Summer School 2025
Pitfalls in the evaluation process and how to avoid them

European Centre of Solidarity
Gdańsk, Poland
(2-4 July 2025)

APPLICA ISMERI EUROPA QUADRANT CONSEIL

Organisation and Rules of the Game

- Topic: Pitfalls in the evaluation process and how to avoid them
  - Block 1 - Pitfalls in the design stage
  - Block 2 - Pitfalls at the implementation stage
  - Block 3 - Pitfalls at the delivery stage
  - Masterclass - Experimental and quasi-experimental methods

- Organisation
  - Plenary sessions introducing general framework on each block
  - Working group sessions – *participants in person only*
  - Panel discussions followed by general discussion

- Working groups
  - 5 Working groups (1-5) - 8 participants in each
  - 5 Working group sessions
  - Work on scenarios and discussion of set of questions
  - Training personnel 'circulating' between groups
  - Panel discussions on working group outcome in plenary sessions – 1 participant from each group
- Online participants
  - Participation in plenary sessions and ‘moderated’ involvement
  - No online working group sessions but zoom chat remains open during working group to allow participants to communicate and network between them

Introduction

A summer school about pitfalls

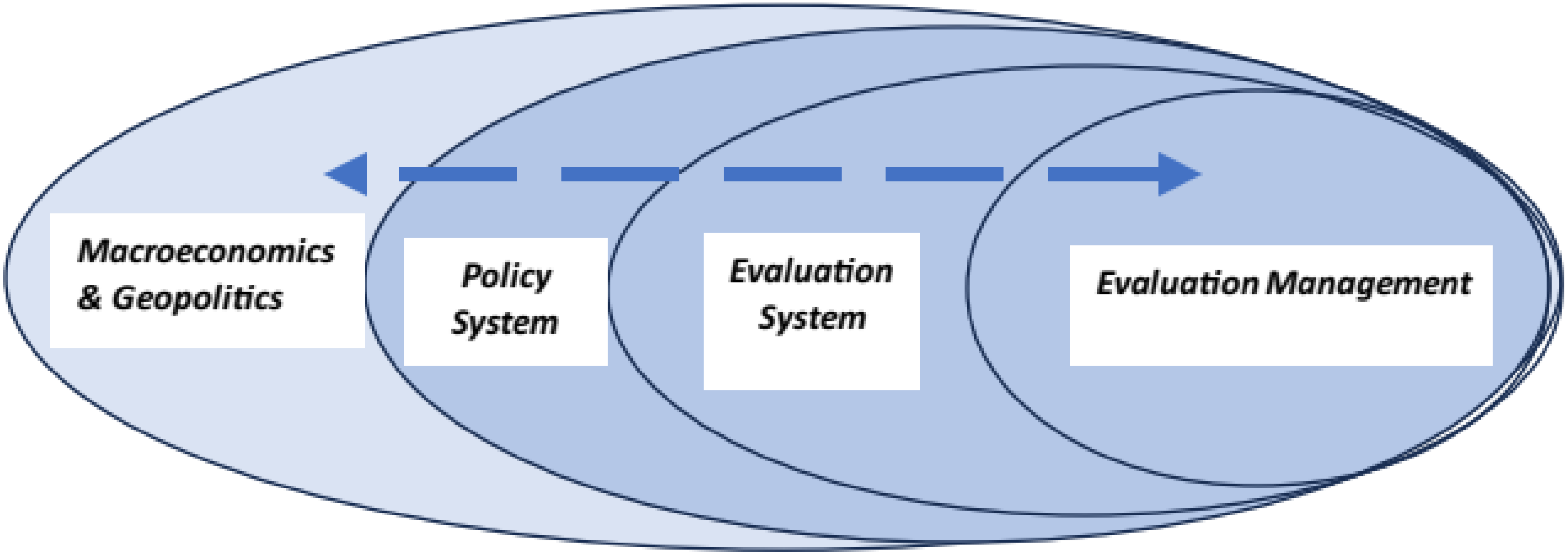
- This summer school is about pitfalls – problems that have the potential to trip us up, fall down a hole in the road – prevent us doing our jobs
- All management involves problem solving, because things always go wrong
- Evaluation management involves a lot of falling down holes in the road and digging ourselves out!
- ….Especially when we work across organisational boundaries ***or*** in unpredictable policy spaces, ***or*** where programmes include different kinds of stakeholder – from public administration, the private sector and civil society; ***or*** in evaluation systems with too few contractors who understand the methods we want to use

- To cope with problems we draw up Evaluation Plans, train specialist staff and develop ‘quality management’ processes which can increase our ability to solve problems, even though pitfalls never completely disappear

- Some problems in evaluation are:
  - routine – *where will I get data!*
  - … more serious, although still knowable, where a complete management process needs to be rethought – *procurement rules prevent me contracting evaluators with the right skills*
  - both serious and unexpected requiring a complete rethink of what we’re doing – like *financial crises, fires and floods, wars in Europe and Covid-19*

- Evaluation is never self contained

- Rather, it is embedded in wider systems:
  - Administrative systems with rules and resources
  - Evaluation systems with knowhow, networks & institutions
  - Policy systems with priorities and interests – that can change while we are evaluating!
  - Geopolitical systems that bring us wars, climate change and pandemics


Macroeconomics & Geopolitics
Policy System
Evaluation System
Evaluation Management

- The word pitfall implies falling down a hole in the road unexpectedly!

- We want to discuss how to make the road a little more predictable – we will be discussing not only:
  - How to *respond* to problems
  - But also how to *identify* and *anticipate* problems and
  - Where possible How to *prevent* problems

- Although we are dividing the Summer School programme into stages of the evaluation cycle these stages are never completely separate
- A badly designed programme is usually associated with a bad evaluation
- A Call for Tender that is unclear about methodology will often lead to contracting the evaluators with the wrong skills
- Evaluators with the wrong skills will not be able to deliver on what they promised in their proposals
- And an evaluation that fails to consult stakeholders is unlikely to focus on their needs and provide results that they can use

How to identify or anticipate pitfalls

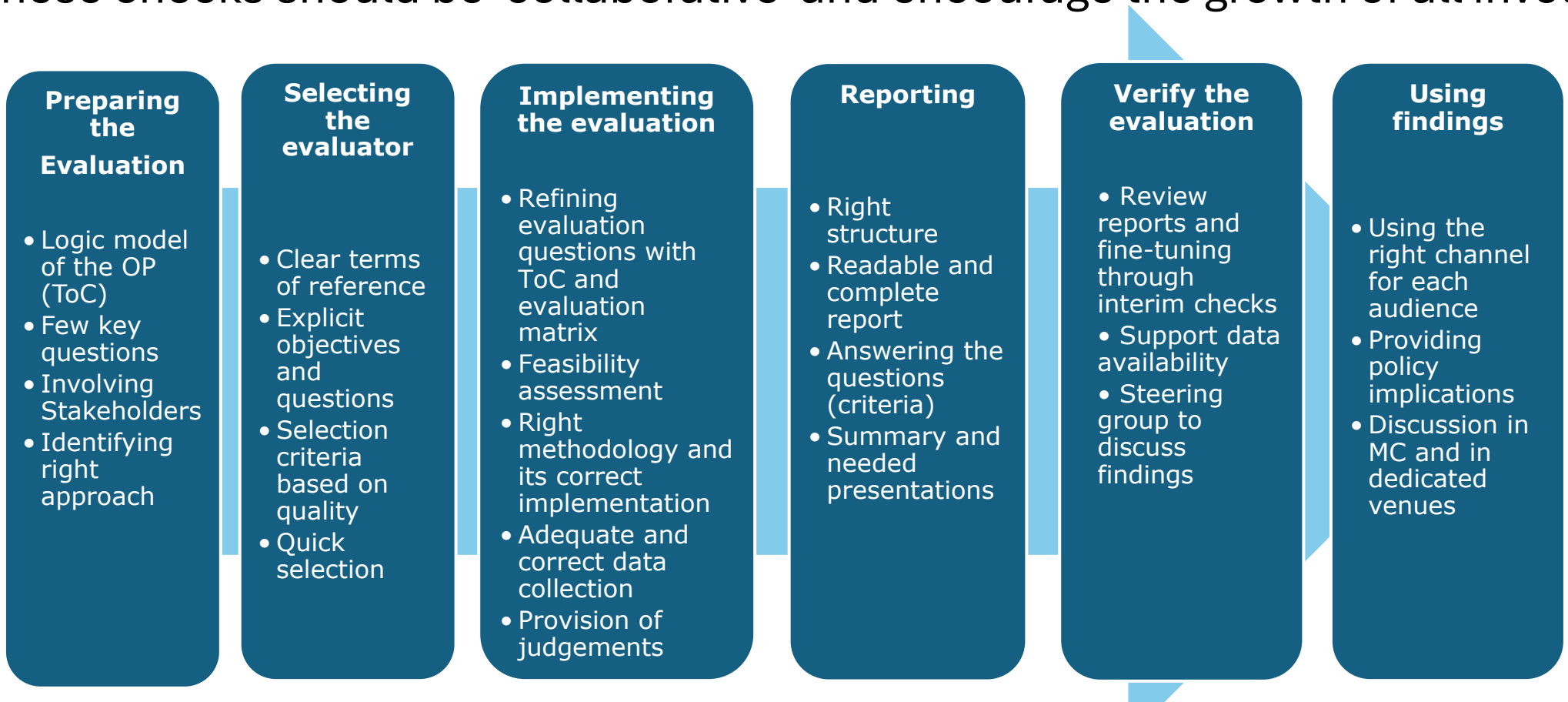
- Identifying pitfalls enables us to react to them.
- However, identifying pitfalls in time is not always simple, as their effects often only become fully apparent at the end of the evaluation.
- The awareness of the **main risks** and a **quality assurance approach** serve this purpose, but they are not always present and may require specific expertise.
- We suggest focusing on the following:
  - The **actors** involved in the process;
  - The **time** at which key checks should be carried out;
  - The necessary **competencies and tools**.

The actors

- Each actor plays a specific role and has different responsibilities, capabilities and interests.
- The first way to anticipate pitfalls is to verify that each actor is doing the right thing.
  - Has the commissioner prepared the evaluation well (ToC, EQs, involvement of the stakeholders, clear type of evaluation, etc.)?
  - Has the commissioner selected the best evaluator according to the right criteria? (if not possible, flexibility and adaptation were planned)
  - Has the implementation of the evaluation been checked?
  - The evaluator has properly designed the methodology, collected the right data, produced well-written reports that answer the EQs, and generated usable findings.
  - Have the proper stakeholders been involved in the key stages and demonstrated the right commitment to follow the evaluation process?

Checks along the way

- A second way to anticipate or recognise pitfalls is to plan specific checks throughout the evaluation process, which is directly under your responsibility.
- These checks should be 'collaborative' and encourage the growth of all involved.



ADRANT CONSEIL

Having competences and tools to anticipate pitfalls

- The Commissioner must have the **basic competencies** to oversee the evaluation process. Several possible strategies: hiring, training, out-sourcing, collaborating with experts and academics, including independent experts in the evaluation teams; etc.
- **Data and methodology** are often the most technical aspects to strengthen.
- Do not underestimate the importance of "**evaluative thinking**" (questioning assumptions, focusing on causalities, using evidence, etc.) which forms the basis of any good evaluation.

Some tools can help

- Having a basic set of tools helps the quality check and the reflection on the evaluation process
- Many **guidelines and toolboxes** are available. Sometimes are schematic and must be contextualised, but are useful
- Among the tools it is possible to remind:
  - Use the **theory of change** to identify the key components of the programme to be evaluated.
  - An **evaluation matrix** is used to verify the coherence between the evaluation questions, the available data, and the methodologies.
  - **Checklists** ensure that reports do not omit or neglect important elements.

- Cutting across these different aims and activities – to identify, anticipate, remediate and prevent - are quite different orientations
- Challenge for the public service: Moving beyond Compliance Management

Safety Officer

- Closest to traditional public administration is compliance management – ensuring that checklists and Quality Control guidelines are adhered to

Navigator

- Anticipation & identification of 'pitfalls' before you fall down the 'pit' requires a good map, strong navigational skills - and the autonomy to take avoiding action. Some call this Quality Assurance, but it is closer to inspectorate …

Crisis manager

- Once you've fallen down the 'pit' guidelines and toolkits rarely help you! You need a crisis manager or at least an escape artist!

Pitfalls in the design stage

Problems and Pitfalls – Design

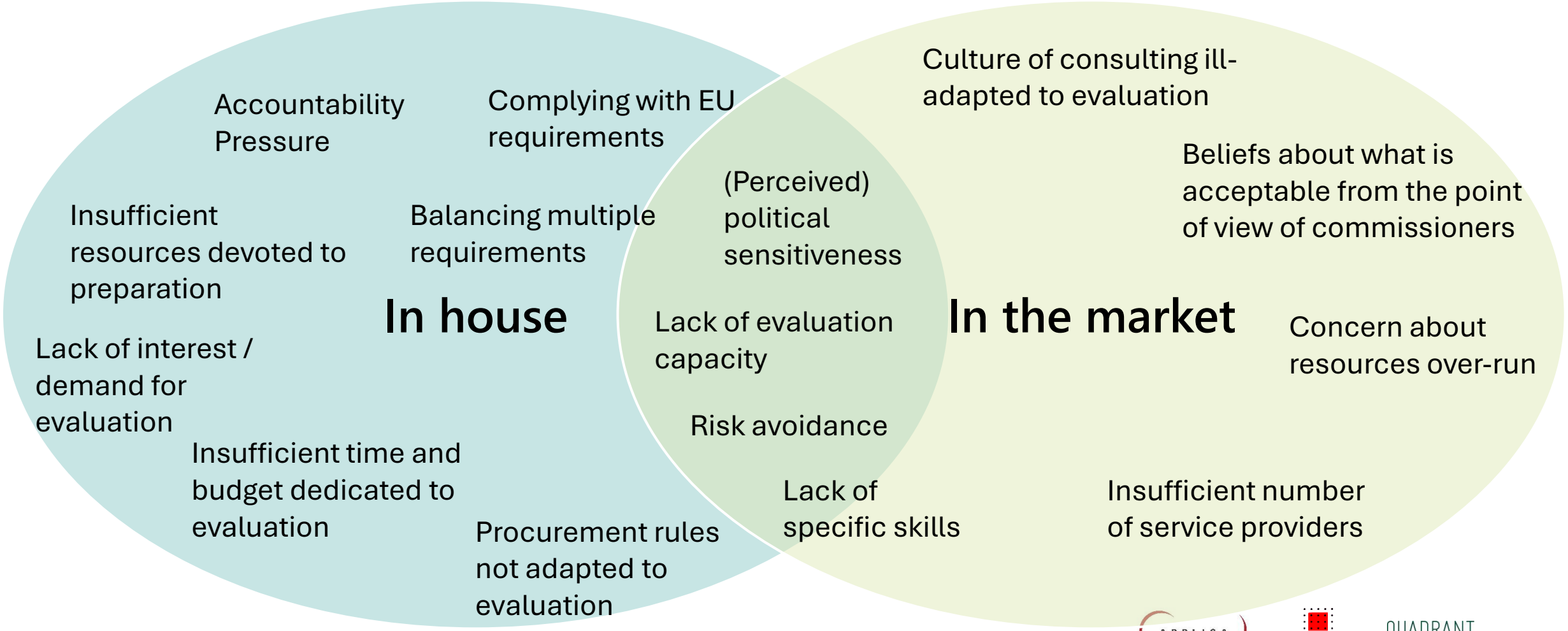
- “Stakeholders have different understandings of evaluation – some formative/to steer programmes; some summative to assess impacts”
- “Overloaded Tor and EQs – every administrative unit want their questions answered”
- “Poor formulation of EQs often disconnected from ToC”
- “Asking EQs we already know answers to”
- “How far should detailed methods be specified”
- “Identifying right questions to match evaluation criteria”
- “Insistence by procurement on detailed, rigid contracts”
- “Choosing good contractors”
- “Long elapse time between drafting Tor and implementing evaluations – Context and evaluation focus can change”

APPLICA ISMERI EUROPA QUADRANT CONSEIL

Common pitfalls of the commissioning process

- ToRs are written (and bidders write a proposal) while the evaluation scope, information needs, and potential uses, **are still mostly unknown**
- Contractual arrangements are made at the beginning of the evaluation on methods, agenda, team… **but may be obsolete early on**
- The selected proposal may not be adapted to the actual needs that appear **after the data collection has started**
- Some important aspects (e.g. quality of collaboration) **cannot be guaranteed through ToRs**
- It is difficult to address quality issues when the choice of service provider was intrinsically inadequate

Other common challenges when commissioning evaluations…



APPLICA ISMERI EUROPA QUADRANT CONSEIL

Some ways to avoid pitfalls

- Before ToRs
  - Find users and uses and translate their needs into evaluation questions
  - Reconstruct intervention logic / draft ToC and find potential bottlenecks
  - Audit available data / Carry on an evaluability study
  - Engage stakeholders in evaluation process early on
- During procurement
  - Write ToRs to attract the type of service providers you want to have
  - Can they do it? Do they have an added value? Is it interesting for them?
  - Include audition / negotiation in the selection process if you can
- Leave possibility to adapt in inception phase

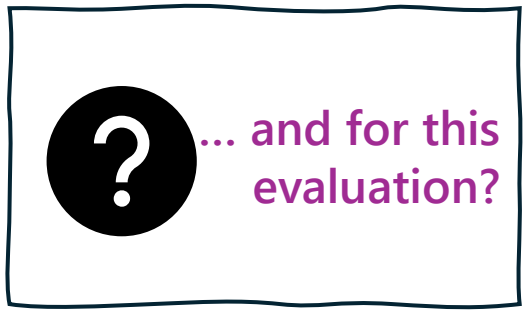
What skills should you expect from your service provider?

Cohesion policy / SF Programmes

Public institutions (who does what, how they work, etc.)

Evaluation / (social) science methods

AI proficiency



Collaboration

Sectoral knowledge (trends, challenges, policies, actors…)

Project Management (multiple actors, topics, streams…)

Policy making (Decision-making, knowledge brokering…)

Others: knowledge of policies in other countries…, disciplinary knowledge? Training in ethics?…

APPLICA ISMERI EUROPA QUADRANT CONSEIL

Flex

- Focused on uses / outcomes
  - Requires good investigation of potential uses
- Good for new topics, tricky results to achieve
- May bring unexpected added value
- Need for capacity on the market

No Flex

- Focused on list of methods
  - Requires initial investment in methodological simulation
- Good for conventional, simple evaluations
- Easier to assess proposals
- Need for internal capacity

⚠️ May both prove inadequate in the end…

What you CAN do

- Commission an **evaluability study** to identify potential scenarios of evaluation
- Ask for **options**
- Include an **audition** in the bidding process / Plan for an **exploratory phase** at the beginning of the evaluation to clarify the scope and use and refine the approach only then
- **Break down** the evaluation process in several steps e.g. through purchase orders

What you should NOT do

- Prescribe approaches or methods if you cannot justify them
- Ask for creativity without clear expected outcomes and a budget range
- Copy-paste ToRs…
  - But you can use them as inspiration!

AI in the evaluation process

- Do you use AI in your work? To do what?

- Do your contractors use AI in their work? To do what?

What are we talking about?

Big Data analysis	• Uses machine learning and data modelling • Generates knowledge from structured and unstructured data sets
Automation of M&E systems	• Standardises data collection and data analysis collected along ToC
Deep Learning	• Meta-analysis of structured knowledge, e.g. information • Generates new knowledge about what works or not
Large Language Models (LLMs)	• Commercial tools used in everyday professional practice: chatbots such as ChatGPT, AI translators or text editors, multimedia publishers…

AI use: challenges and opportunities

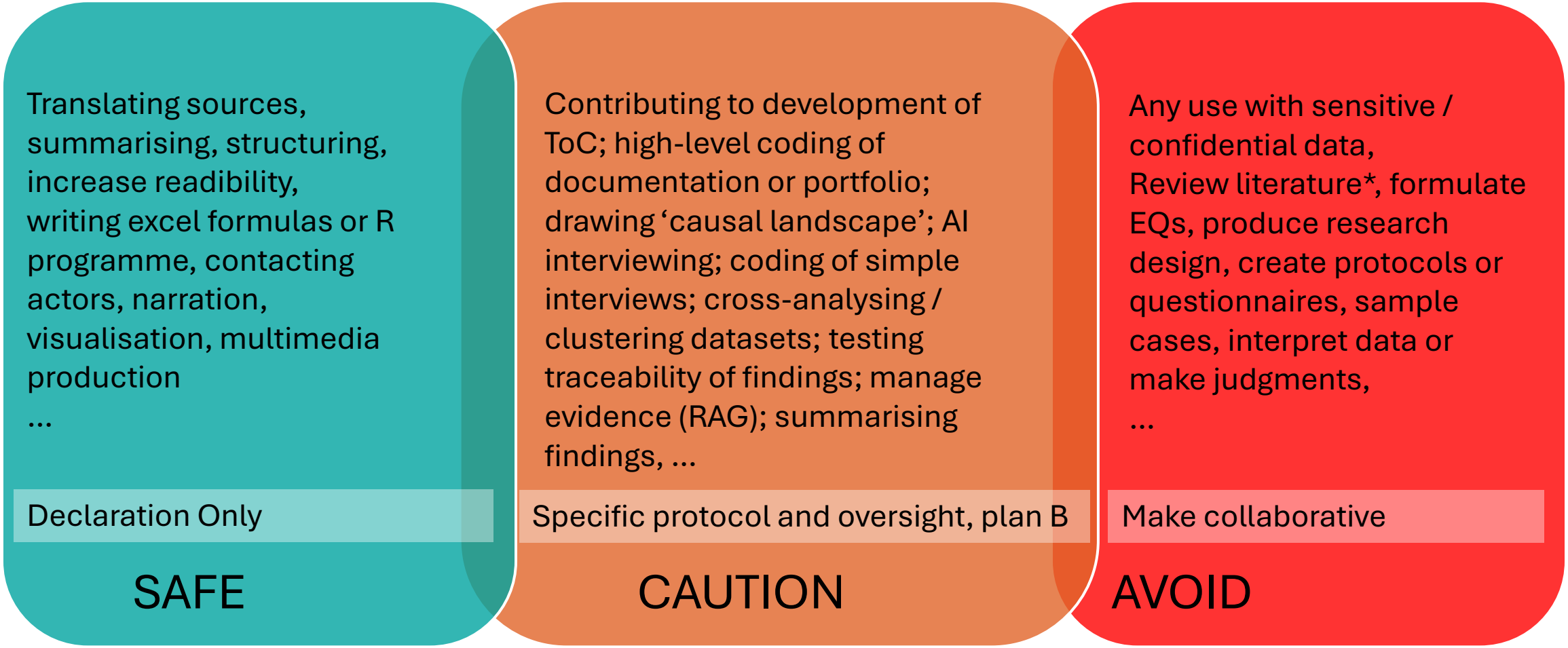
- An LLM produces contextually relevant content… without reference to concepts such as ‘truth’ or accuracy.
- It cannot guarantee the validity of its results
- It is designed to reflect user inputs (=your worldview) and the data it has been trained on (=North-American & European, in English)

- Yet it is being used more and more for many good reasons… by everyone
  - How to ensure relevance of use and validity of results?
  - How to maintain trust in your work and your contractors’ works?

Guiding principles for working with an AI

1. Only use an LLM for production **if you can control** the quality of the process and results
2. For simple tasks, **choose more robust & auditable solutions** (e.g. Excel)
3. Cut complex tasks into **smaller ones** and identify which can be performed by AI, and which by humans
4. Be very **careful about using AI** with private or sensitive data or with vulnerable groups
5. Be aware of **cultural gaps** in input (what a U.S. AI can't grasp) and output (how awkward?)
6. Develop **AI skills**, use AI for tasks for which you or colleagues are skilled
7. Use AI process for **reflexivity** about questions asked and answers received
8. All uses of AI should be **disclosed**

Agree on uses and risks in ToR / inception phase



Some potential uses of AI at design stage

- Before ToRs:
  - Reconstruct the intervention logic, feed a provisional theory of change
  - Explore portfolio of projects or profiles of beneficiaries
  - Consult evaluations on same topic in other languages

- In ToRs:
  - Improve ToR writing – make them more enticing to the service providers you want to attract
  - Simulate responses to questions / to ToRs

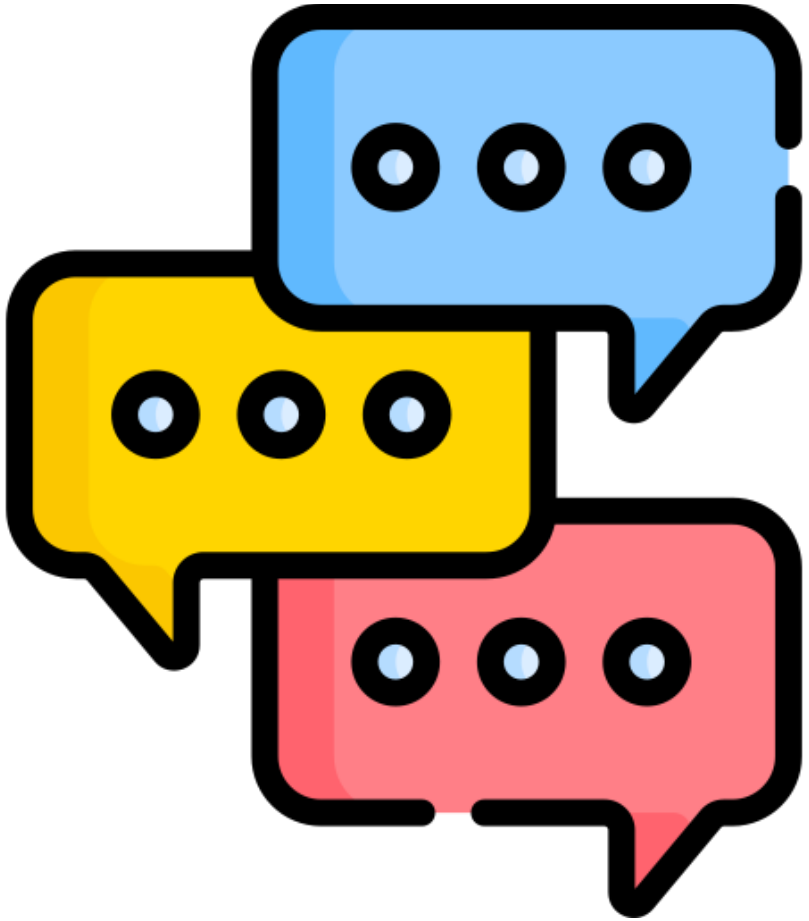
- What should be done with caution or avoided?

Working groups: the scenario

- « You are finalising the ToR of an evaluation of SuCIP, a new ‘Supply Chain Innovation Programme.’ »

WG1

- In relation to each problem identify:
  - Remedies in the short term that will allow you to launch this evaluation as soon as possible
  - Anticipation and Prevention measures that would have avoided the problems now facing the SuCIP evaluation
  - Evaluation capacity improvements so that evaluations of other programmes will not face similar problems in future
- When considering the above think about:
  - What the Evaluation Function can do on its own
  - What Evaluation Function can do with wider MA support
  - What requires collaboration with wider administrative system
  - What requires collaboration with stakeholders



Panel discussion

Presentation of the community of practices

Carlo Amati and Tamina Huber (online)

Thomas Delahais

Propriétaire

Groupe créé : mars 2024

Posts en attente 0

Demandes d'adhésion 1

Gérer le groupe

Modifier le groupe

Récent

- Evaluation Helpdesk - Cohe...
- METHEVAL - Make to learn: ...
- Communauté F3E de consul...
- Monitoring and Evaluation P...
- Causal Pathways

Groupes

- Evaluation Helpdesk - Cohe...
- Communauté F3E de consul...
- Monitoring and Evaluation P...

Voir plus

Evaluation Helpdesk - Cohesion Policy

Privé répertorié

Obtenez un badge de groupe actif

Commencer un post dans ce groupe

Vidéo Photo Sondage

All Recommandé

Épinglé par l'administrateur Ne plus épingler

Thomas Delahais a publié sur Evaluation Helpdesk - Cohesion Policy

Thomas Delahais • Vous
Évaluateur, spécialiste de l'évaluation d'impact, Quadrant Conseil
Administrateur • Administrateur • 1 an(s) •

15 membres

Y compris Virginie Besrest et 6 autres relations

7

Inviter des relations

Tout afficher →

0 ▼100%
Membres actifs

0
Nouveaux membres

0
Posts

0 ▼100%
Vues des posts

Tout voir →

Messagerie

Evaluation Helpdesk 2021-2027

Evaluation Helpdesk on Cohesion Policy https://www.linkedin.com/groups/12999148/

Pitfalls in the implementation stage

- Designing methodological approaches

Frequent pitfalls in using methodologies

- The methodology does not allow to properly answer the EQs
- Data for the chosen methods are not available
- Data collection and used information are redundant and the analysis is efficient
- Available resources are not sufficient to use the chosen methods
- The assumptions at the basis of the method do not fit the intervention to evaluate
- Analyses are not able to examine the causalities triggered by the intervention
- The combination of quantitative and qualitative methods is confused and not integrated

Two main types of error

- **Designing the wrong methodology**
  - Feasibility (in terms of time, data availability, etc.) not verified in advance.
  - It is not designed according to the evaluation questions and the characteristics of the programme.
  - Lack of mapping of causalities to verify (ToC or similar).
  - Not pertinent to the necessary causal inference.

- **Applying the methodology incorrectly**
  - Not enough time or attention is dedicated to data collection and analysis.
  - Insufficient evaluator skill.
  - The methodology does not allow for significant generalisation.

Frequent pitfalls in quasi-experimental methods

- Quasi-experimental methods require a **large amount of data**, such as a control group and a high number of cases with detailed variables, which are not always available.
- These methods rely on **significant assumptions**, such as the parallel trend assumption in DiD or the assumption that relevant variables are continuous at the point where treatment and outcome discontinuities occur in RDD, which are not always realistic
- Although **other treatments** (e.g. national policies) are important, they are not considered.
- **Outcomes** are not well defined (e.g. productivity growth for R&D grants) due to a lack of preliminary analyses and theory of change (ToC).

Frequent pitfalls in using a theory

- The programme theory is **not well-defined** (it is too vague or too simple, has not been shared with key actors, adopts an IL which is inadequate).
- The **assumptions** required to determine the causal effects have not been defined or analysed.
- The **data collection and case studies** have not been defined in accordance with the theory and the subsequent analysis are not well focused.
- The influence of **external factors** is not considered.
- There is **no triangulation** between data or information from different sources.

Frequent pitfalls in using mixed-methods

- Insufficient and **unclear design** (is it convergent, explanatory, exploratory, two-stage, embedded or complementary?).
- The design is **too complex or redundant** in relation to the evaluation questions.
- Inadequate **data collection** does not allow the entire picture to be reconstructed.
- The methods are **not well integrated** and produce confused findings.
- There is **disproportionate reliance** on only one type of data.
- **Contextual factors** affecting different types of data are ignored.
- The analytical methods and final analysis are **unable to integrate and synthesise** different data.

Frequent pitfalls in using case studies

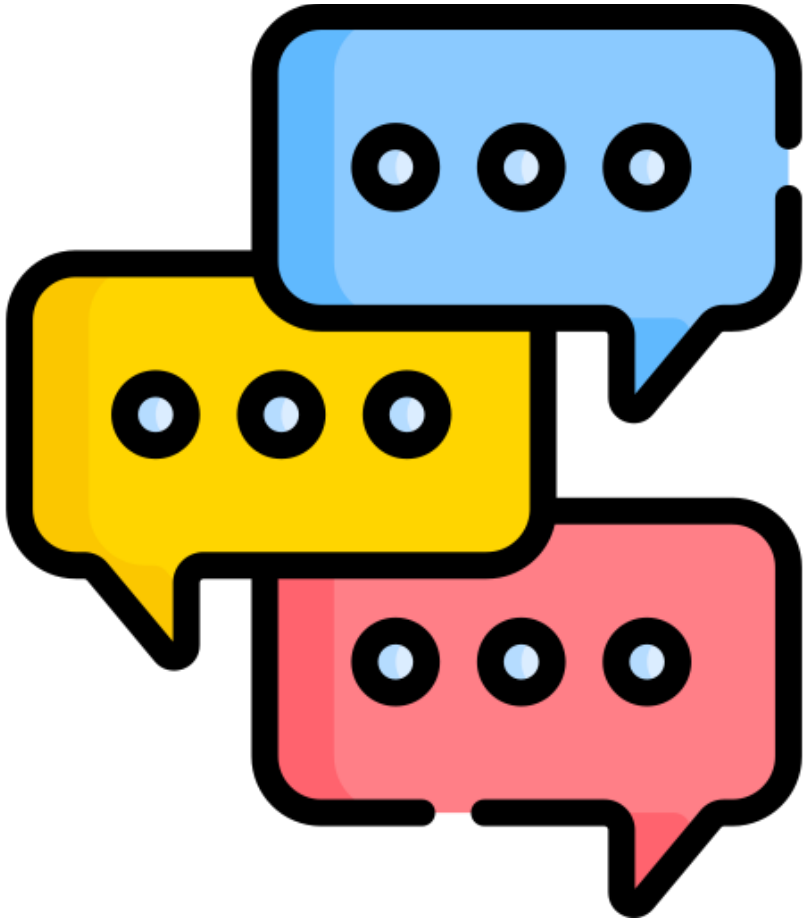
- The **selection of case studies** does not align with the objectives of the evaluation (e.g. exploratory, explanatory or descriptive), which weakens the entire approach.
- The cases **do not focus sufficiently on the evaluation questions** or assumptions of the ToC and remain overly descriptive.
- The **data collection is biased** (e.g. concentrated on a specific type of actor), resulting in biased findings that favour one perspective only.
- The **contextual factors** in the cases are not examined adequately, and the results fail to explain whether the policy was successful or not.
- The **comparison between cases** is inadequate and the generalisation of the results that follows is unconvincing.

How to detect or anticipate pitfalls in methodology

- **Clearly presenting** the evaluation questions and the aims of the evaluation.
- Making **existing data available** to prepare methodological proposals.
- **Selecting a good evaluator** who proposes a good methodological approach.
- Carry out **a preliminary or feasibility** study before launching a complex evaluation.
- **Use experts**, when useful, to support the assessment of evaluation reports and discuss methodological choices.
- Discuss with **implementers and stakeholders** whether the possible results are aligned with their needs.
- Carefully assess the **inception and interim reports** to identify areas for adjustment and improvement.

WG2

- Parallel Working Group 2– Methodological pitfalls
  - What do you think is the best approach to conduct ‘strong’ impact evaluations and Value for Money assessments for a programme like SuCIP – and what do you think are the likely pitfalls?
  - How far would you prioritise a formative/real time evaluation dimension and how could this be undertaken?
  - How far do you think the evaluation team has the skills to tightly specify the methodology for this evaluation or should you ‘leave it to the market’ or contract an external expert to conduct a design study?
  - What other ‘pitfalls’ would you expect when choosing the methods for an evaluation such as this – and how could these be avoided



Panel discussion

Some potential uses of AI in the design of the approach

- Broaden range of data sources and methods that can be used:
  - Text mining / Coding of reports, forms, etc.
  - Network analysis
  - Clustering, modelling
  - Identification of conditions, risks, enablers (that can be later investigated)

  - ⚠️ Use AI to write code, avoid using LLM directly

- What should be done with caution or avoided?

Wrap-up