



Data Ecosystems: a new challenge for official statistics

Session number 20 – Big Data Oriented Systems

Date: 2 June 2016

Donatella Fazio, Istat, dofazio@istat.it
Enrico Giovannini, University of Rome Tor Vergata,
enrico.giovannini@uniroma2.it
Marina Signore, Istat, signore@istat.it



Madrid, May 31 - June 3

Data Ecosystems: a new challenge for official statistics

Web2.0: a new relationship between NSIs and society

Phase1 (until '80s): data collection by sample surveys and census via Questionnaires (on paper) designed by committees of “experts” defining ex-ante the societal information needs

*Citizens, enterprises and Institutions are essentially **respondents***

Phase2 ('80s–'90s): the use of administrative data and Web1.0 tools improve the collection of data and foster the release of statistics

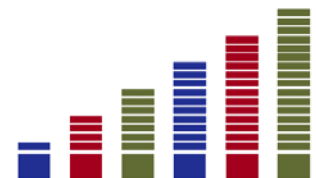
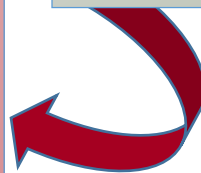
*Citizens, enterprises and Institutions are **respondents** and become more and more **users***

Phase4 (nowadays): the availability of new sourced of data on Internet (**Big Data, Open Data**) coming from satellites, sensors or indirectly generated by citizens (via Social Media, etc.) and the **crowd sourced data** directly locally generated by civil society communities through collaborative platforms, are calling NSIs to open the door to the “data revolution”

*Citizens, enterprises and Institutions are **respondents, users, interpreters** and also **producers of data (prosumers)***

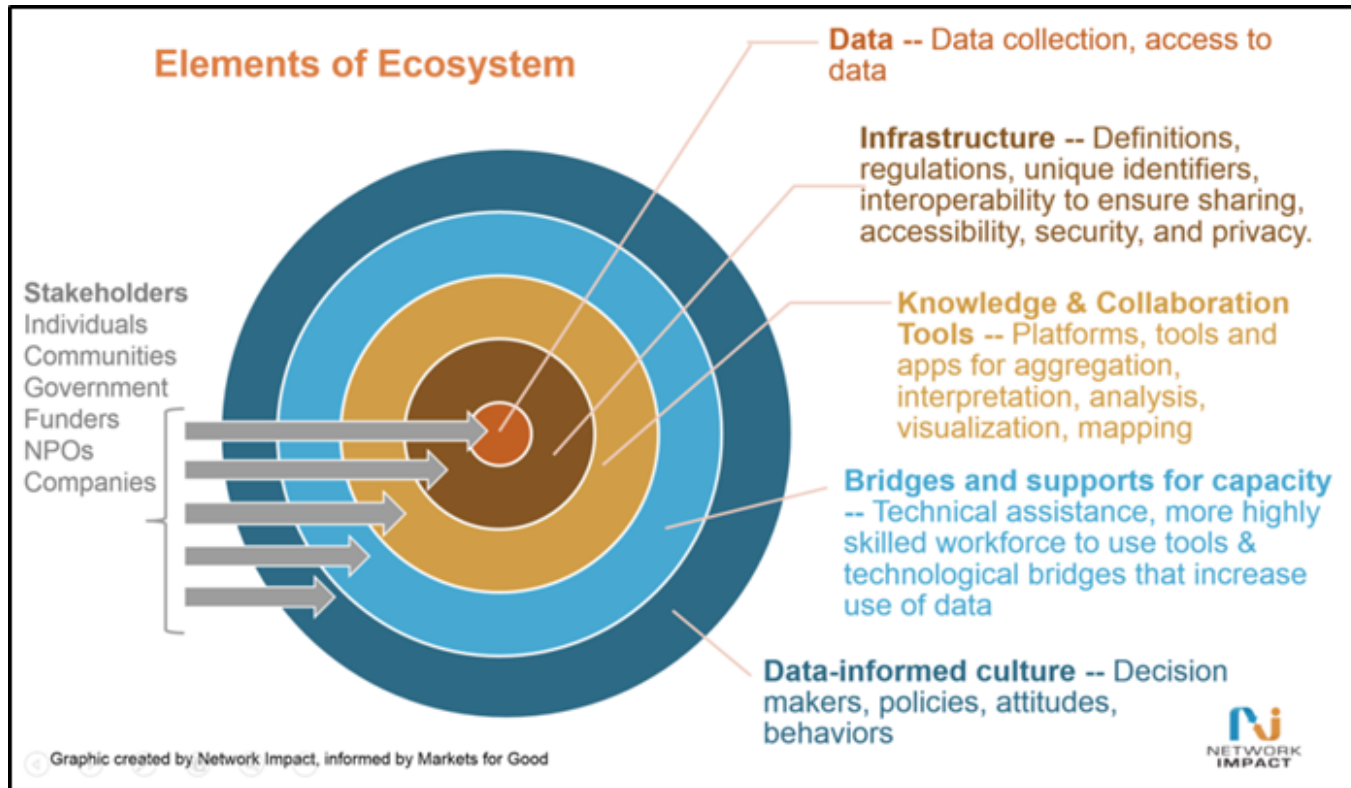
Phase3 (start 2000): the Beyond GDP debate opens the perspective towards the stakeholders' inclusion for the construction of new measurements for societal progress. Facilitated by Web2.0 tools

*Citizens, enterprises and Institutions are **respondents, users,** and now become **co-designers** and **interpreters** of data*

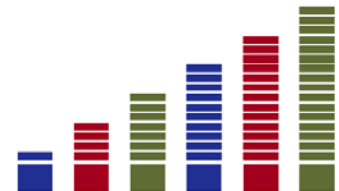


Data Ecosystems: a new challenge for official statistics

The Data Ecosystem approach/1



Source: <http://www.networkimpact.org/leveragingtech/>



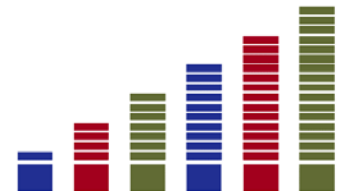
Data Ecosystems: a new challenge for official statistics

The Data Ecosystem approach/2

The management of the Data Ecosystems is based on the concept of *data cycles*, where users/producers of data and information are involved in re-shaping the statistical information in a jointly top-down/bottom-up approach.

So the role of a *passive final user* is replaced by a *proactive community of users*, able to add quantitative and qualitative information that can enrich and possibly correct the official statistics.

In this approach, data management and capacity building can be supported by the producers of official data and by new figures belonging to the community of the Ecosystem, the so called *infomediaries*, i.e. intermediate consumers of data such as builders of apps and data wranglers, that can develop applications to facilitate a better understanding and use of data and information.



Data Ecosystems: a new challenge for official statistics

The construction of Data Ecosystems: two recent experiences

AT GLOBAL LEVEL BY WEB-COSI PROJECT

Web-COSI Wiki of progress stats

Data portal

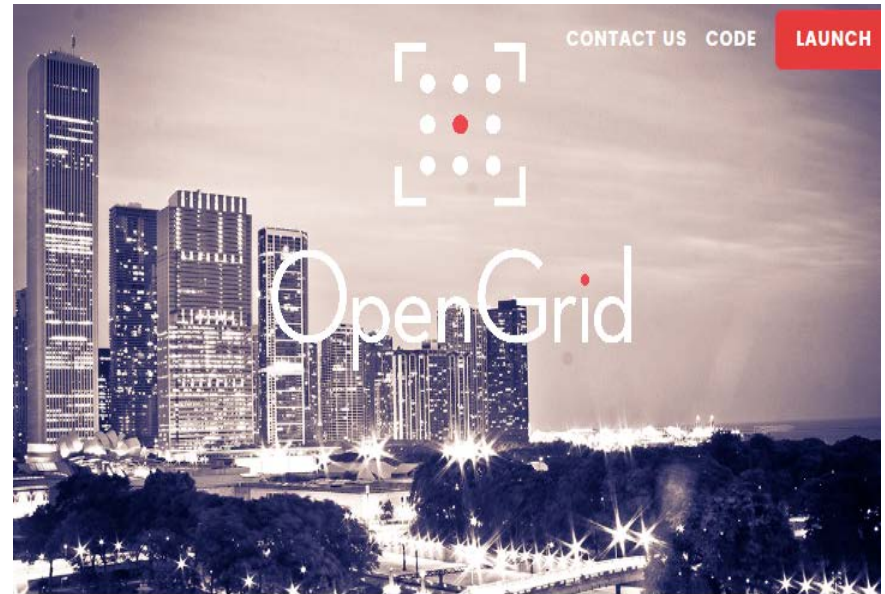
Youth Portal

EU University
Programme

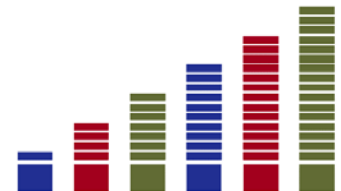
Interactive crowd-sourced map of
organisations and initiatives on well-being

www.wikiprogress.org

AT LOCAL LEVEL BY THE CITY OF CHICAGO



<http://opengrid.io/>



Data Ecosystems: a new challenge for official statistics

A new paradigm for quality for a Data Ecosystem approach

Phase1: Quality as accuracy of sample estimates (optimal sample designs, sampling variances, detail of the estimates). Afterwards, the non-sampling errors were conceptualised and formalised.

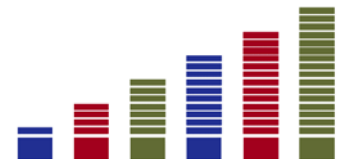
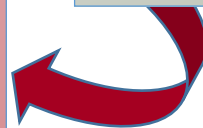
Phase4: Actually the exploitation of new data sources (*Big Data, Open Data, crowd-sourced data*) require a new shift in quality considering the weight of non-official data at global, national and local level.

Mainly at the local level, the problem is no more to manage trade-offs between quality dimensions, but to allow access to data (macro and micro) informing the users on limitations and cautions in the usage of data.

Quality labelling as “Experimental statistics” to be clearly distinguished.

Phase2: Product quality and process quality concepts were developed. The principles of *Total Quality Management* already developed in the industrial context (*continuous quality improvement*) were applied to official statistics. Users and stakeholders were given increasing importance with respect to the process cycle and quality management (from the identification of users’ needs to the assessment of the degree of users satisfaction). Then, quality was expressed by a set of requirements (relevance, accuracy, timeliness, accessibility..). *Code of Practice* (2005, rev 2011)

Phase3: Users of statistics started to play an active role for better data (*Committees of Users*) leading to a dynamic perception of quality balancing and managing trade-offs among different quality dimensions that have different weights for different users and for different uses.



Data Ecosystems: a new challenge for official statistics

THANKS FOR YOUR ATTENTION !

