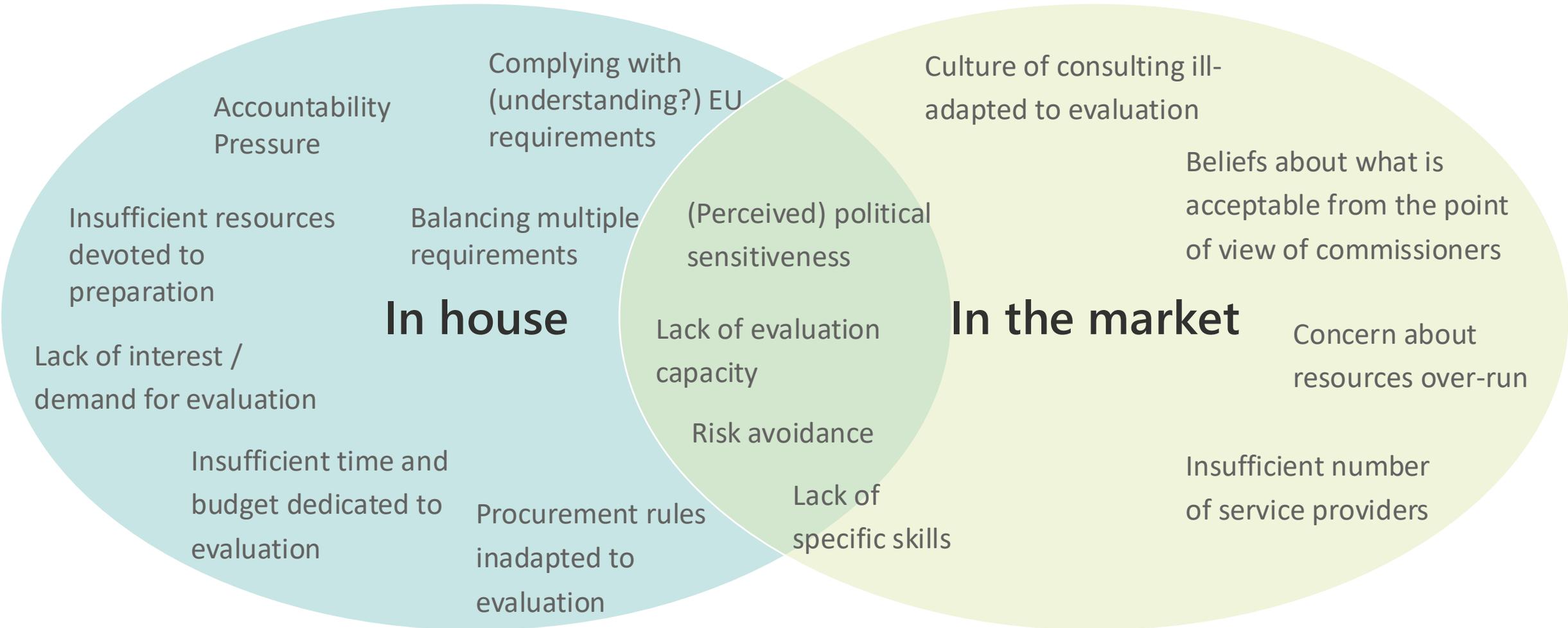**09:30 – 09:50**

Introduction

Introduction by Greek authorities, DG REGIO (Enrico Pesaresi) and DG EMPL (Jorge Villanueva Garcia, Johannes Ziemendorff)

09:50 – 11:00
Plenary

Commissioning the evaluation. Challenges in procurement. Building on existing knowledge. The most important questions your ToR should answer. Anticipating evaluation use.

Common challenges when commissioning evaluations...

In house

In the market

Accountability Pressure

Complying with (understanding?) EU requirements

Insufficient resources devoted to preparation

Balancing multiple requirements

Lack of interest / demand for evaluation

Insufficient time and budget dedicated to evaluation

Procurement rules inadapted to evaluation

(Perceived) political sensitiveness

Lack of evaluation capacity

Risk avoidance

Lack of specific skills

Culture of consulting ill-adapted to evaluation

Beliefs about what is acceptable from the point of view of commissioners

Concern about resources over-run

Insufficient number of service providers

The 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**

For consultants, ToRs are a guessing game. What do you think they are looking for in ToRs?
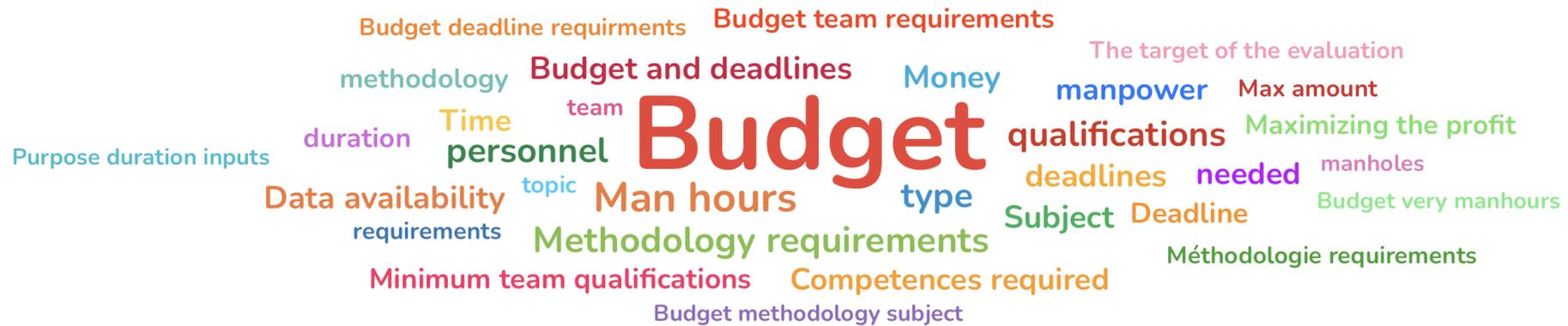
Go to wooclap.com and use code KOCZUC

What do you think consultants look at first in your terms of reference?

Budget deadline requirments

Budget team requirements

The target of the evaluation

methodology

Budget and deadlines

Money

manpower

Max amount

team

Time

duration

personnel

Budget

qualifications

Maximizing the profit

Purpose duration inputs

deadlines

needed

manholes

Data availability

topic

Man hours

type

Subject

Deadline

Budget very manhours

requirements

Methodology requirements

Méthodologie requirements

Minimum team qualifications

Competences required

Budget methodology subject

Definitions…

- Commissioning is the process through which evaluation needs are turned into evaluation services

- It involves:
  - identifying and clarifying those needs
  - considering how they could be answered through evaluation services
  - specifying a series of administrative and technical arrangements through which these services can be delivered
  - assessing different proposals – and purchasing services

- Good commissioning entails preparatory (planning) and follow-up work (quality assurance)

Main advice

- ToRs should be short! 5 to 10 pages is enough. Detailed information can go in the appendix

- ToRs should be thought with their users in mind, **i.e. service providers**. What do THEY need to know:
  - To decide to answer (do they have a chance to win that bid?)?
  - To make a relevant proposal?

- ToRs are a process:
  - Tell users in advance (evaluation plan for year X, PIN),
  - Use questions to clarify ToRs,
  - Offer possibility to counter-argument

The 6 W's of Commissioning Evaluations

Why are we evaluating? For what uses?

Who will be involved?

**hoW** much resources will be invested?

What is to be evaluated?

When are deliverables expected?

hoW should the evaluation be performed?
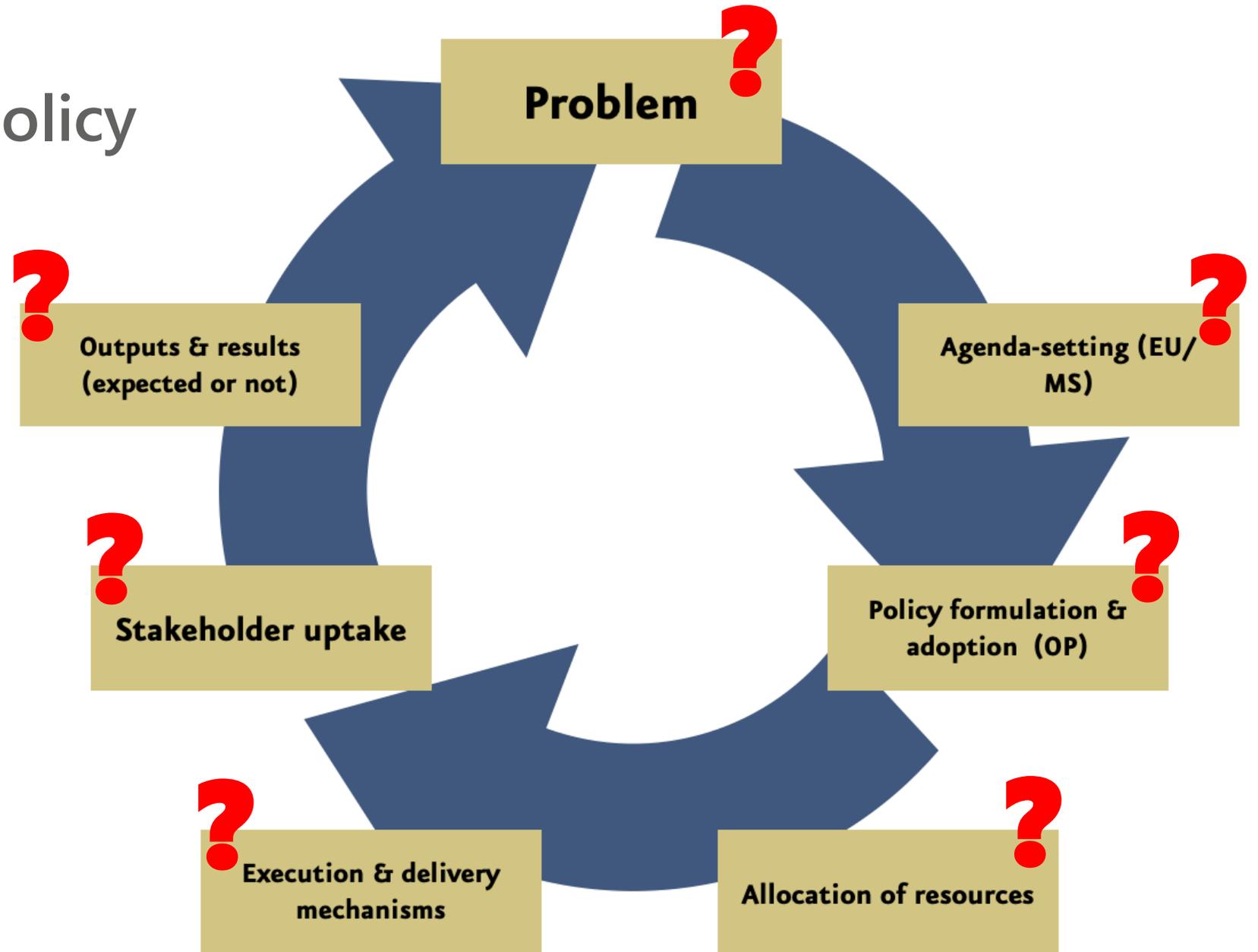
What do you need to do to write good ToRs?

- Definition of object: stakeholder mapping, problem definition, clarification of intervention's expectations
  - What are we talking about? How different from initial plans?
- Consultation of potential users
  - Is there something to learn? What are the current concerns? Is there a decision to be made? Etc.
- Initial portfolio review
  - What is there to evaluate? How does the current state of progress affect the evaluation?
- Agreement on priorities
  - What is most important to evaluate? What should the evaluation focus on?

- Technical feasibility assessment
  - Is there data to evaluate, or can we access that data easily?
- Simulation/planning of methodological approach
  - Do we have an idea of how this evaluation could be carried out?
- Budget simulation
  - Do we have an idea of the cost of this evaluation? How does it fit into our budget?
- Simulation of selection
  - What does good look like, what 'markers' of quality, how does this translate into criteria?
- Consultation with procurement
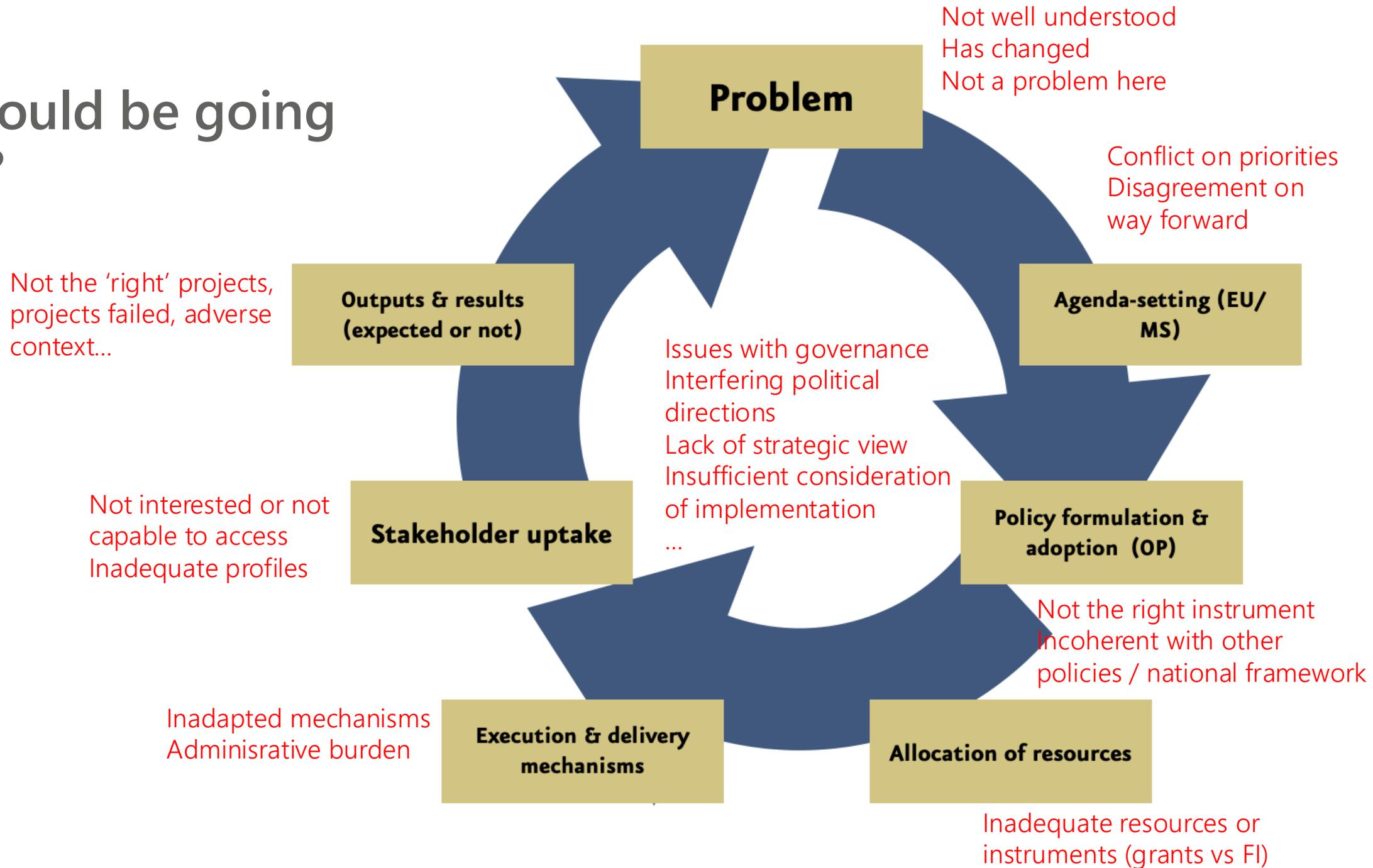 - Does it work – administratively, legally?
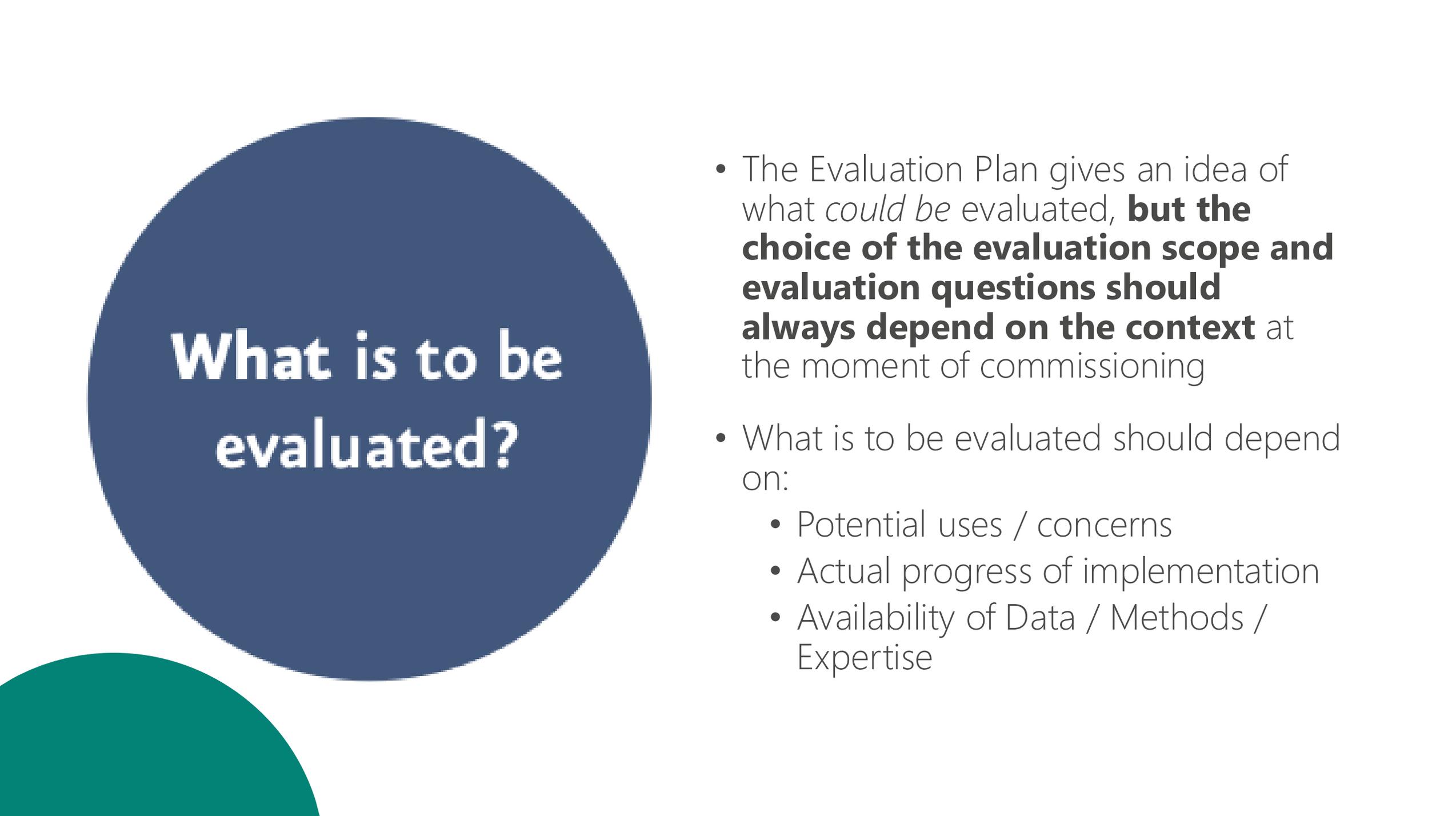
What is to be evaluated?

- The Evaluation Plan gives an idea of what *could be* evaluated, **but the choice of the evaluation scope and evaluation questions should always depend on the context** at the moment of commissioning

- What is to be evaluated should depend on:
  - Potential uses / concerns
  - Actual progress of implementation
  - Availability of Data / Methods / Expertise

Evaluation in the policy cycle

What could be going wrong?

Problem

Not well understood
Has changed
Not a problem here

Conflict on priorities
Disagreement on way forward

Agenda-setting (EU/MS)

Not the 'right' projects, projects failed, adverse context…

Outputs & results (expected or not)

Issues with governance
Interfering political directions
Lack of strategic view
Insufficient consideration of implementation
…

Policy formulation & adoption (OP)

Not interested or not capable to access
Inadequate profiles

Stakeholder uptake

Not the right instrument
Incoherent with other policies / national framework

Inadapted mechanisms
Adminisrative burden

Execution & delivery mechanisms

Allocation of resources

Inadequate resources or instruments (grants vs FI)

Make a case for use

Defend choices
"Defend and justify our choices before decision-makers or partners and revise them if necessary."

Inform strategy
"Stimulate strategic reflection by management (at operative, executive, or top level)."

Learn lessons
"Learn lessons and put good practices to work when carrying out interventions."

Better appraise and design
"Provide better insights at the appraisal phase (better ex-ante evaluation of projects) and improve the design of future interventions (wider range of methods, feasibility, etc.)."

Support discussions
"Substantiate internal and external debates on the best policy guidelines or ways to solve development problems."

Strategic uses
Internal uses or uses that target institutional partners, to support sectoral policies and strategies

Continuous-improvement uses
Primarily internal uses, to improve future projects and interventions

Monitor over time
"Help in project implementation monitoring, especially by providing the data needed to make informed choices."

Management uses
Primarily internal uses, to better monitor and manage ongoing projects and interventions

Dialog uses
Mostly external uses, aimed at strengthening or stimulating partnerships in the projects and policies supported

Dialog with partners
"Facilitate dialog with partners, in particular by building a relationship of trust."

Support decision-making
"Support decision-making on appraisal, progress, or renewal of projects."

Reflect with the team
"Facilitate reflection within teams on the purpose and quality of the action, by providing internal feedback."
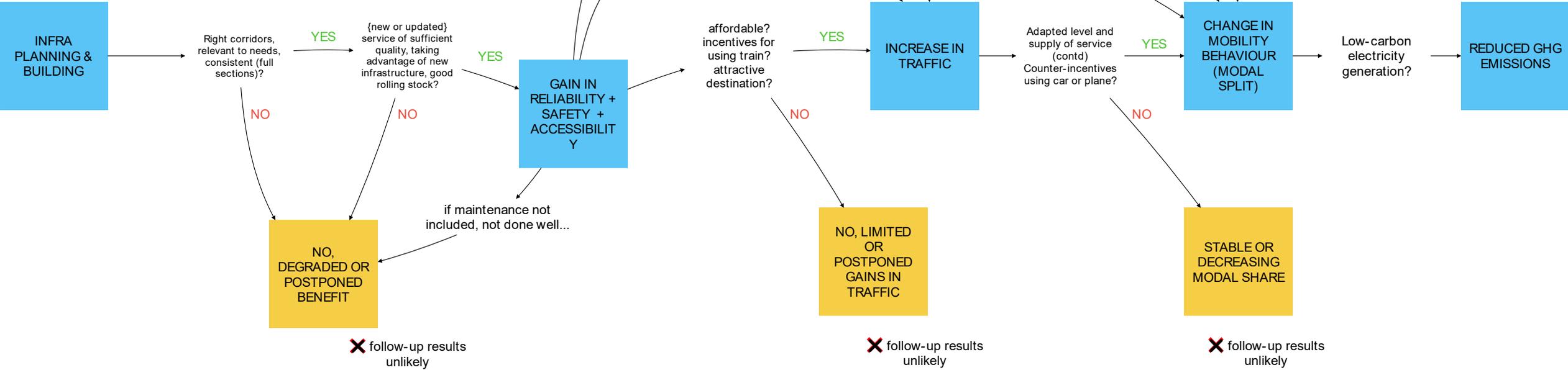
Strengthen capacities
"Help strengthen the capacity of the stakeholders involved"

The following persons helped develop this map of uses: Thomas Delahais, Agathe Devaux-Spatarakis, François Jégou, Jade Joviado, Camille Laporte, Léa Macias, Fiora Noël, Karen Rousseau, Claire Zanuso.

AFD

Clarify assumptions to support scoping

RAILWAYS PLANNING & BUILDING → GAIN IN RELIABILITY + SAFETY + ACCESSIBILITY → INCREASE IN TRAFFIC → CHANGE IN MOBILITY BEHAVIOUR → REDUCED GHG EMISSIONS

TEN-T : A Pipeline model

INFRA PLANNING & BUILDING

Right corridors, relevant to needs, consistent (full sections)?

YES → {new or updated} service of sufficient quality, taking advantage of new infrastructure, good rolling stock?

NO → NO, DEGRADED OR POSTPONED BENEFIT

NO → NO, DEGRADED OR POSTPONED BENEFIT

YES → GAIN IN RELIABILITY + SAFETY + ACCESSIBILITY

if maintenance not included, not done well... → NO, DEGRADED OR POSTPONED BENEFIT

Multimodal arrangements?

New international routes with sufficient quality

affordable? incentives for using train? attractive destination?

YES → INCREASE IN TRAFFIC

NO → NO, LIMITED OR POSTPONED GAINS IN TRAFFIC

Ending investments in car and plane

Change in car/ carbon-intensive culture

Adapted level and supply of service (contd) Counter-incentives using car or plane?

YES → CHANGE IN MOBILITY BEHAVIOUR (MODAL SPLIT)

NO → STABLE OR DECREASING MODAL SHARE

Low-carbon electricity generation? → REDUCED GHG EMISSIONS

✖ follow-up results unlikely

✖ follow-up results unlikely

✖ follow-up results unlikely

Projects A, B ————————→ ✖ (because...)

Projects B, C ————————————→ ✖ (because...)

Projects D, E, F ————————→ ☑ ——————→ ✖ (etc.)

Projects G, H, I ————————→ ☑ ————————————→ ☑ ————————————→ ✖

Etc. ————————→ ☑ ————————————→ ☑ ————————————→ ☑

INFRA PLANNING & BUILDING

Right corridors, relevant to needs, consistent (full sections)? — **YES** → {new or updated} service of sufficient quality, taking advantage of new infrastructure, good rolling stock? — **YES** → **GAIN IN RELIABILITY + SAFETY + ACCESSIBILITY**

afford... incent... using... attra... destir...

NO → **NO, DEGRADED OR POSTPONED BENEFIT**

NO → **NO, DEGRADED OR POSTPONED BENEFIT**

if maintenance not included, not done well... → **NO, DEGRADED OR POSTPONED BENEFIT**

✖ follow-up results unlikely

Break

11:20 – 12:30
Plenary

Scoping the evaluation; choosing a unit of analysis; specifying the overall evaluation approach and a suitable 'menu' of methods: the role of programme theory.

Evaluation Questions and Evaluation Purposes

EQs fall into broad types, related to the purpose (or purposes) of an evaluation. These will include:

- Accounting for Results
- Better Management and Delivery
- Learning and improvement
- Capacity development and sustainability

It is also worth differentiating between 'High Level' EQs from subsidiary EQs – in order to decide on evaluation priorities and where efforts should be concentrated

Accounting for Results

At a *high level* we may want to know:

- Have programmes achieved their objectives?

… In greater detail we may want to know:

- Can results be shown to be consistent with the ToC/Intervention Logic?

- Can we demonstrate that the programme caused the result?

- Who were the programme beneficiaries & how did they assess its results?

- Have results occurred consistently and if not, how can we explain differences across programmes?
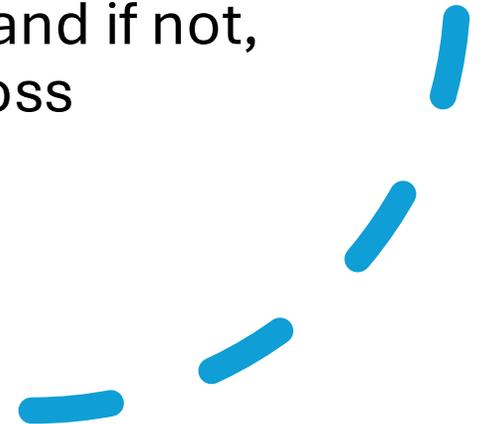
Better Management and Delivery

At a *high level* we may want to know:

- How well-managed was the programme?

In greater detail we may want to know:

- Were resources well-targeted?
- Were resources efficiently spent?
- Did delivery/implementation proceed as planned?
- Were mid-course corrections needed and why?

Learning and Improvement

At a *high level* we may want to know:

What have been the lessons for programming and policy in future?

In greater detail we may want to know:

Could a similar programme be better targeted and delivered if it was to be repeated?

Are there new ideas about how to achieve similar results through other means?

Do these kinds of programmes contribute sufficiently to strategic policy goals?

Capacity development and sustainability

At a *high level* we may want to know:

- Has the programme helped create new capacities and networks?

In greater detail we may want to know:

- Will new capacities continue to generate the kinds of results that the programme intended?

- Are these capacities likely to be self-sustaining and economically viable?

- Have capacities been created in civil society or in public administrations?

- If there is little evidence of sustainability, why is this?

Designing Evaluations
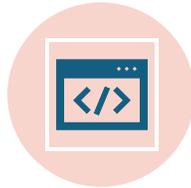
A 'design' is more than a method
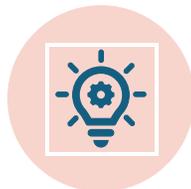
A design involves a deep understanding of what we want to know, the programme context in which we are working and of the capabilities of different families of methods

Interview surveys, case-studies, statistical analyses of labour market data, observational studies, all can be part of different designs

On the basis of this understanding, we can consciously choose which of the many ways we could evaluate any programme

The Design Triangle
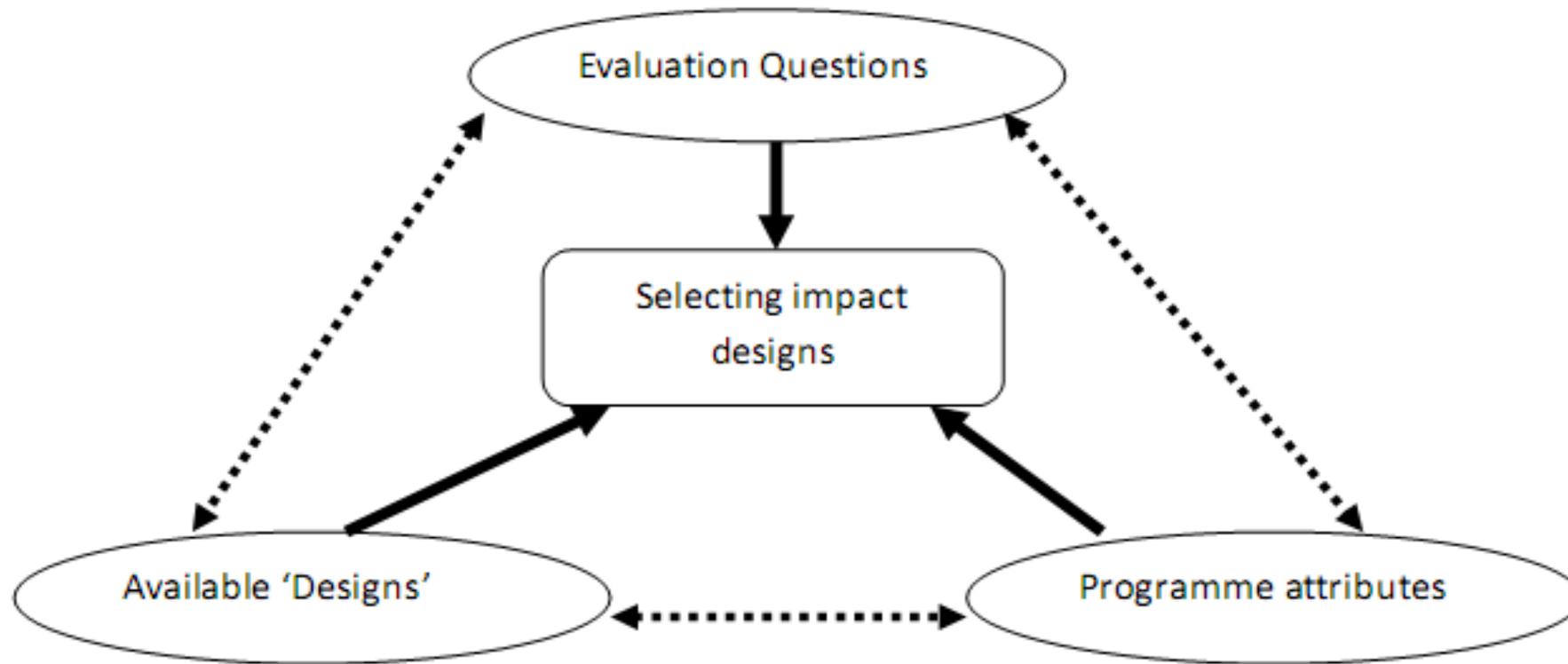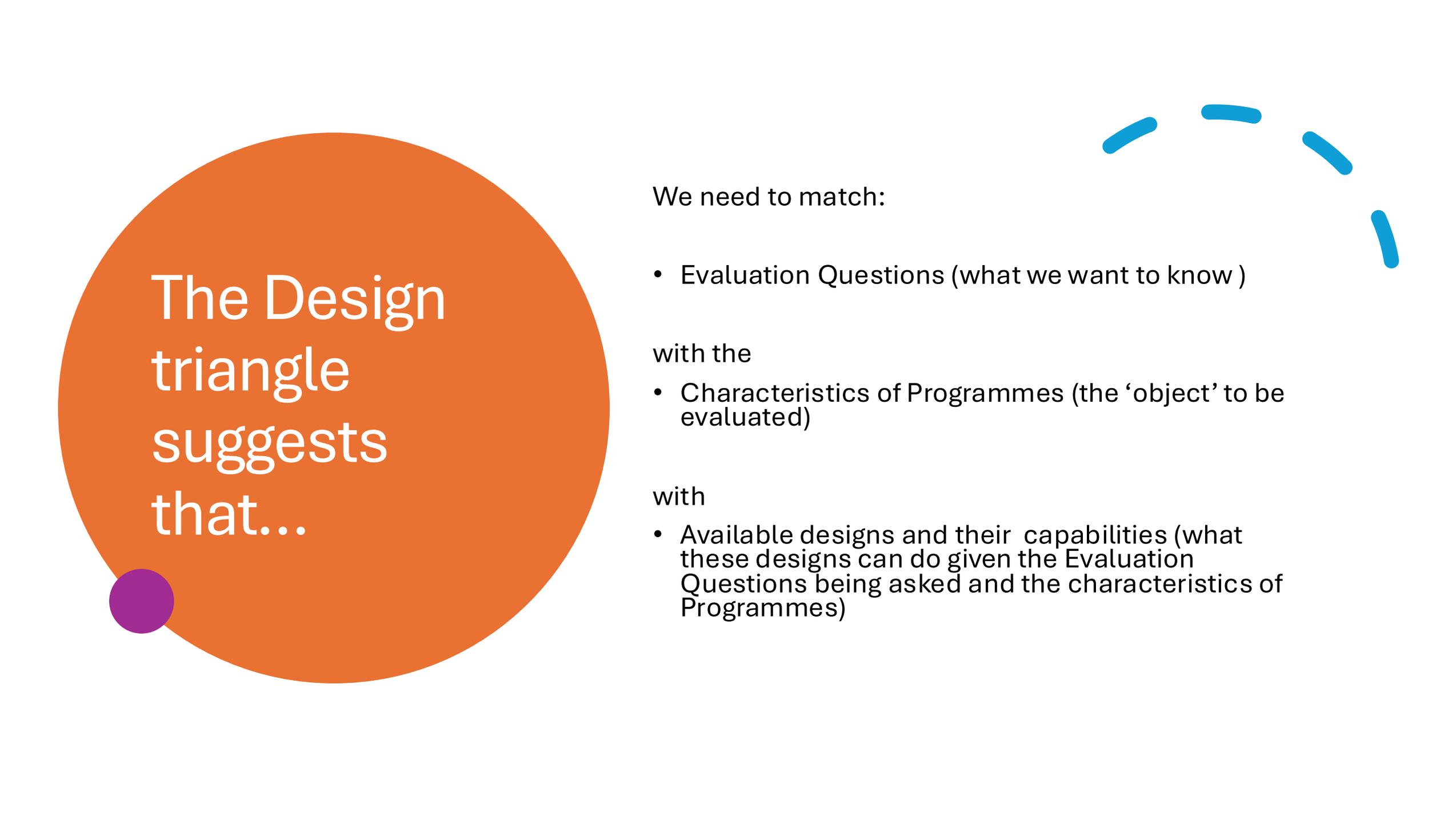
The Design triangle suggests that...

We need to match:

- Evaluation Questions (what we want to know)

with the

- Characteristics of Programmes (the 'object' to be evaluated)

with

- Available designs and their capabilities (what these designs can do given the Evaluation Questions being asked and the characteristics of Programmes)
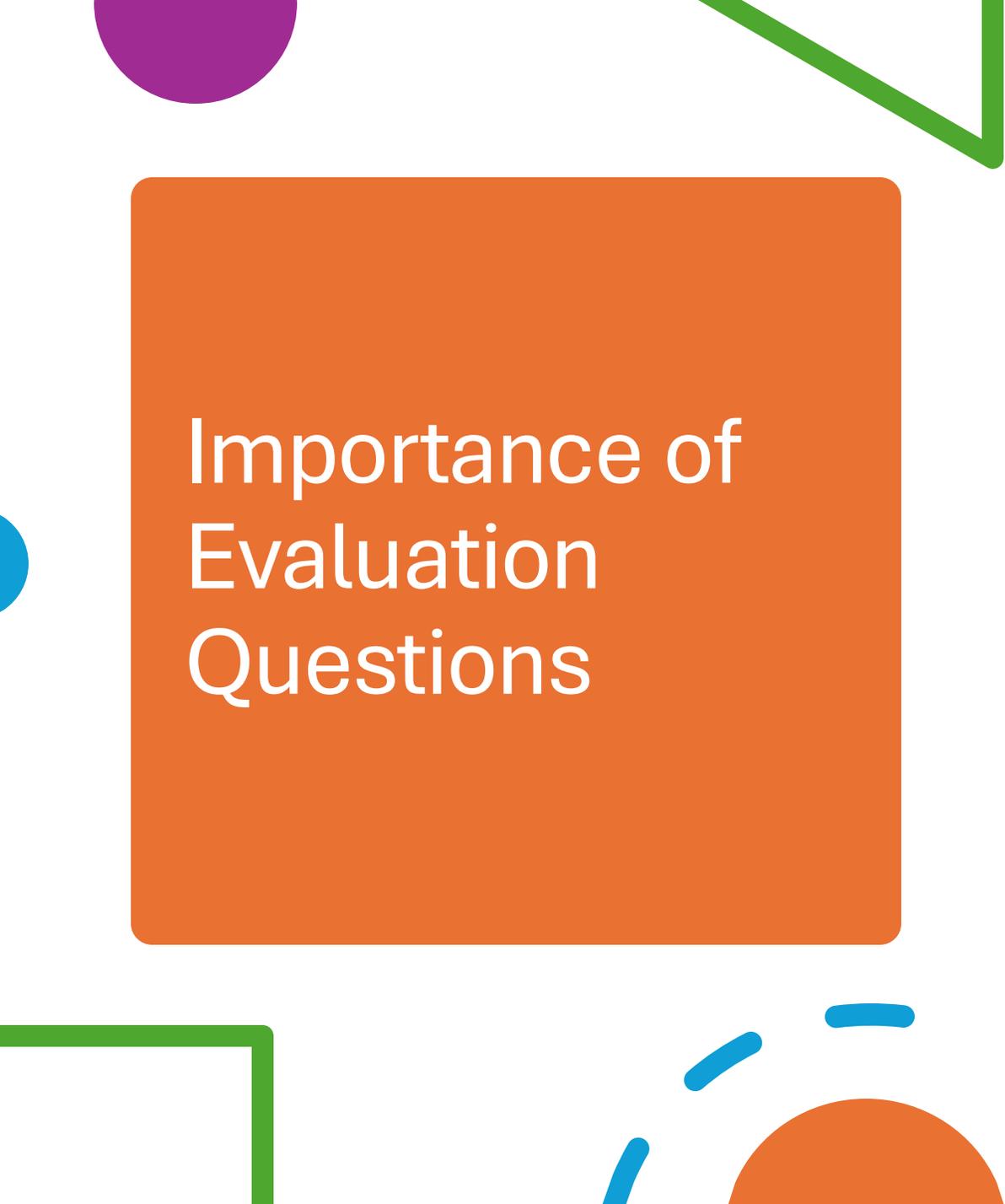
Importance of Evaluation Questions

- Historically evaluations were structured around generic criteria such as Effectiveness; Relevance; Efficiency; Coherence; and Sustainability

- These criteria-led evaluations often produced evaluations that lacked specifics and were not useful for policy purposes. Nowadays we tend to operationalise criteria into more specific *Evaluation Questions*

- Evaluation Questions (EQs) are also a crucial link between evaluation purpose and how we make methodological choices

Evaluations need to answer clear and answerable questions!
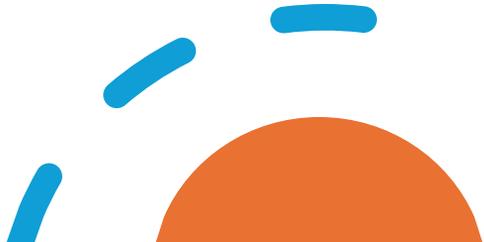
Importance of Evaluation Questions

- EQs are not the same as an 'interview question' – they are 'high-level' question able to be analysed and studied

- EQs focus on the relationship between a programme and real world 'consequences' 'effects' and 'results'

- EQs that only focus on the internal operations of a programme and do not try to illuminate the relationships between programmes and real-world effects are not EQs!

- Most evaluations will have a number of prioritised EQs – each addressing a specific point of interest in a programme

Centrality of Evaluation Questions
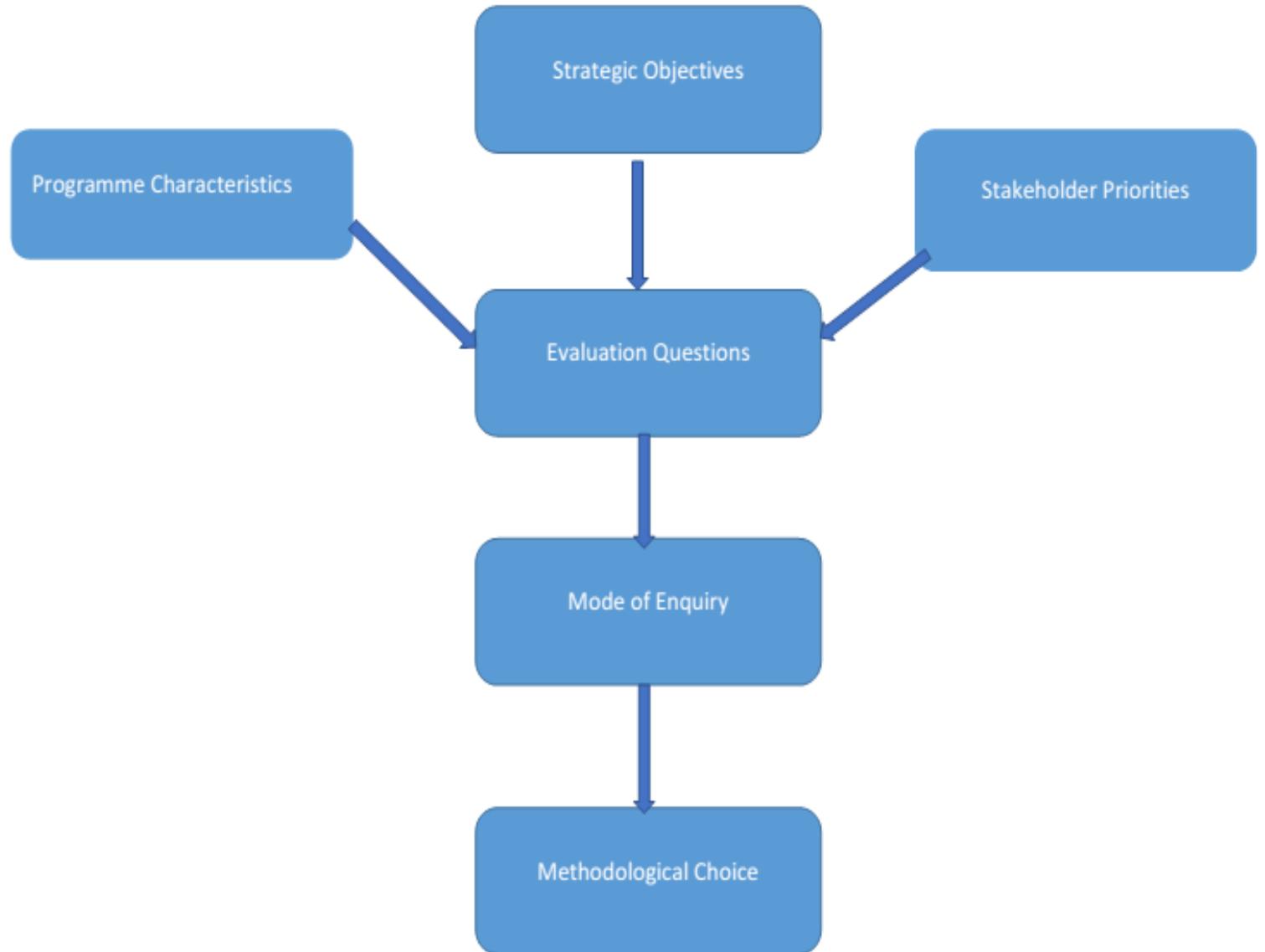
Evaluation Questions and Strategic Objectives

- Strategic Objectives refers to the policy goals that justify a programme – in our case through the various programme priorities – strengthening research; access to ICTs; enhancing competitiveness; promoting social inclusion etc. - that are intended to contribute to 'smart, sustainable and inclusive growth'

- Evaluation Questions need to relate to these goals and priorities although monitoring/indicator exercises that are not necessarily evaluative will also address results

Programme Characteristics

- Evaluation Questions have to take account of programme characteristics

- There are many ways to characterise a programme – in terms of sectors; objectives, their degree of innovativeness etc.

- For the purpose of specifying EQs, characteristics should first be understood in terms of an 'intervention logic', 'theory of change' or 'programme theory' set into a wider context

- Theories of Change can be variously described but usually cover the sequencing of a programmes cycle set into a wider socio-economic and regional context

Programme Characteristics
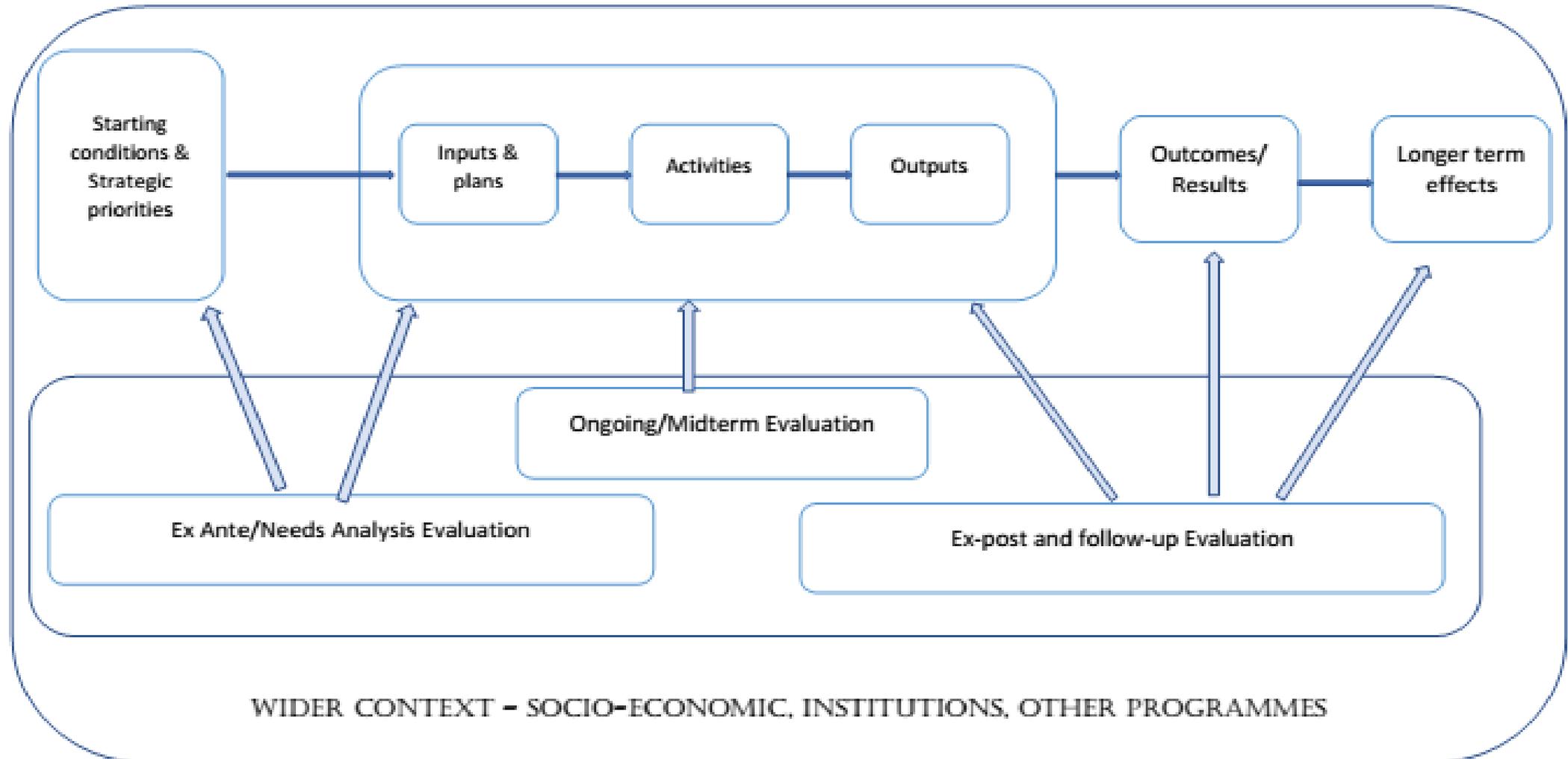
Programme Characteristics

When we discuss methodological choices there are other kinds of programme characteristics that also have to be considered. These include for example:

- Innovativeness of programme goals and delivery

- Simplicity or complexity of the intervention

- Types of outputs envisaged – whether they are material, behavioural, new services, new institutional arrangements

- Timescales and trajectories of change

- How bounded or embeddedness programmes are in relation to other programmes, activities and systems

Stakeholder priorities

- Evaluations like programmes have stakeholders

- Stakeholders will want answers to *their* questions

- Stakeholders are the users of evaluation – they are also often the gatekeepers to evaluation data and provide necessary cooperation for evaluators

- The credibility and legitimacy of an evaluation will often depend on how far the questions of stakeholders are prioritised
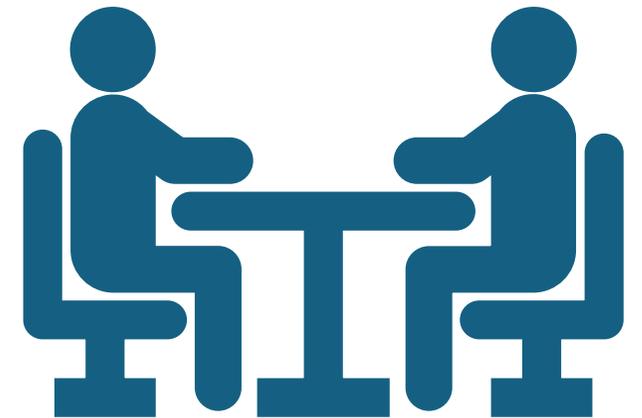
12:30 – 13:00
Introduction to group works

Scoping the evaluation; choosing a unit of analysis; specifying the overall evaluation approach and a suitable 'menu' of methods: the role of programme theory.

Lunch

14:00 – 15:30
Group work

Scoping the evaluation using the programme theory based on case study. Using scoping work including the programme theory to write the ToR

Break

15:45 – 16:45
Group work

Asking evaluation questions jointly with key stakeholders guided by the programme theory. Identifying areas for inquiry. Exploration of methods to answer the questions

16:45 – 17:00
Wrap-up

Day 2

09:30 – 10:15

Discussion

Interim review ενδιάμεση ανασκόπηση of Day 1 by Helpdesk experts and DG REGIO and DG EMPL representatives

10:15 – 11:30
Plenary

Quality Assurance

A good evaluation is an evaluation…

1. That is implemented according to what was initially planned?
2. That follows a set of recognised rules and standards of quality?
3. That implements a cutting-edge methodology/μεθ αιχμής?
4. That provides a fair assessment / καλή εκτίμηση of the situation?
5. That provides new perspectives on the intervention / that makes you think?
6. That is being used?
7. That is an opportunity for dialogue among stakeholders?
8. Which results stakeholders trust?
9. …

Go to wooclap.com and use code KOCZUC

Allez sur **wooclap.com** et utilisez le code **KOCZUC**

A good evaluation is an evaluation that...

|			
4 provides a fair assessment of the situation?	815 pts	**8** which results stakeholders trust?	440 pts
5 makes you reflect on the intervention ?	644 pts	**7** helps stakeholders talk to each other?	431 pts
6 is being used by actors	624 pts	**1** is implemented according to the initial plan?	364 pts
2 follows a set of recognised rules and standards of quality?	565 pts	**3** implements a cutting-edge methodology?	317 pts

Definition of quality is shifting

COMPLIANCE TO TOR?
ΣΥΜΜΌΡΦΩΣΗ

RELIABLE BY DESIGN
ΑΞΙΟΠ ΣΧΕΔΙΑΣΜΌΣ

TRUSTABLE IN PRACTICE ΑΞΙΟΠ
ΕΦΑΡΜΟΓΉ
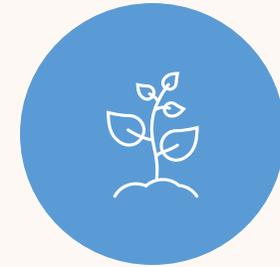
USEFUL

Quality in the evaluation process

Consider potential needs and uses and adapt to them

Relevant scope and adapted resources

Adequate approach and competencies

Respect of method. standards
Consideration of multiple perspectives

Justified adaptations to challenges

Transparent cross analysis
Evidence-based conclusions and recos

| Design Σχεδιασμός | Inception Σύλληψη | Implementation Εφαρμογή | Finalisation Ολοκλήρωση |

Quality starts with good ToR!

✓ Set the evaluated programme in a clear policy context

❓ Ask clear evaluation questions

📋 Give reasonable expectations of what the evaluation should cover

💡 To do all of the above, clarify the Theory of Change behind the intervention

💡 Hint at how the evaluation could be used

The case for quality assurance

- Quality control is like the cavalry; it is always late

- Quality goals are adaptive, they change along the evaluation process and must be negotiated

- Quality as usefulness is uncertain and means seizing opportunities as they arrive
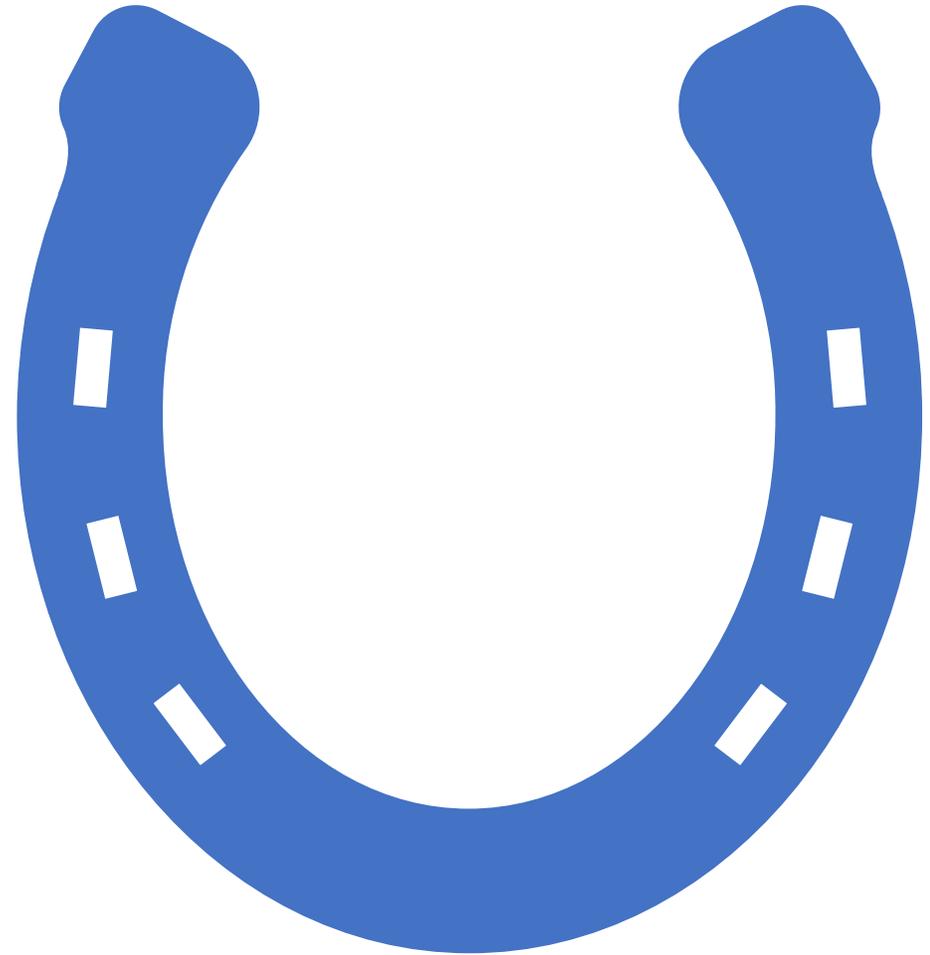
Choosing the best proposal

- As any evaluation process, choosing the best proposal means **formulating explicit criteria**: administrative, technical, financial

- These criteria can be pondered; some criteria can lead to the exclusion of proposals

- Those aspects easiest to assess are not necessarily those most relevant to choose

- Paradox of multicriteria analysis: mediocre proposals on all aspects can be retained over those with a few excellent points

Some criteria to assess the approach

There is an **overall strategy** to answer evaluation questions (not a mere list of tools and methods)

Each tool or method is justified by **information needs in answer to EQs**

The approach is consistent, and you can see how the different phases fit together

There is an **initial discussion of challenges** to the robustness of the evaluation process, or to the generalisability of its finding, and solutions are provided

The approach and methods are **illustrated with examples** or samples of original work
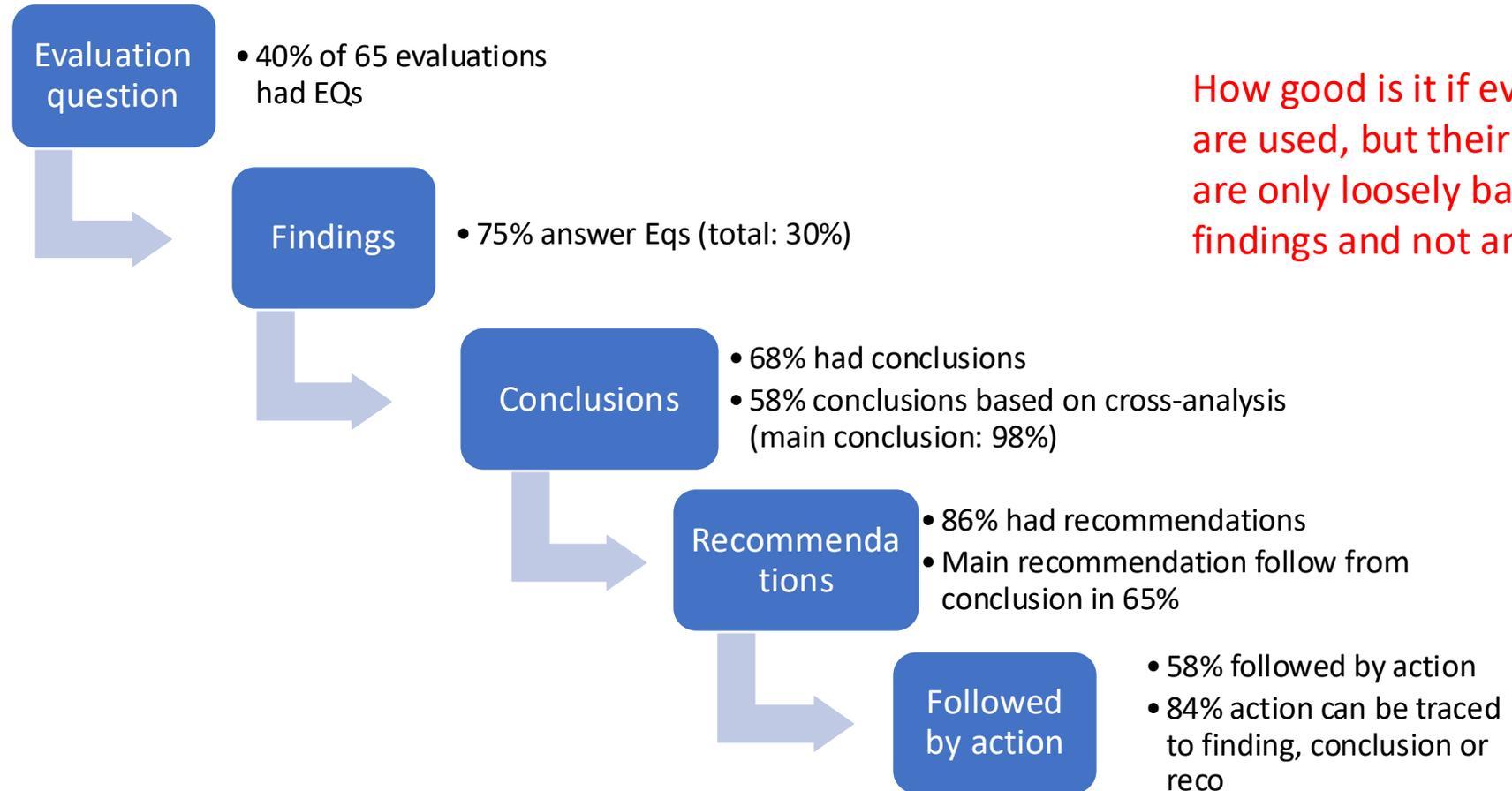
A few tips to receive better proposals

- Inform potential bidders **early enough** ('prior information notice')

- Write ToRs in ways that will **arise intellectual interest of providers**

- Make **documentation available**, including through open data

- Make explicit that **providing original analysis in proposal will count** in selection

- **Encourage** bidders to answer **collectively** when different skills are needed

- Do not forget **feedback to non-retained**

- **Ask bidders** what they think about your ToRs and ways to make them more attractive

A few tips to make better choices

- Use **thresholds** to avoid very bad proposals or bidders competing exclusively on price
- For bigger evaluations, consider **different steps in bidding**, e.g. an open call for interest; publication of ToRs and selection of 3 best proposals; audition and adaptation of proposal; award of the contract
- **Maintain long-term relationships** beyond bids; learn which providers are most likely to respond to which ToRs
- Open the selection process to **a jury** involving different stakeholders

Quality: beware of the weak link!

Evaluation question
- 40% of 65 evaluations had EQs

Findings
- 75% answer Eqs (total: 30%)

Conclusions
- 68% had conclusions
- 58% conclusions based on cross-analysis (main conclusion: 98%)

Recommendations
- 86% had recommendations
- Main recommendation follow from conclusion in 65%

Followed by action
- 58% followed by action
- 84% action can be traced to finding, conclusion or reco

How good is it if evaluations are used, but their conclusions are only loosely based on findings and not answering EQ?

Metaevaluation of 65 evaluations engaged in the framework of the "Public Action Modernisation" in France (2012-2017)

Evaluation Quality as an Organisational Reflex

- The focus of most quality initiatives is on managing individual evaluations or a portfolio of evaluations

- Quality Assurance is also an organisational matter

- Are there organisational procedures, routines and budgets necessary to underpin quality?

- Are guidelines in place about reviewing existing evidence, using peer review panels, choosing methods, involving partners and stakeholders?

- Have ground rules been negotiated – to ensure data access, ensure evaluation independence, disseminate reports

Break

11:45 – 12:30
Group works
Identifying potential quality issues and ways to address them

12:30 – 13:00
# Plenary

Designing for policy use

Making Evaluations Useful

- Using evaluations constructively is often something we think about at the end of an evaluation – how to disseminate results

- In our experience at the Helpdesk this is often too late!

- So what we want to do today is to present the case for planning useful and useable evaluations from the very beginning of evaluations

Making Evaluations Useful

- Evaluations cost much time and many €uro –

- but they are not always useful or used……

- Two reasons:
  - We don't always remember the *full range* of purposes that evaluations can help achieve
  - We don't tend to plan the engagement of stakeholders across the programme & policy lifecycle to take advantage of these purposes
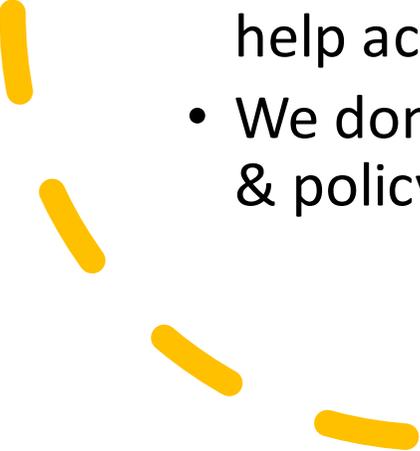
Making Evaluations Useful

- Evaluation is a 'multi-purpose tool' – not just an administrative burden - when used well!

- Purposes can include:
  - Accountability for monies spent
  - Management & implementation support
  - Programme review –mid-course correction
  - Policy strengthening
  - Policy planning

Making Evaluations Useful

Not every evaluation fulfils every purpose!

- Depends on national priorities & capacities
- Extent to which programmes are innovative or uncertain
- Also depends on characteristics of programmes

Programme characteristics matter…

- A standard programme – offering loans or subsidies or delivering a service through a single administrative 'channel'

Or

- An infrastructure investment in road, rail & electricity grids

Or

- An innovative 'Just Transition' initiative that engages many stakeholders – enterprises, citizens & public agencies…

Or

- An RTD investment that depends on new governance arrangements supporting joint decision making & cooperation

- Some programmes are likely to have short term results while others have intergenerational ambitions

- Some programmes are free-standing while others are integrated into national, and European strategies

- Some are inter-regional and transnational

- Some involve communities and 'bottom-up' inputs, others do not

- And most disruptive nowadays is how often circumstances change and policy goals also have to adapt………..

Making Evaluations Useful

! Some programmes/policy instruments more likely to need mid-term review and redirection than others…….

⚠ Although in times of geopolitical uncertainty many more will need *evidence-based* redirection

👥 And different programmes involve different types and numbers of stakeholders in evaluations

🏛 National governments, regional authorities, service providers, SMEs, beneficiaries, universities, municipalities, trade associations, civil society…….

Design principles for useful evaluations….

- Identifying where and when use should be prioritised
- Identifying potential users and stakeholders
- Identifying what potential users want to know
- Matching policy and programme lifecycle with timescales for evaluation 'delivery'
- Designing a process of engagement that identifies both information needs *and* sites for ongoing dialogue

Making Evaluations Useful

- Why 'dialogue' and 'engagement'?

- Often in evaluation we apply a 'learning' lens as well as an 'accountability' lens

  - An accountability lens emphasises reports and analysis at the end of a policy-cycle, but a learning lens emphasises engagement and dialogue
  - Partly because it is difficult to know all the questions to ask when new programmes are launched
  - Partly because interpreting results of evaluations is often a multi-party endeavour

Lunch

14:00 – 15:00
Group works

Strategies for evaluation use

15:00 – 15:45
Feedback

Presentation by 4 groups, feedback

15:45 – 16:00
Plenary
Tips and tricks

APPLICA   ISMERI EUROPA   QUADRANT CONSEIL

Evaluation Helpdesk on
Cohesion Policy
https://www.linkedin.com/
groups/14420317/

Evaluation Helpdesk (2025)