

Moving from Studies to Streams: A More Radical Way to Avoid Floods of Evidence by Channeling Evaluative Efforts in Conditions of Complexity

Tom Ling

Senior Research Leader, RAND Europe

Background: Multiple evaluations of complex problems risk creating “floods” of evidence. Unstructured and unmanageable evaluative findings are hard for decision makers to use, and do not support long-term learning. Using the metaphor of “channeling” evidence to create more manageable streams of evidence, it is argued that the evaluation community can create more usable and useful knowledge. To extend this metaphor, the fenlands of England were drained by channeling water into canals to create land used today for homes and farming; this chapter is concerned with the active steps the evaluation community can take to transform the marshy fenlands of unstructured evidence into navigable and useful canals. “Downstream” systematic reviews and clearinghouses are not enough to address this issue. The problems arise “upstream” from multiple, weakly connected evaluations being conducted in relative isolation, rather than focusing evaluative efforts on the key challenges facing complex policy spaces.

Purpose: It is argued that evaluation policies need to move “upstream” and be consciously designed to both build on existing evaluations and take the needs of decision makers more fully into account. This cannot be achieved by “one-off” improvements to individual evaluations but requires a coherent response that would support effective and meaningful retroductive realist syntheses that reveal the

underlying patterns found in evaluations of heterogeneous interventions.

Setting: Evaluations are almost never at the heart of policy debates today, and the evaluation community should at least be curious about why this is the case. It is argued here that poorly used and unfocused evaluative efforts require “canalization” to replace floods of evidence with coherent sense-making.

Intervention: Not applicable.

Research Design: Review and reflection.

Data Collection: Not applicable.

Analysis and Findings: Urgent action is required across the community of evaluators, commissioners of studies, and users of evaluative evidence to make better use of the resources dedicated to evaluating complex interventions. Those who commission, deliver and use evaluations should work together to ensure that individual evaluations are structured to support retroductive studies and allow coherent assessments of the major challenges facing each field of study, while taking into account the needs of different interest-holders.

Keywords: *evaluation policy; evidence; complexity; sense-making; retroductive realist synthesis; learning; accountability.*

Nicoletta Stame's introduction to *From Studies to Streams* notes that the book title draws on a metaphor (Rist and Stame, 2006). And it is a very productive metaphor. In her introduction Stame suggests that streams might converge on a single result (giving us a coherent set of conclusions) or they may allow for heterogeneity (with the possibility to explore uncertainties and promote organizational learning through a range of more-or-less overlapping insights). The first might suggest "best practice" and the second "good practice." Both could be associated with "learning organizations," but what they are hoping to learn might be subtly different.

In this article I will take the metaphor a little further by suggesting that streams may be more or less "channeled" to shape evaluative efforts to support learning and inform policy especially where it is not possible to identify a unitary best practice. This implies a conscious and strategic effort to shape the way in which evaluative knowledge in a field is developed and used. We need channels to avoid floods and swamps, where the evidence has little coherence and where it is hard to make sense of the direction in which the evidence flows; we need channels so that decision-makers might navigate the knowledge in a meaningful way. Without this we accumulate facts, but we become none the wiser.

It is this "channeling" metaphor that will shape this short contribution; how can the constituencies with an interest in evaluations channel evaluative evidence so that it creates streams of usable knowledge that support better decisions and strengthen learning? And by "better" decisions we mean more inclusive, more comprehensive, and more sustainable decisions that reflect the whole range of available evidence and not just one (potentially narrow or exceptional) study based on one particular time and place.

Ray Rist understood the value of knowledge that drew upon multiple sources, and we concur with the idea that administratively required, narrowly focused monitoring needs to be expanded to support learning. This involves shaping evaluative activities, so they cast light on the challenges and problems being faced by interest-holders. At the same time this requires the independence and rigor in the evaluation process to ensure that, when necessary, unwelcome and difficult lessons are communicated clearly and unambiguously. In short, this involves an intentional and planned approach. In recent decades two important approaches to creating channels of evaluative and research knowledge have been systematic reviews and clearinghouses. However, these are important but insufficient.

Why Systematic Reviews and Clearinghouses Are Not Enough

Stame and Van den Berg, in their articles in this volume, remind us that we should take care about learning from one-off studies. They may be adequate for the task given to the evaluation team, but they may be limited and partial in scope or may merely repeat earlier findings. Sometimes they may deliver transformational findings, as Van den Berg points out. But in complex and dynamic circumstances, they may often not support results-based delivery, which requires a continuous flow of data that can respond to a changing environment and an emergent intervention. As (almost inevitably in complex interventions) results vary over time, the lessons learned must reflect these new data.

One increasingly important mechanism for integrating and synthesizing knowledge derived from individual studies is systematic review. This has been a notable addition to knowledge creation, management, and mobilization in recent decades. Similarly, the significant growth of clearinghouses as repositories of information about the effectiveness of social interventions also indicates a concern to combine findings from multiple studies. The focus has often been on how far clearinghouses meet open science standards in their reviews (Mayo-Wilson et al., 2022). Not meeting standards of replicability, public availability of results, conflicts of interest, study registration, etc., would be a legitimate source of concern but this debate misses the point that not all evaluations can (or should) be channeled by aggregating studies. Systematic reviews and clearinghouses are "downstream" and take place after the evidence has been conceptualized, categorized, collected, and analyzed. Other such efforts include realist synthesis and ethnographic meta-analysis. What is also needed for these efforts to succeed is "upstream" work to create evaluative evidence fit to be integrated and synthesized. This could include identifying also gaps and weaknesses in the evidence. In the following section we consider what larger agencies (such as UN agencies) would need to consider if we were to channel individual studies to coherent streams of evidence that support learning and better decision-making. We then briefly extend this logic to wider system change to achieve this outcome.

How Might Large Agencies Move to Supporting Streams That Promote Learning?

To help understand how this might be achieved, and what benefits might flow, I will draw upon my experience of supporting a UN agency with its evaluation policy. These are my observations (and do not necessarily represent the views of the agency or the final report). Although based on one case, these observations may be relevant to many organizations conducting multiple and heterogeneous evaluations. This includes organizations with “portfolio programs” which have shared aims but different means to achieve them (for example, supporting girls into education).

One challenge facing efforts to channel studies into streams is that evaluation systems respond to several well-known drivers and barriers associated with, among other things, how evaluation needs are identified, ITTs drawn up, contracts awarded, and evaluations managed. The mode of producing evaluative knowledge is always both socially constructed and shaped by institutional and social power. We are not suggesting that these power relationships can be removed; only that they could be used more productively and creatively for all interest-holders concerned.

In the continuously evolving field of international development, these drivers and barriers have a specific shape. First, there is considerable concern to protect the independence of the evaluation function (indeed, to emphasize this, the institutional home of evaluation most often includes the word “independent”). However, too often being “independent” can mean being detached from programmers, rights-holders, and duty bearers. The answer is not to strengthen the independence of evaluators further in relation to the rest of the organization, but to re-think the relationships that exist among the various parts of the agency—to “channel” the evidence from evaluations toward delivering benefits.

Second, an opportunity to build channels results from establishing Sustainable Development Goals (SDGs). It is often asked when commissioning evaluations how well specific programs and projects contribute to these goals.

However, in practice, the temptation is to create a logic model which “shoehorns” SDGs into the theory of change, lacking a rich description of evaluable causal pathways. Instead, most often evaluations should aim to include a detailed understanding of how the intervention contributes to intermediary outputs which, based on wider evidence from outside the program evaluation, can reasonably be supposed to contribute to SDGs.

Almost inevitably, these programs and their intermediary outputs will be very varied, as will the evidence they help create. This is seen to be a barrier to creating streams of evidence from the viewpoint of systematic reviews and clearinghouses. However, from a realist perspective such heterogeneity provides the opportunity to build lessons based on understanding how different programs perform in different contexts (but the conclusions may not meet the standards of generalizability established by clearinghouses). This requires studies that are not only realist in orientation but provide sufficient methodological and descriptive detail to allow causal processes to be revealed through the patterning of data in several heterogeneous interventions (an approach referred to as “retroductive”¹ in the realist literature (The RAMESES II Project, 2017).

Third, it is necessary to socialize the idea that one purpose of evaluation is to contribute to *improvement* rather than to *prove* that one approach is more successful than another. Therefore, the case for channeling evidence is linked to the case for promoting learning as the route to improvement. This requires a culture of curiosity, openness, and at least a willingness to entertain the possibility that existing practices could be improved upon.

Fourth, in theory there are infinite lessons to be learned and things to evaluate. There is a constant risk that we create not “channels” but uncontrolled “floods” of evidence. Instead of well-crafted lessons we may create only random insights. Building streams requires organizations to institutionalize processes through which to identify evaluable questions which, if answered, would make a significant difference to the lives of rights-holders and/or would help to improve the performance of duty-bearers in relation to protecting these rights. This process should be structured around understanding where there are barriers and

¹ The ambition of those using retroduction in evaluation, is to understand how policies or programs produce observed outcomes in the particular circumstances where they were implemented. This involves identifying patterns, and testing different explanations, using both inductive and deductive reasoning. If the evaluation

reports being used in a retroductive review of multiple studies include insufficient information about how data was collected, how concepts were constructed and operationalized, and the contexts within which results were achieved, then they cannot helpfully contribute to a retroductive review.

challenges in delivering the organization's theory of change. Within this process, evaluations could then be independently conducted in ways that make retrodution feasible. All of this would require senior management to be highly engaged in the evaluation activities of their organisation. It would also require a market of evaluation providers who are well versed in this approach and have the capacity to deliver it. Strengthening historically weak approaches to knowledge management would also be needed.

Streams and Systems-Level Interventions

Evaluations are events within a complex system (Hawe et al, 2009). This system includes the thing being evaluated (which the evaluation is intimately connected to) and the systems within which evaluations are constructed, funded, managed and delivered (which provide the wider context of evaluation as a practice). Evaluations as events within complex systems have also changed over time, both in what is being evaluated and how evaluators go about their business. Evaluators are, for example, more likely to be asked to evaluate efforts to achieve transformation at scale (Ling, 2017) or sustainability (Ling and Millard, 2024). Such evaluations themselves tend to be flexible and agile; they are a changing part of a system which is itself in flux. Systems thinking has most often been applied to understanding the dynamics between the evaluation and the evaluand. In this article the focus is on the relationship between the individual evaluation, and the evaluation system within which it sits. This includes repositories of "what works," and clearinghouses in general, which may be substantively focused but often engage with, and inform, evaluation practices. Consequently, we are gaining a better understanding of how evaluation systems can be oriented towards evidence-based policy making.

However, in arguing for evaluation systems that inform decision-making, there should be an understanding of the risk that evaluation usurps other social functions, including the role of democratically elected policy makers, implementing teams, professions, and service users and citizens. So, we must consider the appropriate boundaries that give sufficient space so that evaluation practices can be useful in informing the democratic space without dampening democratic preferences and concerns.

Building and maintaining the channels to shape evidence necessarily involves building new relationships and novel forms of social capital

across multiple interested parties. For example, the What Works Centres in the UK have approached this through developing their own networks as well as collaborating with the Economic and Social Research Council's Parliamentary research leads, knowledge exchange managers, local policy innovation partnerships, policy fellows, and seminar series, and engaging the All-Party Parliamentary Group for Social Science and Policy. The evaluation of this, conducted by Frontier Economics (2022), is broadly positive, but because it is based on the What Works Centres' own linear logic models rather than systems-based evaluation questions, important questions are missed. In particular, it would be hard to demonstrate that the period since What Works Centres were created in 2013 coincided with more evidence-based policy-making in the UK. As we create channels of evaluative evidence, the intersection of evaluation system and democratic processes will always be a place where tensions need to be resolved – the question is how to make these tensions productive rather than destructive.

Conclusion

The need to move away from stand-alone evaluations and toward streams of evidence is compelling and has gained ground since Rist and Stame, (2006) first wrote about this. It remains a fertile way of thinking about how to organize evaluative activities to maximize public benefit. As discussed in this chapter, channels should be curated to promote learning and improvement, involve multiple stakeholders, and ensure that individual evaluations are designed to contribute to a growing understanding of the patterns of change achieved by programs. How evaluative evidence is channeled will be critical to strengthening learning and informing democratic policy-making.

References

- Frontier Economics. (2022). *ESRC investment in what works centres: Evaluation report of the ESRC*. <https://www.ukri.org/wp-content/uploads/2022/08/ESRC-110822-InvestmentInWhatWorksCentresEvaluation.pdf>
- Hawe P, Shiell A, Riley T. 2009. Theorising interventions as events in systems. *American Journal of Community Psychology* 43: 267–76.
- Ling, T. (2017). Achieving equality at scale through system transformation: Evaluating system

- change. In K. Forss (Ed.), *Speaking justice to power*. Routledge.
- Ling, T. and Millard, A. (2024). 'Evaluating sustainability – a multidimensional approach' in Lindkvist I, Bastoe PO and Forss K (2024) *Towards Sustainable Futures: The role of Evaluation* Routledge New York (Open Access available)
<https://doi.org/10.4324/9781032714936>
- Mayo-Wilson, E., Grant, S., & Supplee, L. H. (2022). Clearinghouse standards of evidence on the transparency, openness, and reproducibility of intervention evaluations. *Prevention Science*, 23(5):774–786.
<https://doi.org/10.1007/s11121-021-01284-x>
- Rist, R.C. & Stame, N. (Eds.) (2006). *From studies to streams: managing evaluative systems*. Transaction Publishers New Brunswick, N.J., London, UK
- The RAMESES II Project. (2017). Retrodution in realist evaluation. *The RAMESES II Project—Resources and training materials for realist evaluation*.
https://www.ramesesproject.org/media/RAMESES_II_Retrodution.pdf.
- United Nations Development Program. (2024). *Knowledge management*. Retrieved at: <https://www.undp.org/tag/knowledge-management>