Modernization and agility powered by Communities of Practice

Olav ten Bosch, Matjaž Jug
Statistics Netherlands

UNECE ModernStats World, Geneva, Oct. 2024

Innovation benefits:
new statistics, faster time to market, lower costs, new partnerships and relationships..

Standardisation benefits:
reliability, predictability, safety, lower costs, repeatable processes, consistent measurements..

Modernisation

Communities

Communities are the ***enabler*** for agility and modernization


This wasn’t built in one day!

Communities of Practice (CoP)

Definition

A community of practice (CoP) is a group of people who **"share a concern or a passion for something they do and learn how to do it better as they interact regularly“**. The concept was first proposed by cognitive anthropologist **Jean Lave** and educational theorist **Etienne Wenger** in their 1991 book Situated Learning.

The structural characteristics of a community of practice are again redefined to a domain of knowledge, a notion of community and a practice.

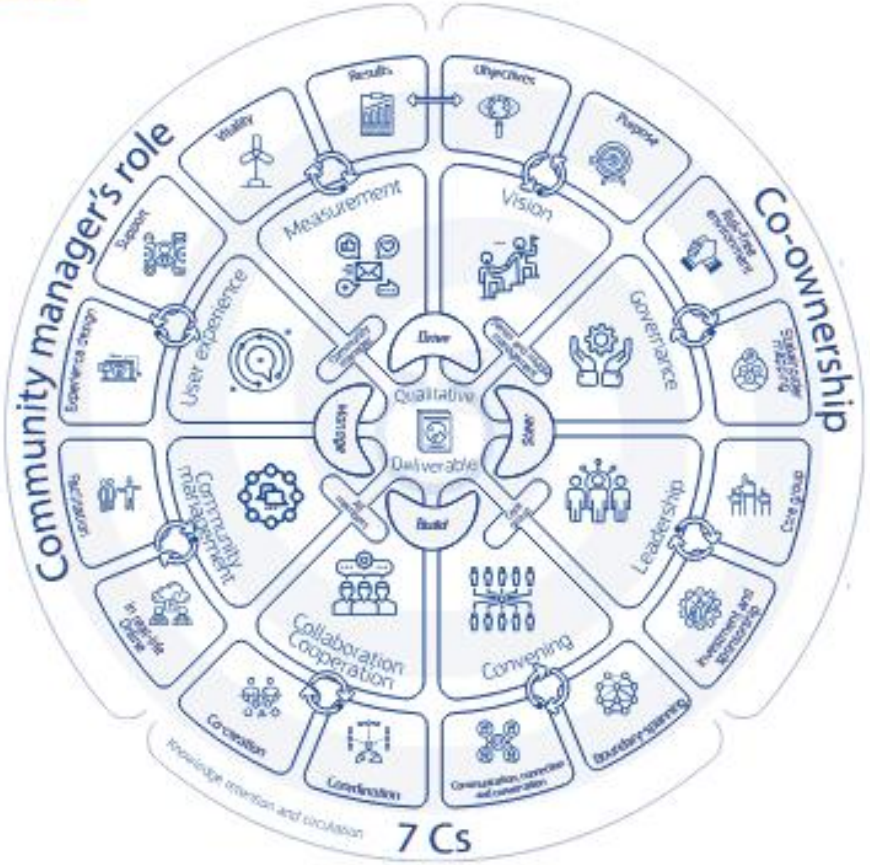
Types: Helping / Best Practice / Knowledge Stewarding / Innovation

Communities in Official Statistics are often based on standards, methodology, technology or modernization topic

Introduction to communities of practice - wenger-trainer
https://www.communityofpractice.ca/background/what-is-a-community-of-practice/

5.1.1. THE COMMUNITIES OF PRACTICE SUCCESS WHEEL

Success factors



[Community of Practice Playbook - openresearch.amsterdam]

Levers

1. Filtering
2. Amplifying
3. Investing & providing
4. Convening
5. Community building
6. Learning and facilitation

slack

GitHub

[Community of Practice Playbook - openresearch.amsterdam]

Example: UN PET Lab

UN PET Lab
United Nations
Privacy Enhancing Technologies Lab

Domain of knowledge: Privacy-enhancing Technologies (PETs)
Type: Helping / Best Practice / Knowledge Stewarding / Innovation
https://unstats.un.org/bigdata/task-teams/privacy/index.cshtml

"..investigating methodologies and approaches to mitigate privacy risks when using sensitive or confidential data, which are collectively referred to as privacy-enhancing technologies (PETs)."

...

2023_UN PET Guide.pdf

UNBigData

Example: SDMX

Domain of knowledge: SDMX standard
Type: Helping / Best Practice / Knowledge Stewarding / Innovation
https://sdmx.org/

"SDMX sponsoring organisations wish to increase and formalise the participation of the **SDMX User Community** (statistical offices, central banks and other national and international organisations dealing with statistics) in the management and development of the SDMX Technical Standards and statistical guidelines."

…

Example: OECD SIS-CC

Domain of knowledge: .Stat data dissemination tools
Type: Helping / Best Practice / Knowledge Stewarding / Innovation
https://siscc.org/

"A reference **open source community** for official statistics, focusing on product excellence and delivering concrete solutions to common problems through co-investment and co-innovation."

…

Example: Open source in Official Statistics

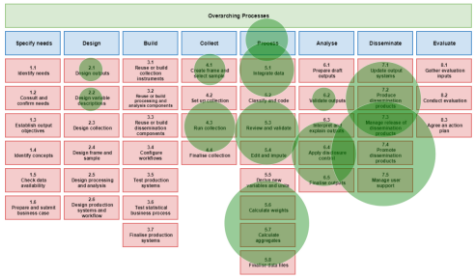
Curated list of software for official statistics

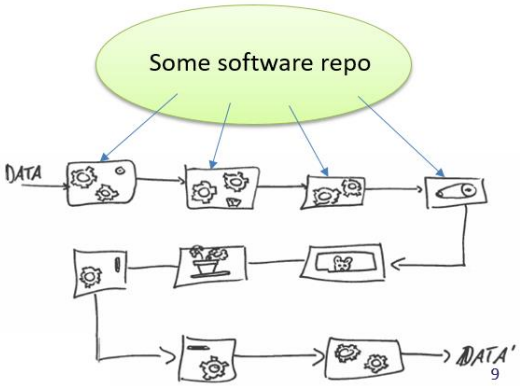
awesome

www.awesomeofficialstatistics.org

Domain of knowledge: Statistical building blocks

Type: open-source tools / Methodology / Knowledge / Innovation





9

Awesome list promotions

- Unece SDE '17
- Unece SCFE '17
- uRos '18
- Unece SDE '18
- Estat Validition Grants kickoff '18
- uRos '19
- Unece modernstats World '19
- Unece modernstats '20 (virtual)
- uRos '20
- ICDSOS '21
- uRos '21
- TF-TSS '22
- Etc. + here

Virtual ☹

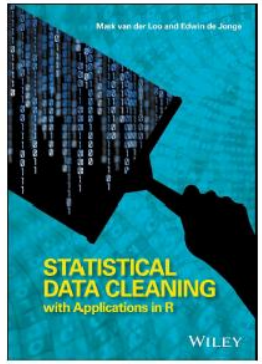
15

Example: Statistical Data Processing

Domain of knowledge: validation, editing, imputation, aggregation

Type: R packages

MPJ van der Loo and E de Jonge (2018)
Statistical data cleaning with applications in R
John Wiley & Sons, NY.



- ***validate***: check data based on validation rules
- ***dcmodify***: change data based on 'if-this-then-that' rules
- ***errorlocate***: locate errors based on validation rules and mark them for correction
- ***simputation***: many different imputation methods
- ***rspa***: adapt numerical records to fit (in)equality restrictions
- ***deductive***: solve errors based on control rules
- ***validatetools***: find inconsistencies and redundancies
- ***accumulate:*** advanced group aggregation
- ***lumberjack:*** standardized logging

Community / current use

Inside Statistics Netherlands:

- Domains: social and economical statistics, agriculture, international trade, education, environment, emmisions, income, shipping, STS, recreation, museums, and many more
- As a validation tool for checking output data before publication

Outside Statistics Netherlands:

- Iceland, Italy, Denmark, Brasil, US and probably many more
- USDA-NASS case: Rule-based Data Validation and Reconciliation of Survey Responses
- Used for training around the world (Denmark in Jordan)
- Used as formal base for implementing data validation in Python

Data science / AI / ML
EU Data spaces / HVD
PETs / sMPC / HE / Federated Learning / Synthetic data
Taxonomies / Knowledge Graphs / Ontologies / Metadata
Citizen science / data donation
Open data / Open models / Open science / Open government / Data stewardship
Green deal / energy transition
Webscraping
Sensor data / IOT / Edge c.
Remote Access/ Microdata
Validation / data cleaning
OS Statistical software
Cloud / Kubernetes

Yin and yang

In Ancient Chinese philosophy, yin and yang is a Chinese philosophical concept that describes how obviously opposite or contrary forces may actually be complementary, interconnected, and interdependent in the natural world, and how they may give rise to each other as they interrelate to one another.
Wikipedia

GSBPM
GAMSO
GSIM
CSPA
CSDA
LIM
MMM

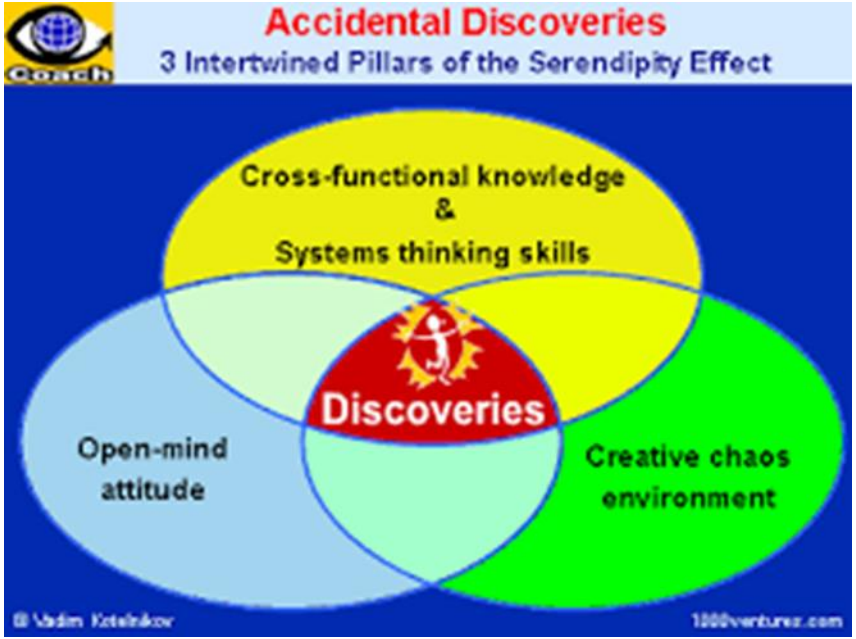
Standardisation enabling Innovation

Innovation helping Standardisation

SDMX
DDI
LOD / RDF / (S)KOS
DOI
DCAT / StatDCAT
SIMS
JSON-STAT
Web retrieval policy
VTL
W3C / ISO
…

Twist - Serendipity effect!

- Not enough to have many communities working on specific topics / standards / technologies
- Connect them in an agile way to reach successful modernisation
- Example: PETs <-> metadata groups <-> data science <-> AI <-> data editing <-> architecture
- …and don’t forget communities outside of HLG-MS & ESS!



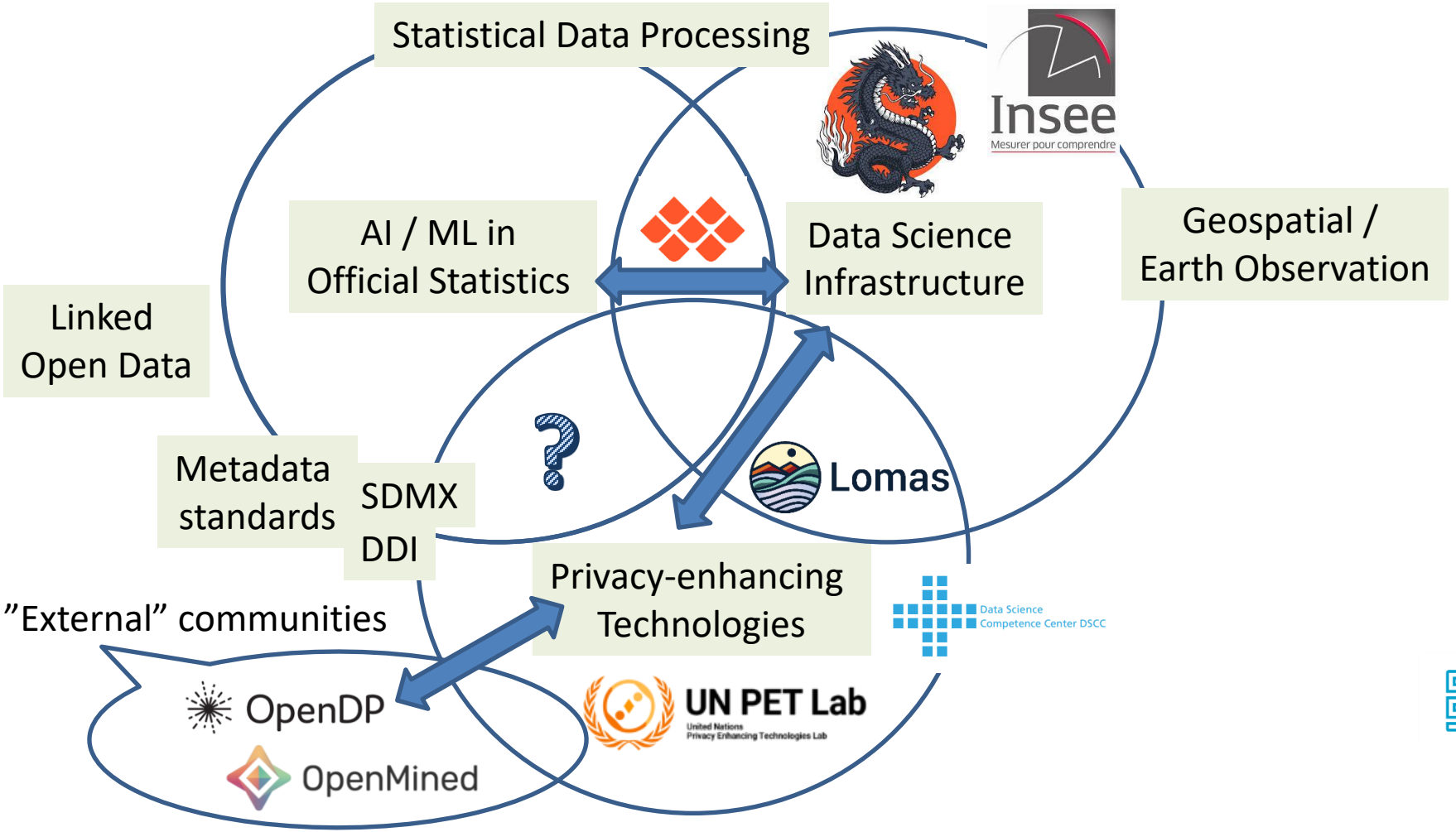
Glue communities together, but how?

The European **One-Stop-Shop** for **Articial Intelligence and Machine Learning** for Official Statistics (**AIML4OS**).

Onyxia: An Open Source **Cloud Native Data Science Platform**

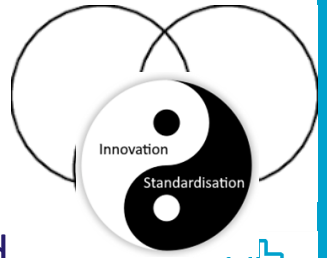
Lomas

Lomas is open-source, portable, modular, easy to use, scalable client-server platform for **private data science**


Statistical Data Processing
Insee
Mesurer pour comprendre
AI / ML in
Official Statistics
Data Science
Infrastructure
Geospatial /
Earth Observation
Linked
Open Data
Metadata
standards
SDMX
DDI
Lomas
Privacy-enhancing
Technologies
Data Science
Competence Center DSCC
”External” communities
OpenDP
OpenMined
UN PET Lab
United Nations
Privacy Enhancing Technologies Lab

Recommendations

- Continue using core UNECE ModernStats standards (GAMSO, GSBPM, GSIM..)
- Continue updating UNECE/ESS reference architectures to support innovative approaches (examples CSPA, CSDA, BREAL)
- Partner with Academia and Industry sectors to gain knowledge and resources
- Organize and support Communities of Practice (examples: Data Science, Geospatial, PETs..) that use standardization in practice
- ..
- Q: how can we effectively connect Communities of Practice and sustain the serendipity effect?



Questions, ideas, suggestions

Olav ten Bosch o.tenbosch@cbs.nl

Matjaz Jug m.jug@cbs.nl

and keep an eye on:

awesomeofficialstatistics.org awesome Star 183 Fork 47