



METODIKA

Context & key concepts of the new Evaluation Methodology

**Erik Arnold & Bea Mahieu -
Technopolis Group**

www.metodika.reformy-msmt.cz



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



The EM in its historical context

- **Evolution after the reform of 2008:**

- Progressive restriction of the **scope**: research outputs only
 - Reduction of the complexity of performance to an overly simple category of outputs
 - Lack of consideration for disciplinary differences and for the missions of ROs
- Narrowing of the **function** of evaluation: only as component of the PRFS
 - Concept of evaluation as part of a policy cycle providing strategic information is not perceived; it provides information that is at the best of limited relevance
- Increasing **breadth** of coverage: from funding bodies to individual researchers
- Metodika 2013-2015: an improvement but still focused exclusively on outputs

- **Result:**

- Evaluation is perceived as counting outputs = points = funding
- It constitutes the key factor for R&D management throughout the entire system
- It is detached from any discourse on policy and strategy related to the national R&D system



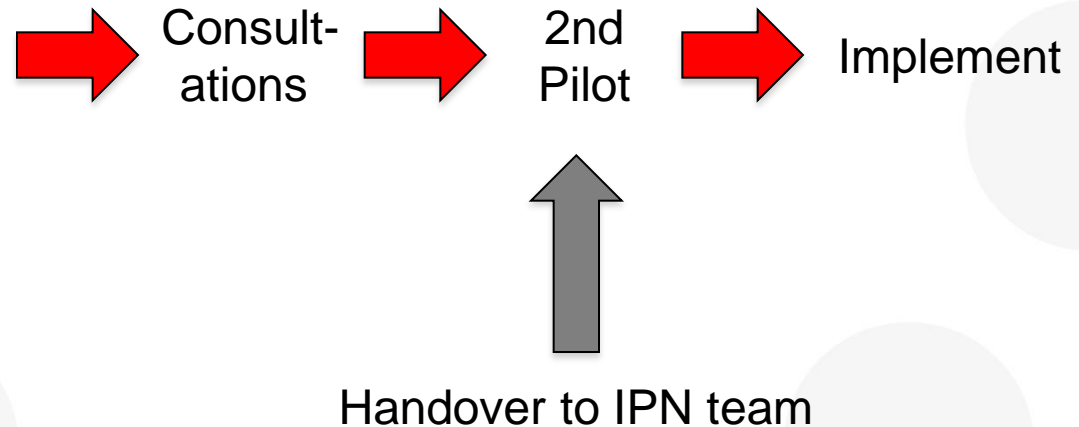
The mandate for this study

- **Objectives: to develop an evaluation methodology**
 - Conducted on a **national** basis at the **institutional** level
 - Evaluation results inform the institutional **funding** system (PRFS)
 - Providing strategic **information** for the actors at all levels in the R&D system
- **The expectations:**
 - A peer-review evaluation process
 - Fulfil formative and summative functions
 - Cover outputs, impacts, and institutional projections of research development
 - Take into consideration the different missions of research organisations and the field specifics
 - Evaluation processes resistant to clientelism and conflicts of interests
 - Take into account 'gaming'
 - Total costs do not exceed 1% of public institutional support for R&D in a five-year time period

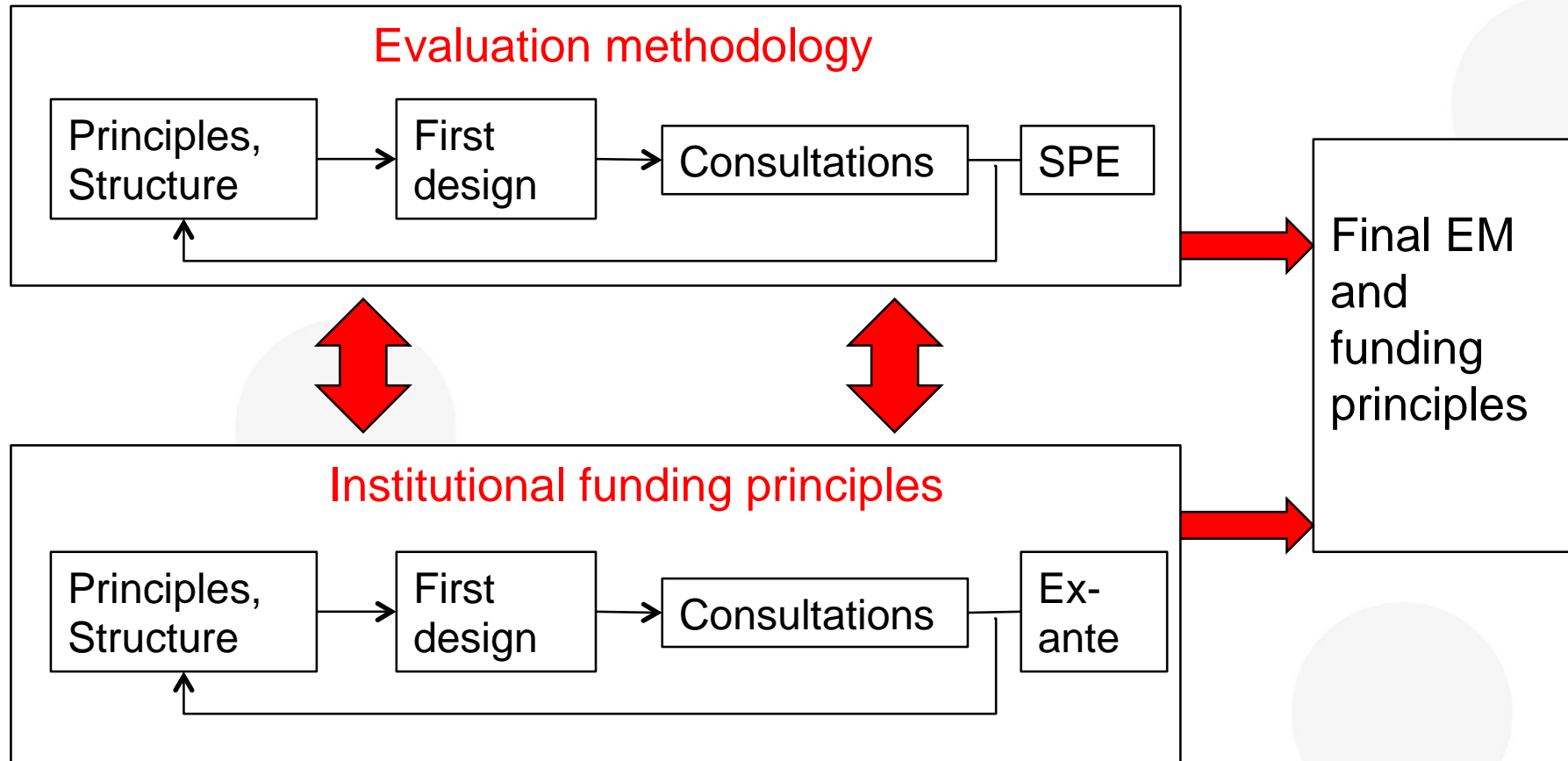
The project



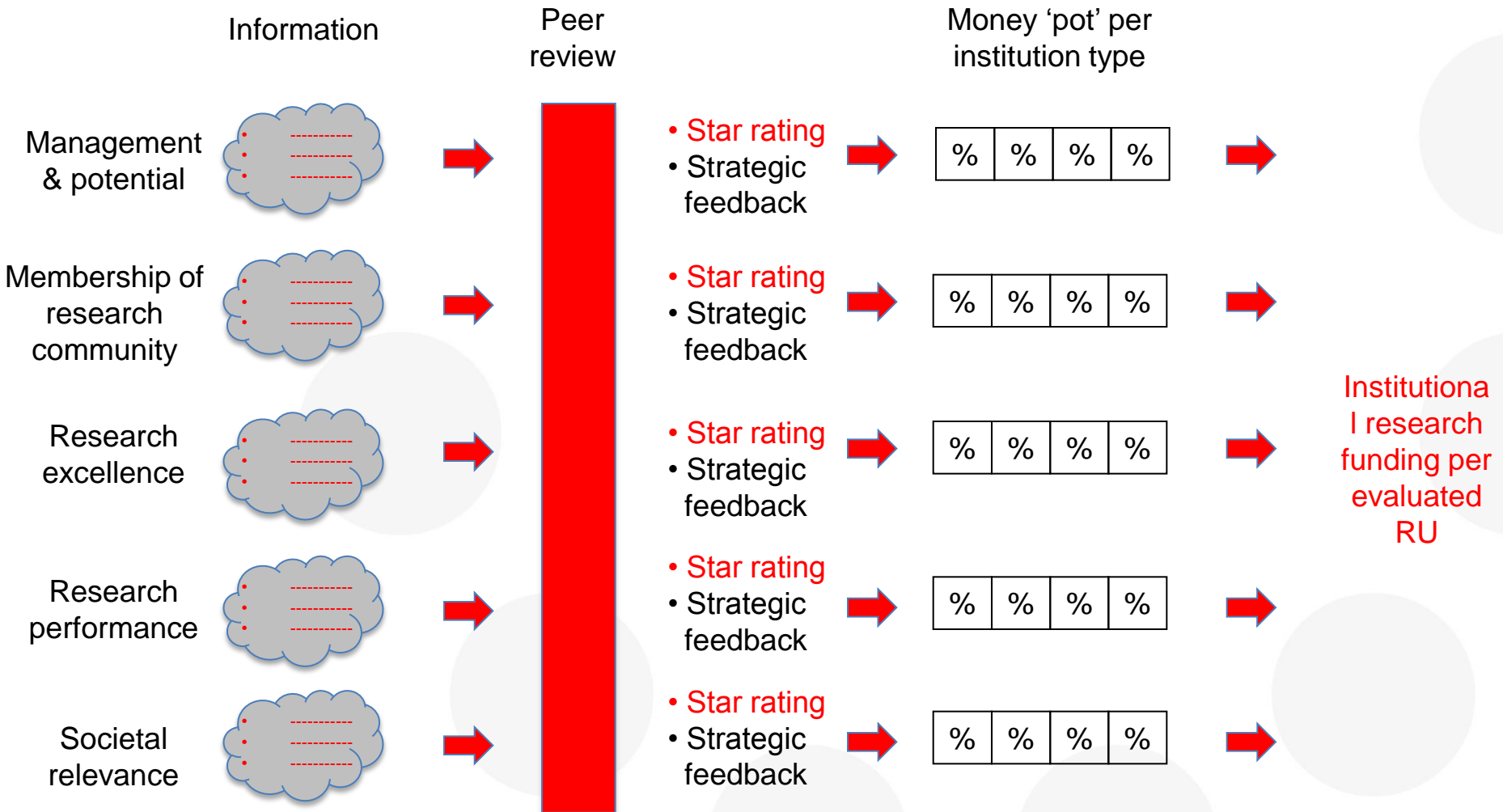
PHASE 1
WP1 - EvU evaluation structure determination
WP2 - EvU field-specific eval. methodology (FEM)
WP3 - EvU evaluation process rules
CONCLUSION PHASE 1
PHASE 2
WP4 - Evaluation information support
WP5 - Definition 'institutional funding' - intl practice
WP6 - SWOT Czech institutional R&D base
WP7 - Draft models FF & PBC
WP8 - Model exp. impacts new inst. funding system
CONCLUSION PHASE 2
PHASE 3
WP9 - Small Pilot Evaluation
CONCLUSION PHASE 3 / STUDY



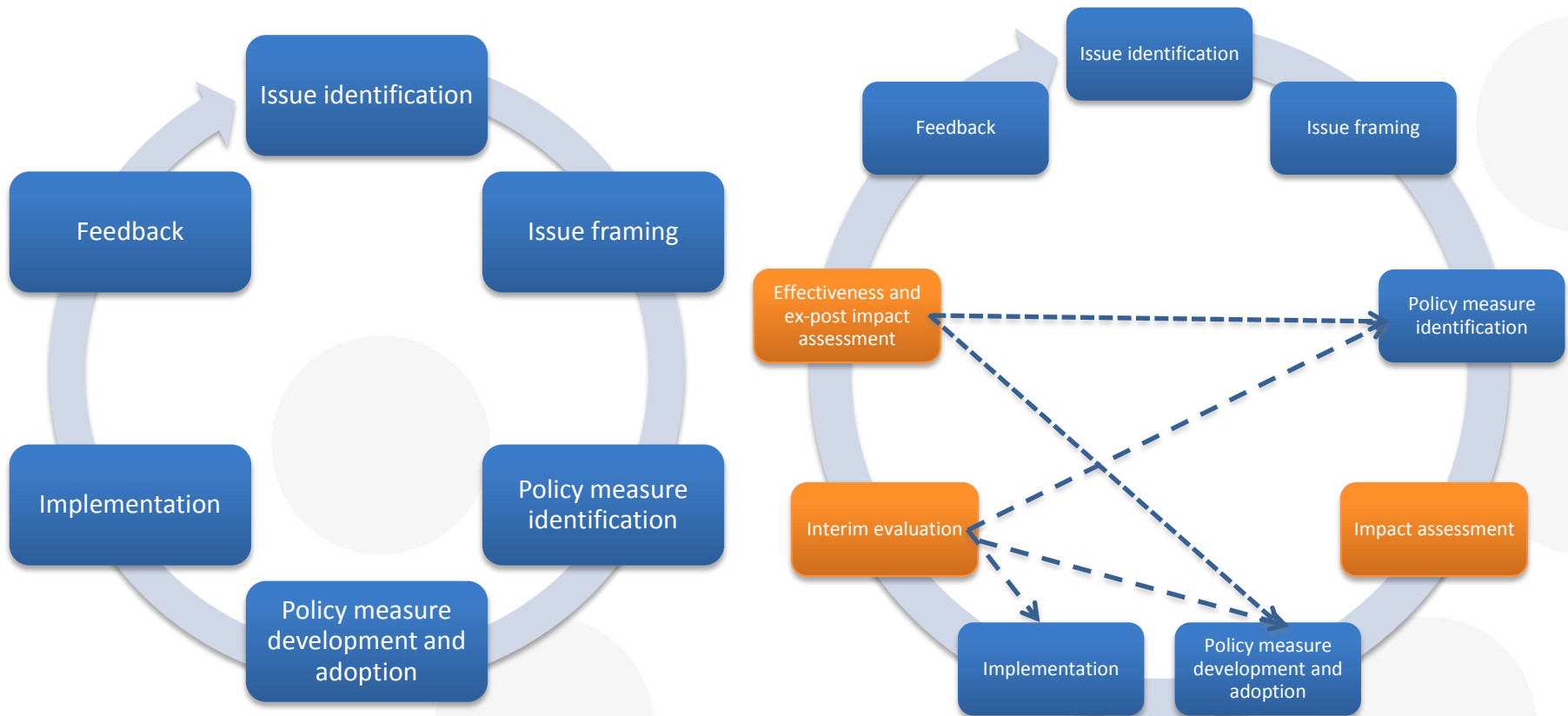
Evaluation (assessment) in the context of funding



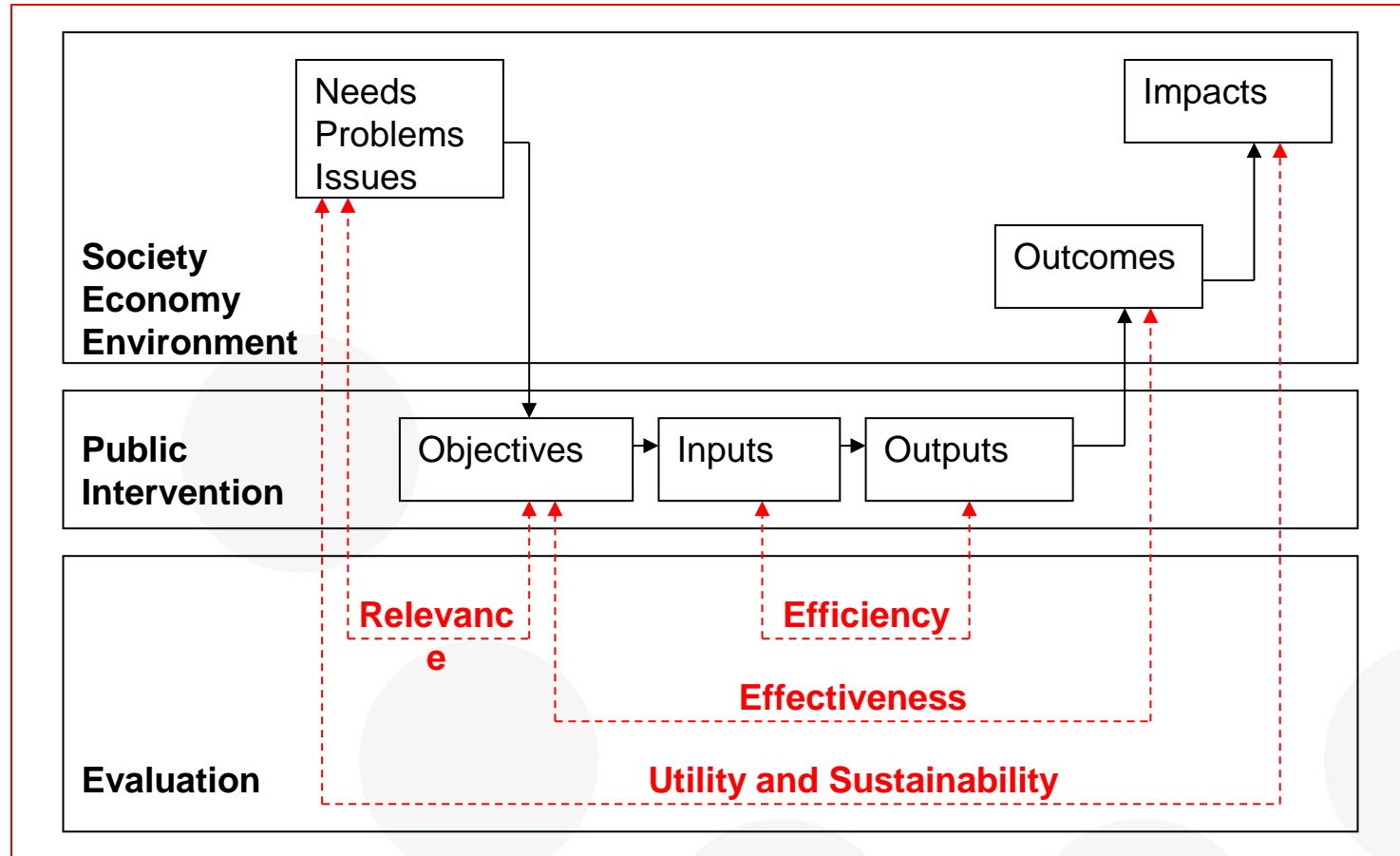
Evaluation in the context of a funding mechanism



Evaluation in general should inform the whole policy cycle



Evaluation should analyse societal effects of intervention, not just focus on outputs



Evaluation (assessment) for a PRFS is necessarily more narrow



'General' evaluation	Evaluation in a PRFS
Informs policy making on the failures in the system and recommends possible policy interventions	Is part of a policy intervention : it acts upon previously identified failures and steers research behaviour to tackle these by providing incentives
Has no effects directly linked to the evaluation	Is intended to create effects
Has no consequences of gaming or unintended effects	Has inevitable consequences of gaming and may lead to unintended effects
Gives information on the positioning of the evaluated objects in the national/international context	Sets the evaluated objects in competition to each other

Who does what



	Australia (2014)	Austria (2014)	Belgium (FL) – BOF (2014)	Finland (2014)	Italy (2014)	Netherlands (2014)	New Zealand (2014)	Norway / evaluations (2014)	Norway / PRFS (2014)	Sweden (2014)	UK – (2014)
Purpose											
Performance assessment	X	X	X	X	X	X	X	X	X	X	X
Inform funding			X	X	X		X		X	X	X
Main function											
Formative	X	X		X		X	X	X		X	
Summative			X		X				X		X
Formative function											
National R&D governance	X		X		X			X	X		X
Institutional R&D management	X	X		X		X	X	X		X	
Summative function											
R&D quality	X	X	X	X	X	X	X	X	X	X	X
R&D capacity building	X		X	X	X		X	X	X		X
Research excellence	X										X
Societal relevance	X				X	X					X
Notes											
		Performance contracts	Performance agreements	Performance contracts		Performance contracts		Separate PRFS – simple, 10% of funding	New PRFS in 2015		

National and institutional evaluations tend to be distinct



- **National evaluations typically**

- Are commissioned by national research funders or their principals
- Compare performance with other countries at the field level
- Collect strategic information to support research policy interventions
- Are summative with respect to the institutional level
- Lean heavily on bibliometric methods – with selective use of peer review

- **Institutional evaluations**

- Are commissioned by institutions or their principals
- If they benchmark (often they don't), compare with other institutions
- Try to explain performance at a level of detail that is formative, in order to support institutional research management
- Tend to use informed peer review (though note that modern research management routinely monitors performance by bibliometric means)

It's unusual to compare different types of RO



- **Types of organisation**
 - Universities / Higher Education Institutions
 - Scientific research institutes
 - Research and Technology Organisations (RTOs)
 - Government laboratories
 - NB that **hybrid** forms are increasingly appearing
- **Different types of RO are rarely benchmarked against each other**
 - Differences in mission or function
 - Different principals and accountability
 - Different mix among types of research and between research and non-research activities
- **Legal form is irrelevant – evaluation is based on function**
 - See, for example, the Norwegian institute PRFS, which separates institutes from universities but ignores legal form

Typical institutional research evaluation approaches by RO type



- **Universities**
 - Focus on quality, interaction with teaching
 - Preference for peer review
- **Scientific research institutes**
 - Ditto, but no emphasis on teaching unless PhD education is provided
- **RTOs**
 - Economic performance, social impacts, customer feedback, research quality
 - Predominance of social scientific methods; peers if quality is a major issue
- **Government labs**
 - Mission performance, usefulness to policy formation processes, social impacts, research quality
 - Predominance of social scientific methods; peers if quality is a major issue



Why use peers?

- **Legitimacy in the scientific community**
- **Ability to address context, for example at institutional level**
- **Understanding inter-field differences**
- **Ability to be formative**

- **Drawbacks include**
 - These evaluated generally have to do work as part of the process
 - Need to ensure independence (research funders devote significant effort to this)
 - Expensive, especially if site visits are involved
 - Shortage of peers
 - Occasional misbehaviour at the individual level – typically constrained by the use of panels



Why use bibliometrics?

- **Comparatively low cost – and the cost is declining**
- **Need not involve work by those evaluated**
- **Can produce fine-grained ratings and rankings – though arguably this is spurious precision**
- **However**
 - Inter-field comparisons are problematic
 - Use of indicators is generally unsophisticated and sometimes problematic (for example, Journal Impact Factors)
 - Lack of context indicators and conventions for handling them
 - Hence, they provide a much more partial picture than peer review
 - Unhelpful for formative evaluation
- **Emerging conclusion: use ‘informed’ peer review**



Some effects of PRFS

- **Positive**

- Improved research and management attention to productivity and quality
- Performance focus and individual and collective levels
- Restructuring towards high-quality performers
- Greater transparency about performance

- **Negative**

- Focus on producing indicators rather than science, encouraging 'gaming'
- Narrowing of the career path, with growing influence of research managers
- Squeezing out the heterodox and interdisciplinarity
- Matthew Effect encourages lock-ins
- Research indicators often used in relation to teaching



Unit of analysis

- **Depends on the purpose of the evaluation**
- **Tendency to avoid the individual level – that produces erratic results and undermines institutional autonomy**
- **Some systems require tactical decisions about including individual researchers, with intra-institutional and personal consequences**
- **Decisions to include organisations are normally policy-based. You can't 'volunteer' (and usually you can't avoid assessment, either)**



Pitting discipline against discipline

- **Most PRFS reallocate a small proportion of total institutional funding and are therefore tolerant of some inaccuracy or margin for error**
- **By and large, disciplines are incommensurable – one of the most interesting activities in scientometrics is the struggle to overcome this fact**
 - The nature of the research enterprise differs ('What is research?')
 - Differing notions of quality
 - Publication behaviours differ significantly
 - Different social roles of disciplines
- **It is not necessarily desirable to change the proportions of national effort among disciplines for reasons of performance only – policy needs to play a part**



Interdisciplinarity

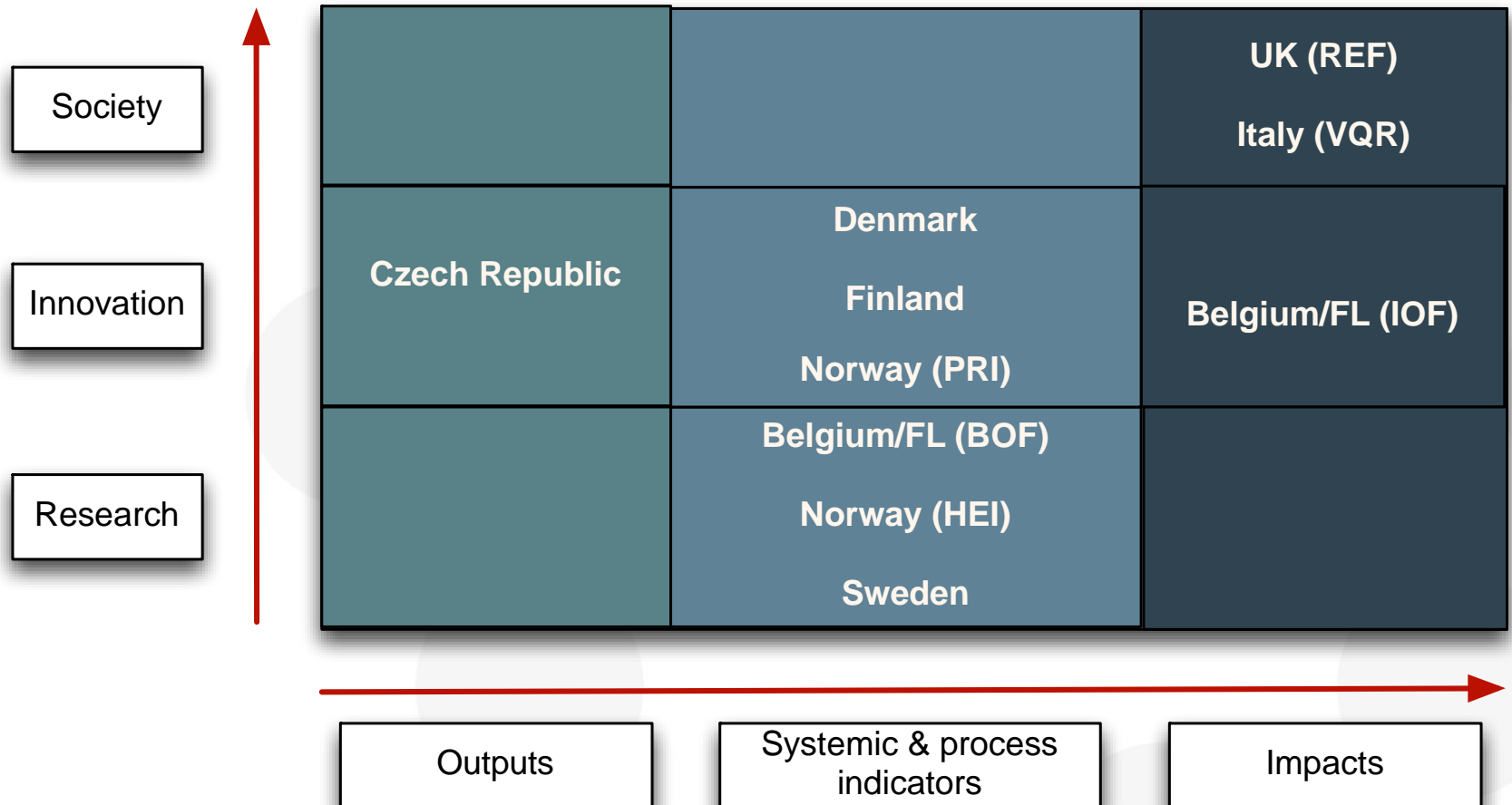
- **All research funders (and therefore evaluators) struggle with this**
- **Interdisciplinarity is important**
 - Emergence of new disciplines and approaches
 - Ability to do Mode 2 work, addressing industrial and societal challenges
- **But it's hard to measure**
 - Discipline redefinition implies changes in the criteria relevant for measurement
 - Interdisciplinarity requires use of expertise and indicator categories that cross existing boundaries
 - For example, multiple peer panels
 - In metrics-based approaches, the interesting things start happening in undefined or 'other' categories
 - As with any change, there is often professional disagreement about standards, criteria, approaches

Different indicators, different purposes



	Research productivity	Research quality	Relevance of research	Efficiency/ value for money	Quality/ sustainability of national research systems
Input Criteria		X	X	X	
Systemic indicators			X		X
Process indicators	X		X		X
Research outputs	X	X	X		
Impact indicators			X	X	

Recent Czech focus on outputs is unusual





Ratings tend to use 4-5 points

Score	Meaning	Research quality	Relevance to society	Viability
1	World leading/ excellent	The research unit has been shown to be one of the few most influential research groups in the world in its particular field.	The research unit makes an outstanding contribution to society.	The research unit is excellently equipped for the future.
2	Very good	The research unit conducts very good, internationally recognised research.	The research unit makes a very good contribution to society.	The research unit is very well equipped for the future.
3	Good	The research unit conducts good research.	The research unit makes a good contribution to society.	The research unit makes responsible strategic decisions and is therefore well equipped for the future.
4	Unsatisfactory	The research unit does not achieve satisfactory results in its field.	The research unit does not make a satisfactory contribution to society.	The research unit is not adequately equipped for the future.

Example: Netherlands

Implementation issues for peer/panel approaches



- **Robustness**
 - Clear and universal assessment guidelines
 - Professional management and support organisation
- **Fairness**
 - Clientelism and nepotism are avoided
 - Scholarly bias is tackled
 - Appropriate representation
 - Interdisciplinarity is addressed
 - Consistency of assessments
 - Transparency of the process
- **Cost**

The Evaluation Methodology





The key principles

- **The EM**

- Reflects the **strategic policy objectives** for the R&D system
- Functions = to act as source for **strategic information** & directly inform public **institutional funding** for research organisations

- **The evaluation**

- Is at the level of field-defined **Research Unit** (RU) within an Evaluated Unit, i.e. a research organisation or in the case of the public HEI, a faculty
- Is a process of **informed peer review**: metrics inform but do not substitute for judgment
- Covers **all research organisations** of a critical size, on a voluntary basis
- Is a **fair and egalitarian** system: a single framework for assessment while allowing for a reasonable level of field- and RO typology-specific variations
- Is **comprehensive**: all dimensions of the research activities and its outputs, outcomes and impact
- The **minimum possible** cost and burden to deliver a robust and defensible process



The role and function of the EM

- The key function of an evaluation system is to **support** public R&D governance in the attainment of its strategic objectives, i.e.
 - To strengthen **R&D capacity**
 - To foster **excellence** in research
 - To foster research alignment with the **societal needs**
 - To support the **growth and competitiveness** of the Czech Republic
- **By:**
 - Assessing past research performance
 - Supporting future performance improvement
 - Provide **strategic information** for policy making and R&D management
 - Provide **incentives** for positive change in the R&D system reflecting the strategy

How does the EM provide incentives for positive change?



- **Indicators** used & their structuring into a set of **assessment criteria** is guided by the **policy objectives**

Institutional management & development potential

Quality adequacy of the physical research environment

Research strategy & management (including HR management)

Membership of the (world) research community

(International) research presence & collaboration

Scientific research excellence

The quality of the selected 'best' outputs

Research performance

Research productivity

Research capacity building (PhD)

Overall quality of the research activities

Societal relevance

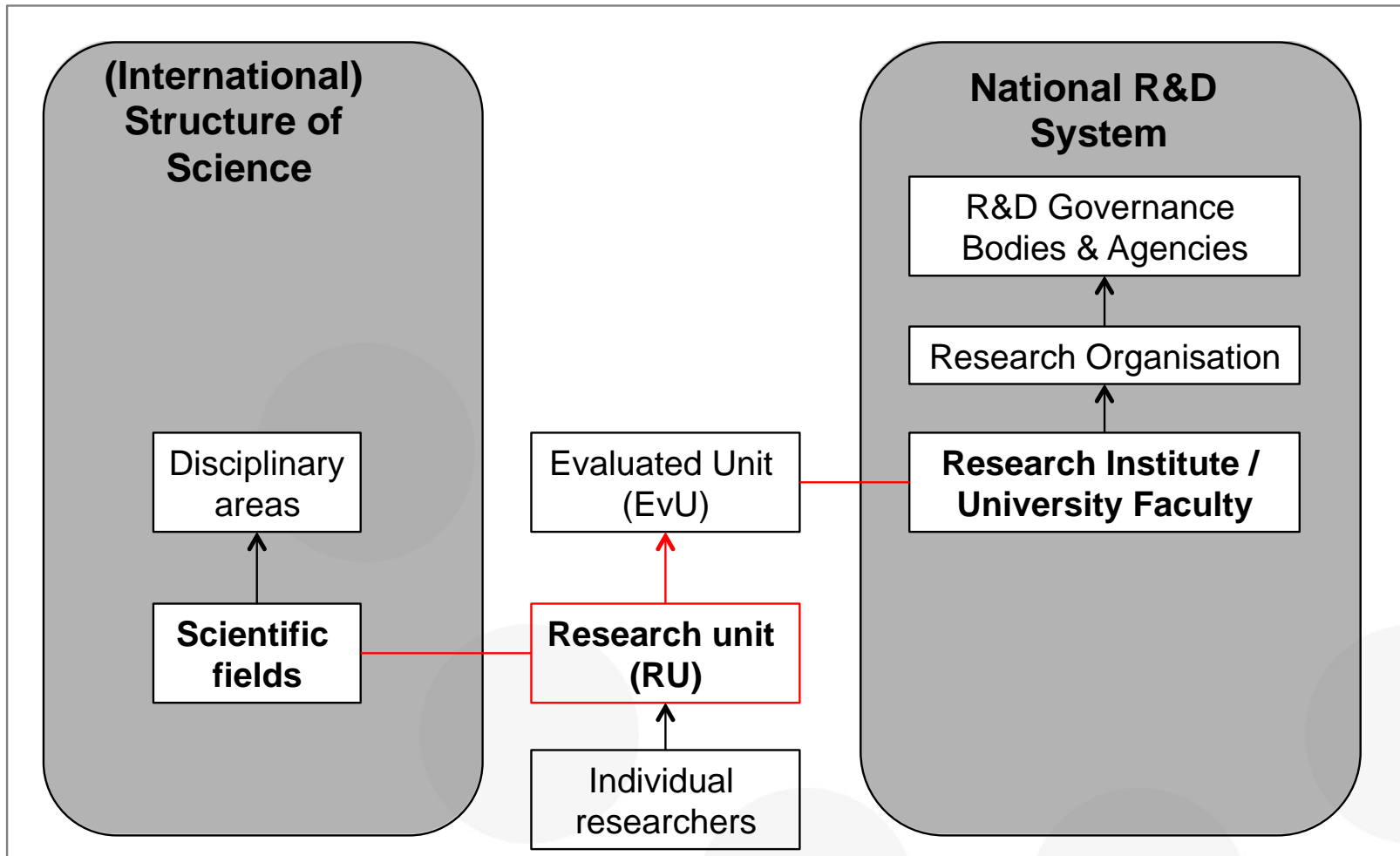
Importance of (potential) societal impacts

How does the EM act as a source of strategic information?



- The EM is **not** an arithmetic exercise → **quantitative & qualitative** input & evaluation results
- **Quality** information thanks to a **detailed** and **comprehensive approach**
 - Focus on the evaluated actors' role, positioning, and competitive value in the **national** R&D and innovation system as well as in the **international** R&D landscape
 - **Strengths and weaknesses** of the different actors against all dimensions of the research activities, i.e. the research quality, research strategy, the research environment (i.e. the institutional conditions), the research outputs, outcomes and impacts
 - It will allow for the identification of the factors upon which action is needed in order to **improve** research performance, at the national as well as institutional level.
- **Primary unit of analysis** = the elements that worldwide constitute the fundamental structure of research, i.e. **scientific fields** → **Research Unit**

Matching the scientific with the institutional dimension





A process of informed peer review

- Based on a **mix** of appropriate **quantitative and qualitative** data to support their professional judgement
 - international bibliometric data, data from the RD&I system, quantitative and qualitative data provided by the evaluated Research Units through self-reporting
- **Expert panels** allow for in-depth consideration of **differences in scientific fields and sub-fields**
 - Type of outputs, channels, productivity, costs, collaboration, internationalisation ...
- Strict rules govern **conflicts of interest**, against nepotism and 'clientelism', combined with **auditing mechanisms** and rules for punishing cases of **fraud**
- The evaluation will be conducted in **full transparency**
- Evaluation results: **quality scores & explanatory texts, conclusions & recommendations**

Covering all research organisations in the CR of a critical size



- **A broad & varied R&D base:**
 - HEIs, public research institutes, state organisations, private research organisations, organisational units of the CR, non-profit organisations, and associations of legal persons
- **Categorisation based on their mission for society related to research:**
 - **Scientific Research Organisations**, including the scientific research institutes, HEIs and university hospitals
 - **RTOs**, i.e. research organisations that have as *primary* function to provide knowledge transfer services to the industry sector.
 - The Aerospace Research and Test Establishment, Research Institute of Building Materials, etc
 - **Public Service Research Organisations**, i.e. research organisations that have as primary function to provide services to government or society.
 - Research Institute for Labour and Social Affairs, the Institute for International Relations, the Czech Metrology Institute, etc.
 - **Infrastructure Research Organisations**, providing infrastructure & information to research
 - CESNET, the National Library of the Czech Republic, etc.
- **Critical size:** 50 research outputs in 5 years – the minimum for robust data analytics

Research outputs taken into account



	Threshold	Scientific research excellence	Other indicators
Scholarly outputs			
Papers in peer-reviewed journals (J)	X	X	X
Conference proceedings (D)	X	X	X
Monographs, books and book chapters (B), <i>provided they are identified with an ISBN number</i>	X	X	X
Non-traditional scholarly outputs			
Results used by the funding provider, i.e. projected into legislation or norm, projected into non-legislative or strategic documents (H)	X		X
Research report containing classified information (V)	X		X
Certified methodologies, art conservation methodologies, specialized map works (N)	X		X
Patents and other IP			
Patents and patent applications (P)	X		X
Plant/ breeders rights (Zodry & Zplem)	X		X

Handling interdisciplinary research



- Interdisciplinary research **between sub-fields**: handled by **1 subject panel**
 - Facilitated by the categorisation of scientific disciplines in broad fields based on the OECD FOS (internationally recognised): 6 disciplinary areas – 36 fields – 190 sub-fields (e.g. Medical & Health sciences – Basic medicine – Neurosciences)
- Interdisciplinary research **within a disciplinary area: cross-referrals** among the subject panels in 1 disciplinary area
- Interdisciplinary research **across disciplinary areas**: application for **Interdisciplinary Research Unit**
- **For cross-referrals & Interdisciplinary Research Units:**
 - At least 30% of research activities take place across disciplinary areas
 - “Proven” through research outputs/bibliometrics or the scientific background of the researchers in the unit
 - Decided upon by main panel & subject panel chairs

Using a fair and egalitarian system



- A **single framework for assessment** across all disciplines and types of research organisations
 - The same assessment criteria are used for all research organisations & fields
- Allows for **full comparability** of the evaluation results, independently of the scientific fields and wherever the research is conducted
 - Based on a detailed Evaluation Protocol, setting common procedures and providing standard definitions
 - The main panel chairs act as auditors and guarantors for the comparability
- A first step in the evaluation process is the **calibration exercise** related to the assessment criteria
 - **Field-specific definition** of the terms: originality, rigour, significance, and reach
 - **Weighting** the importance of sub-criteria for the field and for the different types of research organisation

Assessment criteria & indicators (1)



- **Institutional management & development potential**
 - Quality adequacy of the research environment
 - Research infrastructure, use of national/international RI, shared or collaborative use
 - Research capacity: funding profile (institutional, competitive, contract; national/international), staff profile & size (including HC versus FTE)
 - Research strategy & HR management
 - Quality of the research strategy: strength of the research plan (long/short term, objectives & activities, adequacy, feasibility)
 - Quality of processes for career development (objectives, competency framework, appraisal criteria, frequency performance reviews, etc), policy & practice concerning PhD & postdocs (support & supervision, training, etc)
- **Membership of the national & global research community**
 - National & international
 - Intensity & quality of collaboration & partnerships, co-publications, international mobility
 - Reputation & esteem

Assessment criteria & indicators (2)



- **Scientific research excellence: peak quality**
 - In terms of **originality**, **significance** and **rigour**
 - Two-stage assessment process:
 - Review results (highly selected scholarly outputs)
 - Panel assessment: aggregating review results + citation impacts: nr and % publications among the top 10%, and 25% most cited publications (world, EU28)
- **Research performance: overall quality of the research activities**
 - Research productivity: outputs/FTE researchers
 - Contribution to capacity building: PhD students/awarded & trained
 - Overall quality of the research activities:
 - Publishing profile, citation impacts, publication channels, etc (incl. relative data on positioning in field)
 - Overall competitiveness (funding, strategy & alignment trends, infrastructure etc)
 - Narrative on main value for research

Assessment criteria & indicators (3)



- **Societal relevance**
 - In terms of **reach** & **significance**
 - Knowledge & technology transfer activities: intensity & quality
 - Income from societal relevant research (competitive & contract research)
 - Income from commercialization activities
 - National collaborations and partnerships beyond academia
 - Reputation and esteem beyond academia
 - Non-traditional scholarly research & IPR-related outputs (national/international)
 - Societal impact: narratives



Limiting the costs

- **Sharp focus on key information needed only**
- **Maximal use of the information available in the RD&I IS (IS VaVal)**
- **Remote review and remote assessments, complemented with panel meetings**
- **Limited number of subject panels & panellists**
- **No site visits**

- **The higher the level of sophistication, the higher the cost**



METODIKA

Děkujeme za pozornost!

www.metodika.reformy-msmt.cz



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ