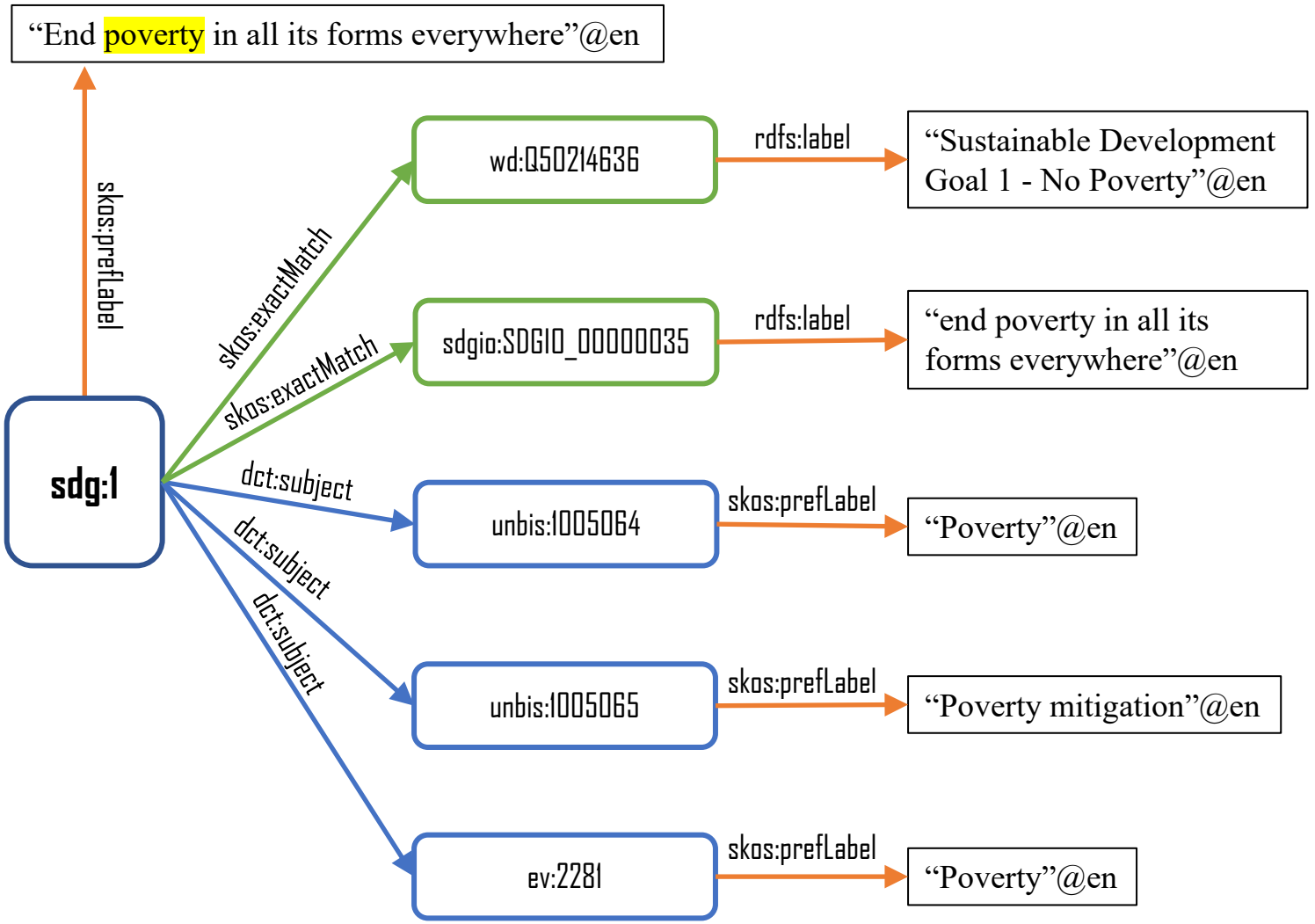
Goal 7 Ensure access to affordable, reliable, sustainable and modern energy for all

Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

Goal 9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

Goal 10. Reduce inequality within and among countries

SDG taxonomy



Multilingualism

- SDG taxonomy:
  - English
  - French (~50%)
  - Spanish (~50%)
- UNBIS
- Eurovoc

Multilingualism

our world: the 2030 Agenda for Sustainable Development

Goal 1. End poverty in all its forms everywhere

1.1 By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day

1.2 By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions

1.3 Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable

1.4 By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance

1.5 By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters

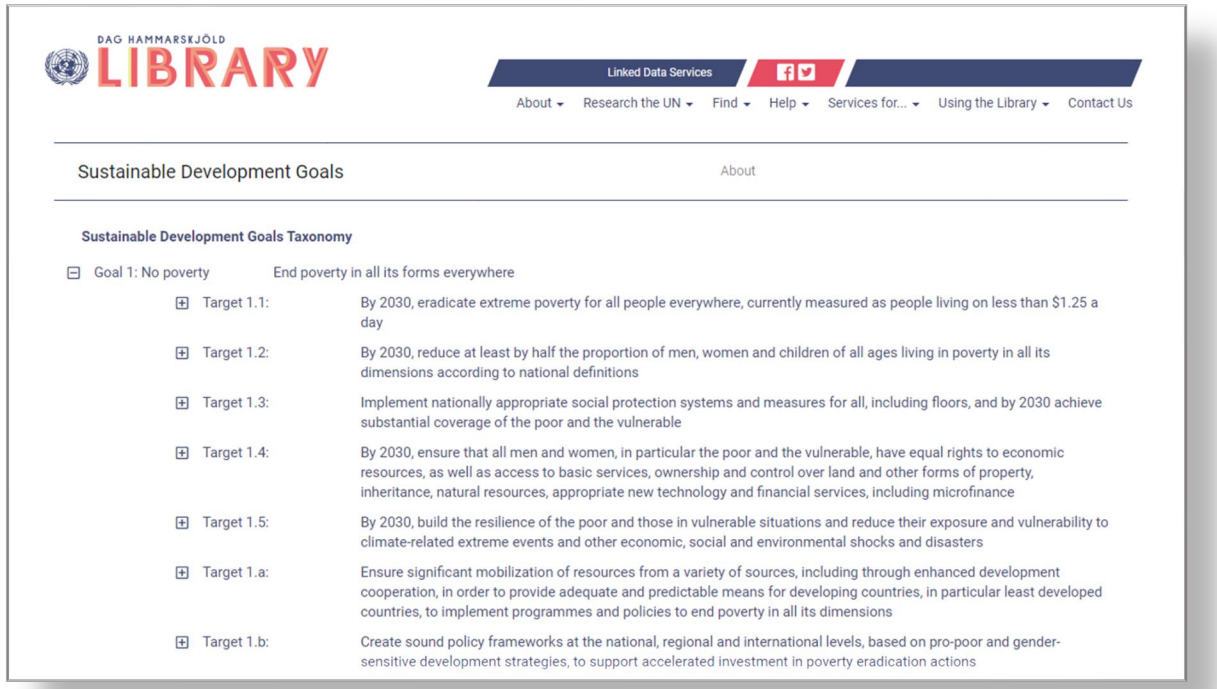

```
{
    "@context": {
        "dct": "http://purl.org/dc/terms/",
        "skos": "http://www.w3.org/2004/02/skos/core#",
        "sdgo": "http://metadata.un.org/sdg/ontology#"
    },
    "@id": "http://metadata.un.org/sdg/1.1",
    "@type": [
        "sdgo:Target",
        "skos:Concept"
    ],
    "skos:prefLabel": [
        {
            "@value": "By 2030, eradicate extreme poverty for all people everywh
              people living on less than $1.25 a day",
            "@language": "en"
        },
        {
            "@value": "De aquí a 2030, erradicar para todas las personas y en to
              extrema (actualmente se considera que sufren pobreza extrema las p
              de 1,25 dólares de los Estados Unidos al día)",
            "@language": "es"
        },
        {
            "@value": "D'ici à 2030, éliminer complètement l'extrême pauvreté da
              actuellement du fait de vivre avec moins de 1,25 dollar des États-
            "@language": "fr"
        }
    ]
}
```

metadata.un.org

United Nations' platform for Linked Data Services

United Nations' platform for Linked Data Services

The **UNBIS Thesaurus** and the **Sustainable Development Goals taxonomy** are hosted by the Dag Hammarskjöld Library at http://metadata.un.org/



About metadata.un.org: How did we get here?

- From Lotus Notes to Linked Data
  - Origin: External cooperation with EU Publications office (2013) to map UNBIS Thesaurus to EuroVoc
    - Get the data out of Lotus Notes
    - Transform it to a Linked Data structure: Basic SKOS + some EuroVoc for hierarchies
    - Align
  - Result: Alignment between EuroVoc and UNBIS Thesaurus

About metadata.un.org: How did we get here?

But that was just data. What about the website?

- URIs all pointed to the old Thesaurus website
- Pressure to rehost the terms as Lotus Notes was being phased out
- We needed a new website
  - Several projects over several years to build a modern interface
  - Metadata.un.org domain attracted attention
  - 2019: Parallel projects to finish metadata.un.org alongside deployment of SDG taxonomy

About metadata.un.org: How did we get here?

- None of this was possible without the work that had come before.
- While there were technical challenges, they were secondary to the data challenges
    - That Lotus Notes data? Library Taxonomists had used ISO 25964 for structural guidance.
    - Transformation followed the W3C's SKOS recommendation.
    - Taxonomists in the Library specialize in maintaining this data, including its structure.
    - The Library's Taxonomists also consulted on mappings between UNBIST and the SDG taxonomy.

The main point? There are standards you can use. There are professionals you can consult. You are not alone.

From application to *Software as a Service*

From a working demo to Software-as-a-Service

- Linked SDGs is a distributed system with a **React front-end** application, **five API-based micro-services**, and an **Apache Jena Fuseki triple store database**
  - Initially, the code was separated into distinct code repositories
  - This made sense from a development perspective as it allows services to be loosely-coupled and pieces could be swapped out for others or upgraded easily
- However, from a deployment perspective, this presented challenges:
  - Configuration, routing, dependencies, would all need to be carefully documented and managed manually
  - If the overall system went down or had a disaster event, restoring service would take a very long time
  - The system may not scale well or easily, and a service crashed, it would need to be manually restarted or separately automated
  - System monitoring and observability would need to be handled separately
  - The application is experimental and may need to be migrated to other cloud providers or to in-house; redeploying manually would be a nightmare

Containers to the rescue?

- There was a realization that we needed to **containerize the application**, and that meant we'd need to use a container orchestrator like Kubernetes
  - **Application container:** a standardized unit of software that packages code and dependencies so that several applications requiring distinct dependencies can be run on one operating system
  - **Kubernetes:** a portable, extensible, open-source platform for managing containerized workloads and services that facilitates both declarative configuration and automation
- Deploying LinkedSDGs on Kubernetes would be challenging, but would provide a few **key benefits**:
  - **Self-healing.** If a microservice dies, Kubernetes will restart it automatically.
  - **Configuration management**. Container and system configuration, including networking, storage requirements, application resource limits, and other things all become part of your code base, decoupling infrastructure from application from infrastructure concerns
  - **Runs the same on a laptop as it does in the cloud.** This means you can get the application ready for production in your local environment and minimize unnecessary cloud spend
  - **Ease of patching:** Updating the software each microservice runs is as simple as re-building the container
  - And more…

Plan of attack

- Phase One: containerize each microservice using Docker
  - Automate the installation of dependencies
  - Move configuration from the application to environment variables that could be passed on runtime (find hard-coded database URLs, etc)
  - Ensure that each application can build to a container and send these to a Docker registry
- Phase two: build Kubernetes deployment templates
  - We started with docker-compose, and used a CLI tool to translate these into draft Kubernetes objects
  - Modified and customized the auto-generated templates
  - Deployed to an isolated namespace on the UNGP Kubernetes cluster on GCP into an autoscaling nodegroup

What did we learn on this deployment journey?

- Make “data engineering” (or DevOps engineering, etc) an explicit part of data science innovation projects: using containers, deploying to Kubernetes simplified the task of moving this application from concept to production
- Multi-disciplinary teams were essential: it would not have been possible to refactor the code base with just one person. We needed to involve ontology experts and data scientists alongside data engineers to understand the code base and make it container-ready
- Code literacy was essential: all members of the team were comfortable using code to perform their duties, whether it was project management, statistics, data science, or engineering tasks. This provided a common language and made work very efficient. Nobody used the word “IT” to abstract technical things that need to be done.

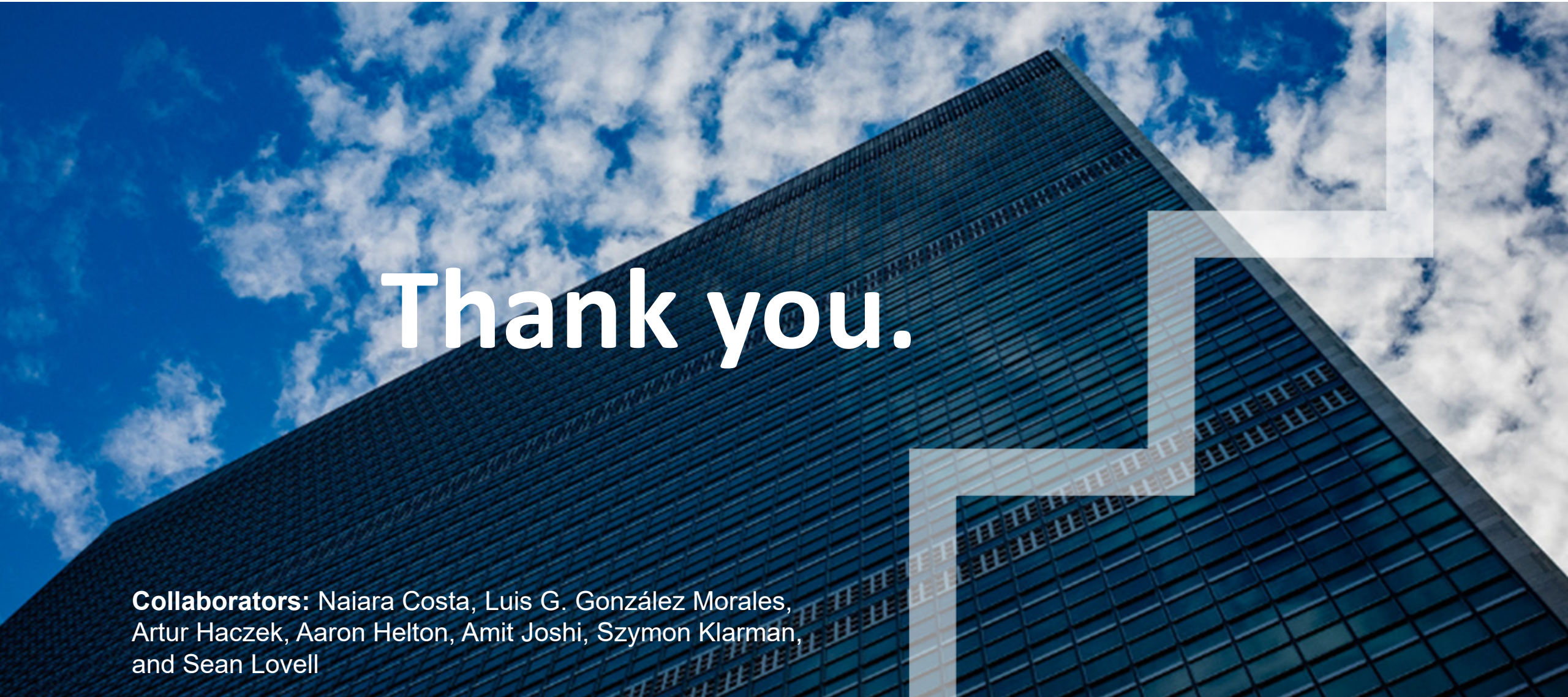
Next Steps

- Establish LinkedSDG as a platform to support interoperability of SDG data among data publishers, including data published from citizen scientists and National Statistical Offices.
- Provide stakeholders' alternate means to streamline their data management workflow by providing RESTFUL web APIs and exposing SDG ontology and KOS through specific endpoints for programmatic consumptions.
- Expand the platform to allow data publishers to store SDG data in the platform and publish to Linked Open Data (LOD) cloud to make SDG data openly accessible.
- Research use cases from different groups of stakeholders and citizen scientists working on SDGs and identify the areas of improvement
- Build upon the existing work on ontologies and KOS, and map to other publicly available vocabularies and ontologies.
- Expand the existing work to use RDF Data Cube Vocabulary for representing statistical data, including mappings to SDMX vocabulary.
- Seek volunteers and partners to build the next version

Final thoughts

- The adoption, survival and spread of particular metadata schemas and knowledge organization systems among a community of data producers and consumers is a social phenomenon (Sicilia, 2014)
    - It involves not only technologies and tools, but also procedures, people and organizations.
    - It requires endorsement and adoption of technical standards by a user community.
- Simplicity and ease of use of the technical solutions is crucial for adoption

United Nations | Department of Economic and Social Affairs
Statistics • SDG Knowledge



Thank you.

Collaborators: Naiara Costa, Luis G. González Morales, Artur Haczek, Aaron Helton, Amit Joshi, Szymon Klarman, and Sean Lovell