



9th CONFERENCE

on the Evaluation
of EU Cohesion Policy

Shaping Transitions with Evidence

16-17 September 2021
Porto, Portugal

Workshop 5C:

The use of data for evaluation

Data sources for evaluation and the links among them

Thursday 16 September 2021

16:15 – 17:30 GMT+1

#CohesionEval2021

Regional and
Urban Policy

Housekeeping

- **In the room:** microphones available
- **Online:** Post questions in the Zoom “chat”
Moderator will ask online questions to panelist.
- Video and presentations will be online here: [Conference Website](#)



Twitter: #CohesionEval2021 #CohesionOpenData @RegioEvaluation

Chair / Moderator: John WALSH,
DG REGIO.B2 - Evaluation and European Semester

PANELISTS:

- **Veronique MENEZ**, Coordinatrice des fonds européens, Agence Nationale de la Cohésion des Territoires, FR
- **Daniele BONDONIO**, Professor of Statistics for Economics, Università del Piemonte Orientale, IT
- **Audronė SADAUSKAITĖ**, Researcher with Visionary Analytics, LT
- **Emanuela SIRTORI**, Development and Evaluation Unit CSIL, IT



DG REGIO perspective

Context: 2014-2020 ex post impact evaluation by EC and Member States

- Data from programme monitoring has an important value in evaluations; provides an “information map” / raw material to focus evaluation work; 2014-2020 saw improved monitoring and open data tools;
- [Cohesion Open data Platform](#) – [Data stories](#)



DG REGIO perspective (2/2)

- But programme monitoring does not solve all evaluation data needs;
- A close matching of data availability, evaluation methods and questions is critical for an evaluation to deliver robust results;



Data ... the oil of evaluation ?

Financial data - Process indicators - Outputs indicators

Results for beneficiaries (outcomes)

Geographic data - Administrative registers - Micro data

Statistical, context, impact indicators

Survey data (qualitative data on opinions / attitudes)

Big data (behaviour, movement, consumption pattern, etc.)



Opening remarks from panellist

- Experience and recent data work
- Insights, challenges and interesting upcoming developments



Veronique MENEZ



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

AGENCE
NATIONALE
DE LA COHÉSION
DES TERRITOIRES

A good evaluation starts on a robust analysis of monitoring data



Importance of robust monitoring data analysis 1/3

Excessive / over programming rate	RELEVANCE: Needs underestimated? Deadweight effect? EFFECTIVENESS: results can be observed
Insufficient programming rate	COHERENCE: Competition with other measures? IMPLEMENTATION: Low visibility? RELEVANCE: Needs over estimated? EFFECTIVENESS: No effects / impacts to expect
Variety of individual project cost	EFFICIENCY: Overfunding / Underfunding of operations?



Question stakeholders during the evaluation with these hypotheses
(managing authorities / targets / experts)

Importance of robust monitoring data analysis 2/3

- Feedback : Study on 2014-2020 EU funds utilisation in France (2019) – State of play
- Data analysed per thematic objective & Investment priority
 - Programming rate
 - Payment rate
 - Number of projects
 - Profile of beneficiaries
 - Profile of co-financers
 - Localisation of operations
 - Type of operations
 - Size of projects

Ex: Programming rate for one IP per region



Information collected: successful measures, reached beneficiaries, successful funding partners, appropriate size of projects, etc.

Lessons learned from a robust monitoring data analysis 3/3



What are the successful implementation schemes?

- Size of projects, cofinancers, profile of beneficiaries, type of operations, etc.



Which best practices can be identified for future programmes?



Which questions should be further investigated during the evaluation?

- EX: Deadweight effects, low visibility, absence of impacts, lack of certain beneficiaries initially targetted, etc.



On which topics can we expect impacts?

- No need to cover all aspects
- Rather focus on limited topics



Daniele Bondonio



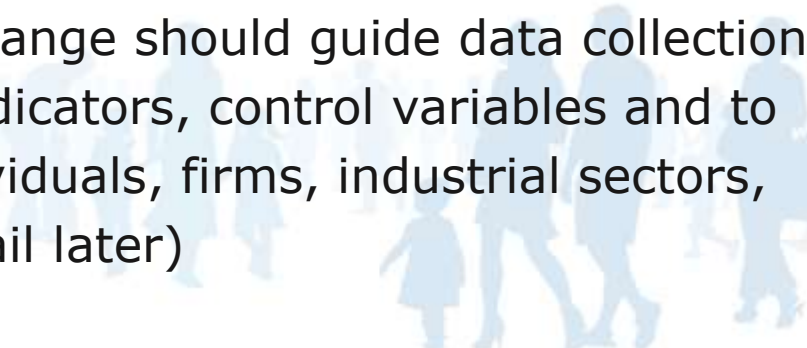
UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE


- **Scientific Expert** for DG-Regio Evaluation Helpdesk services
- **Recent evaluation projects** (principal investigator / scientific advisor):


-The impact of support to innovation projects on job quality (the study pioneers the linkage between 3 data sources: firm-level administrative records, employee-level data, program-activity data)

-Challenges in the use of Structural Funds expenditure data to estimate NUTS 2 regional growth (DG REGIO, Eval. Unit research project). Problems with geographic allocations of EU Funds expenditures and temporal lags between expenditures and measurable outcomes

Challenges / lessons / opportunities

- **Incorporate evaluation needs into the planning of monitoring activities**
E.g. provision monitoring system that maintain records of: non-supported applicants, dates of actual completion of the supported investments, ...
 - **Logic of intervention** / theory of change should guide data collection activities for the relevant outcome indicators, control variables and to choose the level of the analysis (individuals, firms, industrial sectors, municipalities, provinces..., more detail later)
- 

- **Primary data collection** by means of interviews and questionnaires can be an important tool to complement other data sources, also for performing quantitative analyses (including counterfactual approaches).
E.g. data on innovation outputs, innovation collaboration networks, and quality of life are often missing or unsatisfactory in databases from national statistical offices or administrative/commercial databases
 - Key issue: **how to incentivize better response rates**. Institutional options vary. Sometimes is possible to include in calls for applications the requirement to answer to questionnaires/interviews
- 

- **Big data** have a great potential for future causal impact evaluations in the context of individual-level data or for estimating the impact of infrastructural projects (e.g. recording bike path usage, measuring travelling times and congestions on roads, etc..);
 - For support to SME innovation projects and firm-level impact evaluations in general, it is crucial to **optimize access and availability of administrative sources**.
- 

- **Incorporating the needs for impact-evaluation data into the designing of the programmes**

If there is a high degree of uncertainty about the results of the policy intervention, the designing of the program could incorporate some features of a pilot design.

E.g.:

- the programme is implemented only in certain areas or for certain individuals or firms. Similarity to the medical field: when the treatment is experimental you do not treat all the patients at the same time in a same area)

- eligibility boundaries overlap those of available data

- Possible **institutional solutions for pilot programmes** to generate usable data:
 - delay of treatment schemes



Audronė SADAUSKAITĖ

Recent relevant assignments:



VISIONARY
ANALYTICS

Technical Assistance to the Managing Authority on Implementation of EU Cohesion Policy – preparations for 2021-2027 programming period	<i>Ministry of Finance of Lithuania</i>
<u>Evaluation of SME competitiveness measures under the Operational Programme 2014-2020</u>	<i>Ministry of Economy and Innovation of Lithuania</i>
<u>Ex-ante evaluation of digitisation of society measures for 2021-2027</u>	<i>Ministry of Economy and Innovation of Lithuania</i>
<u>Strengthening Evaluation Capacities in Lithuania</u>	<i>Ministry of Finance of Lithuania</i>

See more of our work at the EU and national level at [visionary.lt](https://www.visionary.lt)


- **Value of e-Cohesion data.** Example uses:
 - Defining scope of analysis
 - Directly informing some evaluation criteria (e.g. relevance, efficiency)
 - Assessing geographic spread of support
- **Importance of dataset merging for measuring impact**
 - Need for data on control group (n/a via e-Cohesion)
 - Use of surveys, web scraping, access to third-party databases
- **Importance of open (public) data**



Emanuela SIRTORI



CENTRE FOR INDUSTRIAL STUDIES

- Data collection:
 - Primary data collection, including through text mining, semantic analysis, webscraping
 - Experience in matching ERDF and H2020 project and beneficiary data (e.g. for DG DEFIS and DG RTD)
 - Data analysis:
 - from theory-based to counterfactual impact evaluations
 - Recent experiences with 2 EU level evaluations involving significant data inputs (for DG REGIO)
- 

Ex-post evaluation of 2007-2013 ERDF support to RTD infrastructures and activities

[#CohesionOpenData link](#)

Final report forthcoming

- **53** ERDF OPs in **18** Member States
- Fields of interventions: **01 & 02**
- Collected data on **9,800 operations** and **11,400 beneficiaries** and their financial information
- Added info on types of operations and beneficiaries (mainly manually)
- **Descriptive analysis** to understand how EU funds were used

Ex post 14-20 – Preparatory Study on the monitoring data on ERDF and Cohesion Fund operations in the 2014-2020 period

Ongoing

- **220** ERDF/CF OPs and **76** ERDF Cooperation Programmes in **28** Member States
- Thematic Objectives: **all**
- Collected data on **562,000 operations** and **1.15 million beneficiaries** and their financial information
- Added info on types of operations and beneficiaries (esp. from **matching** with other databases)
- Data on operations matched with **output indicators**
- Descriptive analysis and **cluster analysis** to understand how EU funds were used and assess data reliability
- **Quality assessment** of the whole monitoring system

A look into the databases

DB Beneficiaries	
1. General information	[...]
2. Programme and operation identification	CCI number
	Official operation Identification Code
	Official and ad-hoc operation code
	Operation name
	[...]
3. Beneficiary and intermediary identification	Beneficiary name
	VAT code (validated), BvD code, CORDIS id
	Role (leader / partner / final recipient (of financial instruments))
	Intermediary name
	[...]
4. Beneficiary's characteristics	Type of beneficiary (from cluster analysis)
	Size and ownership
	Location (NUTS)
	List CORDIS projects
	[...]
5. Financial information on the beneficiary	Total eligible expenditure to the beneficiary (allocated, paid)
	Public eligible expenditure to the beneficiary (allocated, paid)
	EU contribution to the beneficiary (allocated, paid)

DB Operations	
1. General information	Currency and conversion rate, cut-off date
2. Programme identification	CCI number
	[...]
	Official operation Identification Code
3. Operation identification	Official and ad-hoc operation code
	Operation name and description
	OP Priority Axis
	OP action/measure
	[...]
	Operation type (from cluster analysis)
4. Operation's characteristics	Thematic Objective, Investment priority, Field of Intervention
	Form of finance
	Location (NUTS)
	Territory type
	Economic activity
	PPP, Major Project, Joint Action Plan
	N. of beneficiaries
	[...]
5. Operation's financial information	Total project cost
	Total eligible expenditure (allocated and paid)
	Public eligible expenditure (allocated and paid)
	EU contribution (allocated and paid)

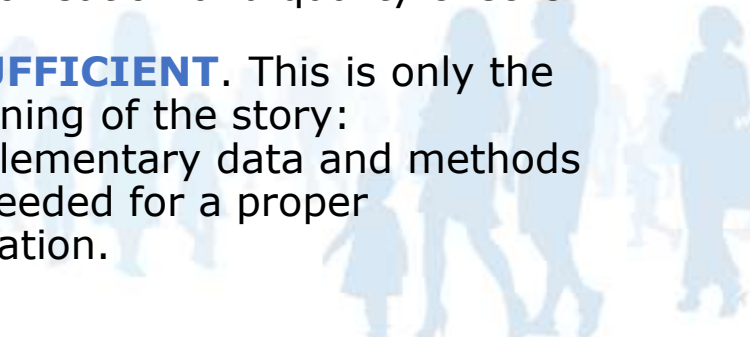
DB Output indicators	
1. General information	[...]
2. Programme and operation identification	CCI number
	Official operation Identification Code
	Official and ad-hoc operation code
	Operation name
	OP action/measure
	Operation type
	[...]
3. Output indicator identification and background information	Indicator unique identifier and title
	Description, unit of measure
	Target, achieved value
	Financial information
	[...]

Lessons learned

Opportunities

- **ACCESSIBLE.** Rich micro data is available in the Managing Authorities' monitoring systems.
- **USEFUL.** These data can be used for insightful comparative analysis.
- **SCALABLE.** Their potential can be fully exploited through matching with external data sources (e.g. Cordis data on H2020 projects, Orbis company database, etc.).

Challenges

- **IMPERFECT.** Data collected more for auditing and administrative monitoring, than for evaluation.
 - **HETEROGENEOUS.** Significant efforts for data cleaning, harmonisation and quality checks.
 - **INSUFFICIENT.** This is only the beginning of the story: complementary data and methods are needed for a proper evaluation.
- 

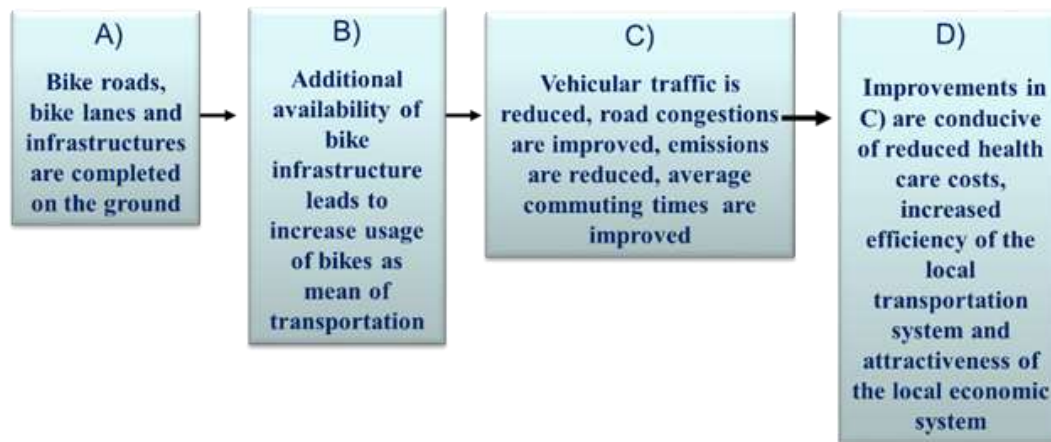
Questions / discussion

- **Questions from participants?**
- **Links between data sources and evaluation methods?**
- **Issues to be addressed to improve data acquisition and its use in impact evaluations?**



Daniele BONDONIO

- Intervention logic / programme theory should be the foundation for data acquisitions and evaluation methods to be used in quantitative analyses
- chains of causal links, enables the sorting out of the different data needs and methods for the analysis

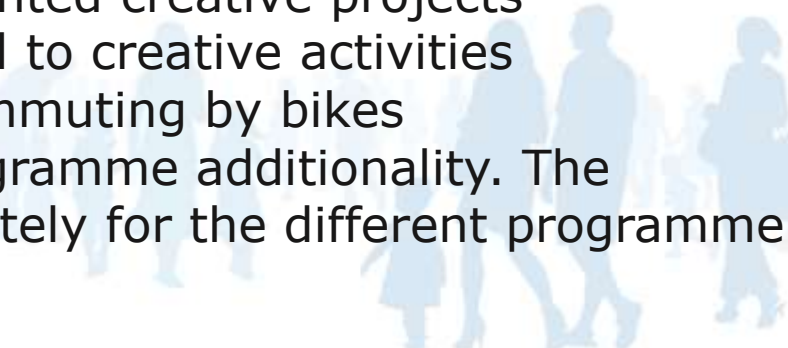


- **Indicators A)** are considered in monitoring tasks and they do not display any “spontaneous change” due to independent factors: simple before-after comparisons or benchmark comparisons are acceptable
- For **Indicators B-D)**, instead, before-after or benchmark analysis is not reliable (spontaneous change can occur). They can be divided into three categories:

I) Proximate/local result indicators E.g.

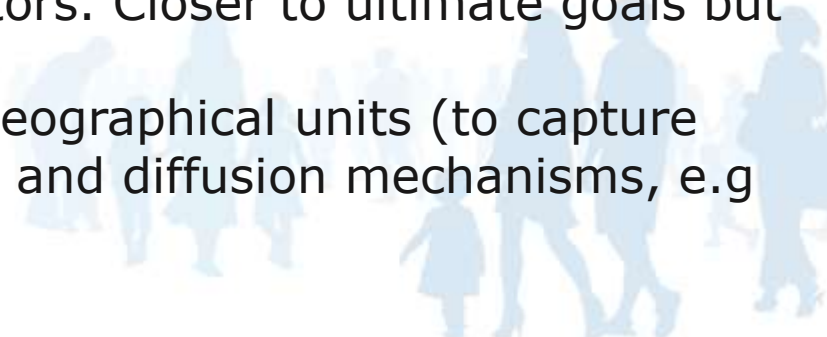
- investment tied to implemented creative projects
- n. of start-up in sectors tied to creative activities
- percentage of residents commuting by bikes

Are best suitable to estimate programme additionality. The analysis can be performed separately for the different programme interventions



- II)** Distal/global result indicators E.g.
- number of jobs in the local economy;
 - exports and innovation outcomes;
 - air quality indicators;

It is necessary to analyze jointly all programmes. The analysis aims to comparatively assess the causal impact of the different interventions (e.g. Bondonio and Martini 2019 and Bondonio and Greenbaum 2014)

- III)** Local economy/community indicators. Closer to ultimate goals but the analysis is more challenging.
Units of observation in terms of geographical units (to capture negative/positive spillover effects and diffusion mechanisms, e.g. Bondonio and Greenbaum 2018)
- 

Audronė SADAUSKAITĖ

Issues to be addressed to improve data acquisition and use in impact assessments:

- **GDPR:** Authorities should set up appropriate processes around data from beneficiaries so that the data can be transferred to evaluators.
- **Complexity of eCohesion monitoring systems:** Evaluators have to define a (precise) list of indicators to request from eCohesion monitoring systems. The systems can be quite technocratic, not always optimized (e.g. several indicators measuring the same thing).
- **Data on financial instruments:** Data on final beneficiaries of financial instruments is particularly difficult to obtain (since the beneficiary/applicant is the financial institution which receives EU funding). With the increasing push for more use of financial instruments, accessibility of this data for evaluation purposes will become increasingly crucial.


More questions

- **How to address capacity issues and reduce barriers to access and use of data?**
- **How to integrate more complex quantitative / data hungry techniques into national evaluations?**
- **How to incorporate the needs of impact evaluations into the policy design and monitoring systems?**



Audronė SADAUSKAITĖ

How to address capacity issues?

- **Contracting Authorities** should have a basic understanding of the variety and added-value of the more complex quantitative evaluation techniques.
 - Sometimes resources dedicated to an evaluation does not permit for use of complex techniques even if they are feasible (data is available).
 - Strengthening evaluation capacities of Authorities is key:
 - Methodological guidance to Contracting Authorities
 - Trainings to public servants
 - Conferences for exchange of good practices (better use of evaluation results and take-up of innovations in evaluation)
- 



Audronė SADAUSKAITĖ


How to reduce barriers to access and use of data?

- **Timing of evaluations. There is a paradox between:**

- 1) need for **timely** evaluation results before starting programming of the new period and
 - 2) more **robust** evaluations with more and better quality data
- Issue with 1): evaluating measures that are in the middle of their implementation (projects not ended = small samples in small countries such as LT, no mid- and long-term impacts can be measured).
 - Issue with 2): limited integration of lessons learned when programming new measures.
 - **Key implication: timely implementation of measures is a prerequisite for timely (1) and robust (2) evaluation results.**

ADDITIONAL MATERIAL: DANIELE BONDONIO

Issues to be addressed to improve data acquisition and its use in impact evaluations?

- Interviews and questionnaires can be also used as tools to acquire crucial data for quantitative analyses (including counterfactual approaches).
E.g. data on innovation outputs, innovation collaboration networks, quality of life are often missing.
 - Key issue is how to incentivise better response rates: institutional options vary. Sometimes is possible to include in calls for applications the requirement to answer to questionnaires/interviews? (Sometime applicable also to non-supported units)
- 



ADDITIONAL MATERIAL: DANIELE BONDONIO

References:

- The impact of support to innovation projects on job quality (the study pioneers the linkage between 3 data sources: firm-level administrative records, employee-level data, program-activity data):

Downloadable WP: https://repositorio.iscteul.pt/bitstream/10071/11932/1/DINAMIA_WP_2016-01.pdf

Citation: Bondonio D, Farinha T and Mamede R, Does EU Enterprise Support Boost High Quality Jobs? Evidence from Linked Employer-Employee Microdata and Natural-Experiment Conditions, Evaluation Review (under final review)

- Impact evaluations to comparatively assess the causal impact of the different programme interventions on different global/distant result indicators:
 - Bondonio D. and Martini A. (2019) "Are they worth it? A Counterfactual Impact Evaluation of a Decade of Investment Subsidies to Italian Firms", in Italian Journal of Regional Science, ISSN1720-3929, n. 2/2019.
 - Bondonio D. and Greenbaum R. (2014) "Revitalizing regional economies through enterprise support policies: An impact evaluation of multiple instruments", in European Urban and Regional Studies (Impact Factor 2.07, Group A anvr), vol.21 n.1, pp.79-103, 2014.



ADDITIONAL MATERIAL: DANIELE BONDONIO

- Impact evaluations to comparatively assess the causal impact of the different programme interventions on different global/distant result indicators:
 - Bondonio D. and Martini A. (2019) "Are they worth it? A Counterfactual Impact Evaluation of a Decade of Investment Subsidies to Italian Firms", in Italian Journal of Regional Science, ISSN1720-3929, n. 2/2019.
 - Bondonio D. and Greenbaum R. (2014) "Revitalizing regional economies through enterprise support policies: An impact evaluation of multiple instruments", in European Urban and Regional Studies, vol.21 n.1, pp.79-103, 2014.
- Impact evaluation with aggregated data at the level of local economy/community in which the programme intervention is available (units of observation for the analysis are in terms of geographical units in order to capture negative/positive spillover effects, diffusion mechanisms etc.)
 - Bondonio D. and Greenbaum R. (2018) "Natural Disasters and Relief Assistance: Empirical Evidence on the Resilience of U.S. Counties using Dynamic Propensity Score Matching", in Journal of Regional Science, vol. 58 n. 3, pp. 659-680, 2018.

