

Research vision

– DESTinY –

Digital Evidence Synthesis Tool Innovation Yielding Improvements in Climate & Health

This project vision explains the rationale, and describes the overall structure of the work program of this consortium. Additional details on project impacts, policy and practice ambitions, public engagement, team composition, and outputs are provided elsewhere in the proposal.

1 Background

Climate change impacts already cause widespread human suffering affecting millions of people across the globe [1, 2] and health risks will further escalate with continued warming [2]. Urgent climate change mitigation and adaptation—often associated with substantial health co-benefits [3–5]—are imperative to protect the health of future generations and to ensure climate-related health hazards stay within the capacity of health systems [2].

Decision makers urgently need access to timely knowledge to take impactful, evidence-based, health-centred climate action. We realise this through the co-production with evidence producers and tool developers. Living evidence provides a radical new model to rapidly fill relevant evidence gaps [6–8]. However the vast and fast-growing evidence base poses fundamental challenges to conventional (usually manual) synthesis methods [9].

Digital Evidence Synthesis Tools (DESTs) have been heralded to make the evidence synthesis process faster and cheaper without compromising quality [10]. However, the last 20 years of research and application [11] have not lived up to this promise. Advances have been marginal, limited to individual tasks, and there has been a lack of attention to researching their safe and responsible use at scale [12, 13]. Recent fundamental advances in artificial intelligence, particularly large-language models, promise a step-change across a broader range of complex evidence synthesis tasks [14, 15]. The next generation of DESTs require re-defined artificial-intelligence-powered human-machine interactions that can be applied safely and responsibly without compromising methodological standards and trust [16]. Co-production is critical to effectively address the needs of decision makers in low- and high-income countries.

2 Project vision and approach

DESTinY will co-develop a new generation of digital evidence synthesis tools (DESTs) and showcase their transformational power for the delivery of rigorous living evidence in climate and health that matters to policymakers and other evidence users. This defines who we work with, how we work, and the technology we use to make evidence synthesis dramatically more useful. In particular, our project will:

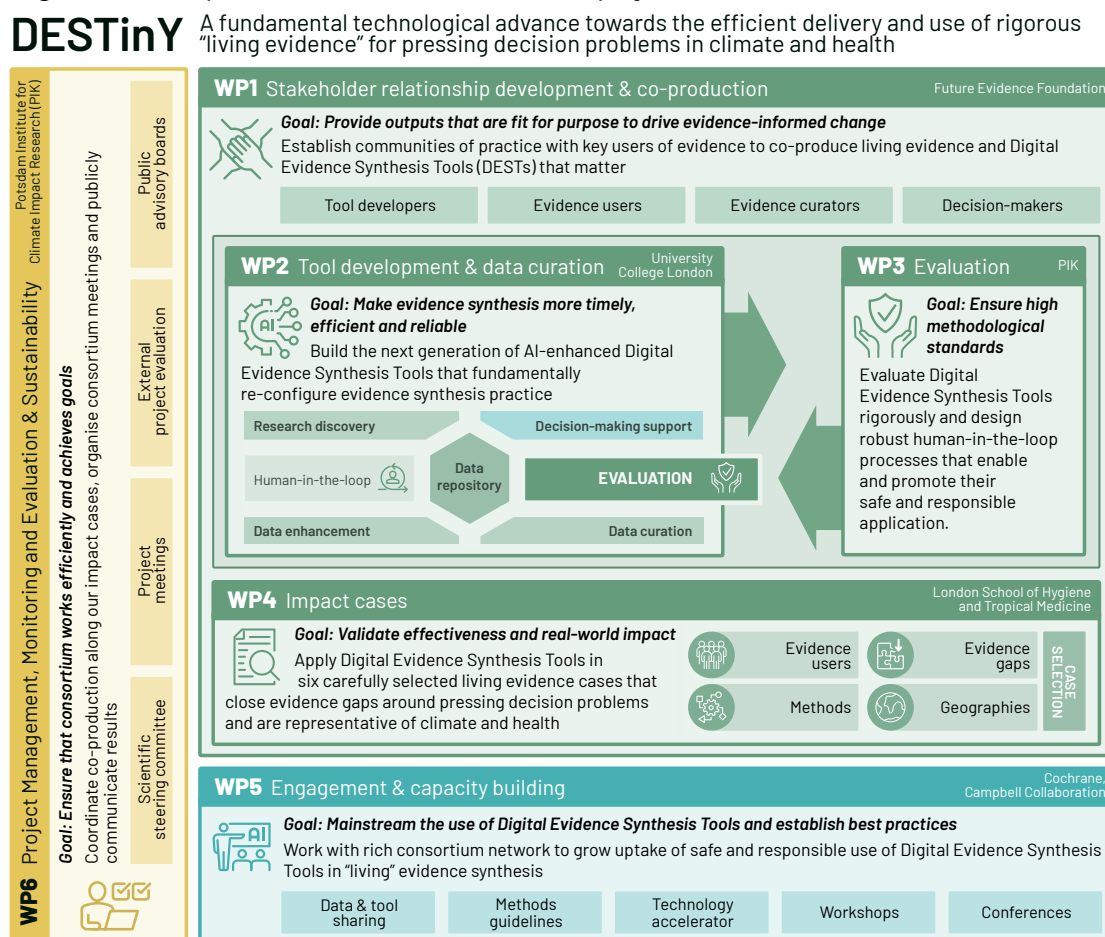
- **Leverage recent AI advances** to develop new DESTs, enhancing evidence synthesis by reconfiguring human-machine interactions (**WP2**);
- **Engage in DEST evaluation** to design safe and responsible applications without eroding methodological standards (**WP3**);
- **Build communities of practice** with decision makers around the globe and across scales to ensure DESTs are fit-for-purpose, work for all and are applied to six impact cases that matter (**WP1, WP4**); and
- **Support users, producers, and funders** of climate and health evidence synthesis to establish best practices (**WP5**).

Technology is transformative through its users. Our co-production processes use the proven Alive model and builds on decades of experience from our global consortium, including leading climate and health teams (LSHTM, PIK, ASCEND), DEST developers (PIK, UCL, FEF), and collaboration experts (FEF/ALIVE, ASCEND, eBASE). We co-produce evidence and DESTs with urgent needs, including international organisations, national and local governments as well as NGOs (**WP4**).

3 Work packages

Figure 1 provides an overview of how we will deliver the overarching goals of the project (see Section 2), structured by six work packages. The subsequent sections explain each work package in more detail.

Figure 1: Conceptual overview of the DESTinY project



WP1: Stakeholder relationship development and co-production

Institutions: Future evidence foundation (FEF) (*all partners*)

Aims and activities

Co-production is at the heart of the project. We will use the Alive (Alliance for Living Evidence) partnership model (Figure 2) to ensure DEST development is responsive to the needs of users involved in driving evidence-informed climate and health policy and action—particularly policy-makers and their advisors, and affected and vulnerable communities. This will ensure the next generation of DESTs maximise their positive impact on human health and climate.

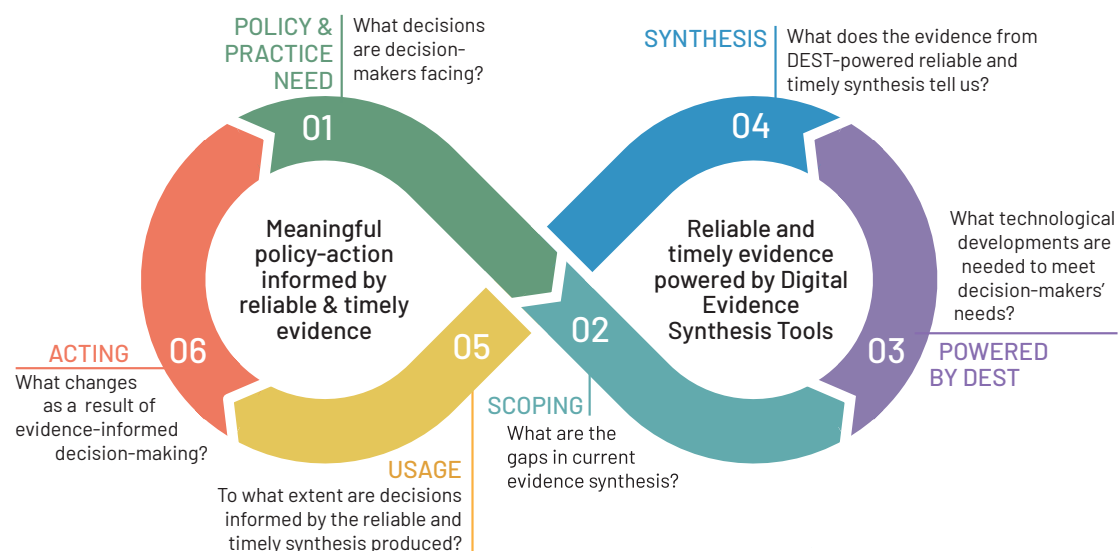
We will convene partnerships between decision makers, advisors, intermediary organisations and evidence synthesis groups, supporting them to work together to identify, synthesise and engage with living evidence for climate and health decision-making. We will facilitate ongoing evidence-informed discussions between evidence users that will build engagement and alignment around effective climate and health action. This places evidence users at the centre of tool development (**WP2**) evidence generation (**WP4**) and real-world impact (**WP4**).

This work will be supported by the long-standing relationships with regional and local decision-makers of our consortium members LSHTM, eBASE, ACRES and ASCEND, combined with experience in brokering actionable evidence for impactful decision-making.

Figure 2: Graphical summary of the Alive model

ALIVE MODEL - PROVIDING RELIABLE & TIMELY EVIDENCE FOR DECISIONS

The Alive Model will play a key role in ensuring decision-makers' needs are met through reliable and timely evidence synthesis



Deliverables / Milestones

- **Living evidence partnerships** between key groups affected by or able to influence climate and health decision-making, evidence synthesis providers and DEST developers.
- **DEST and living evidence priorities** of key groups that can guide development of both the next generation of DESTs and the evidence needed for prioritised climate and health decision-making.
- **Decision-maker engagement** with continuous cycles of prioritised living evidence outputs.
- **Enabling environment for evidence-informed decision-making** driving better decision-making to protect health and climate, and creating the conditions for aligned action and enhanced impact.

WP2: Development of tools and curated data

Institutions: University College London (UCL) (PIK, FEF, ACRES)

Aims/activities

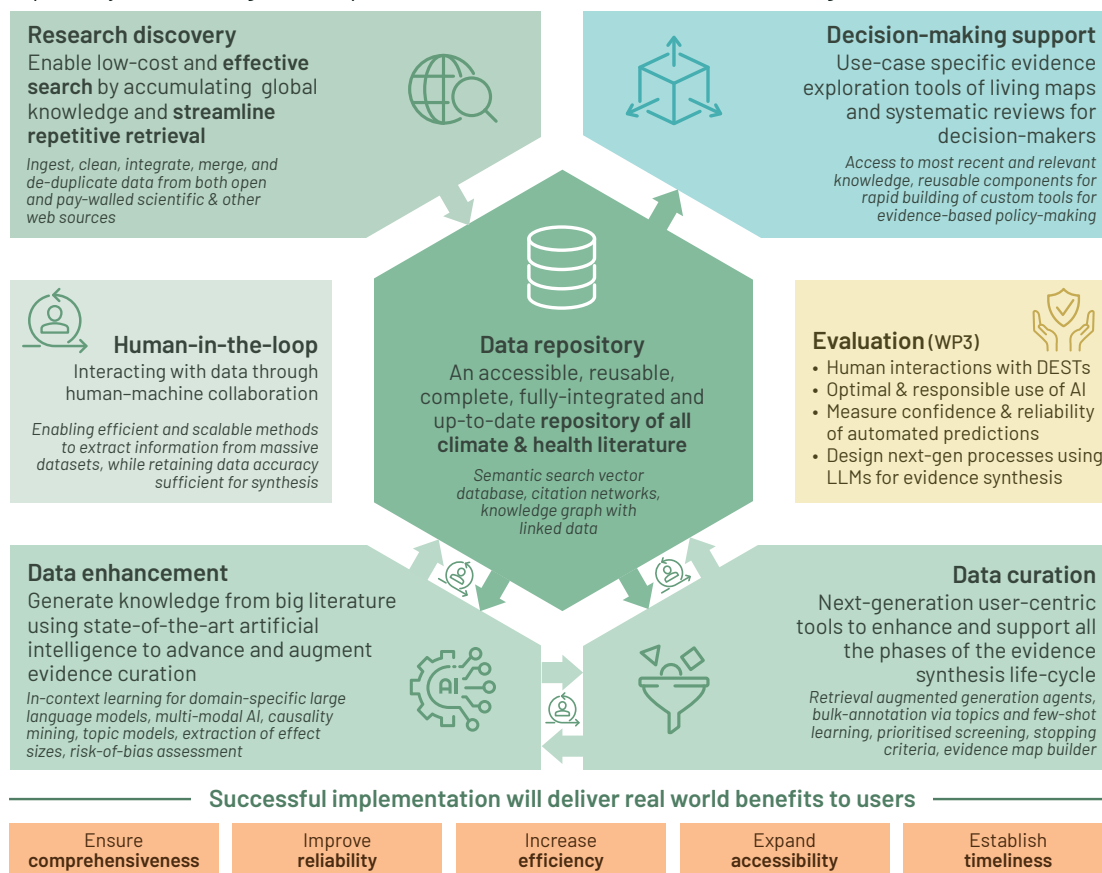
Tool development will be driven by an overriding objective to deliver a step-change in making evidence synthesis faster, cheaper, timelier, and more useful—not just small incremental improvements to existing tools. We will prioritise the automation of complex and resource-intensive tasks, including study discovery, data extraction and harmonisation, critical appraisal, and synthesis. Outputs will be modular, open and FAIR (through APIs, code, and apps) to ensure innovations are shared widely, drive adoption by third party tools and catalyse growth in the research and development of machine learning for evidence synthesis and use. All tool and method development will be conducted within robust research and evaluation frameworks (e.g. ‘studies within a review’ [17]) in order to build a cumulative evidence base for tool development and use.

WP2 systematically addresses **four key bottlenecks** that inhibit the production and use of climate and health evidence in decision-making, driven by our extensive expertise [18–28].

Figure 3: Conceptual overview of the proposed technical infrastructure.

WP2 BUILDING NEXT-GENERATION TECHNOLOGIES TO IMPROVE EVIDENCE SYNTHESIS

Accelerating evidence synthesis with interoperable state-of-the-art tools around an integrated data repository, establishing their responsible use (evaluated in WP3), and delivering real-world benefits



First, bibliographic data are fragmented across many databases. This wastes limited resources constructing, translating and running searches across databases, and deduplicating results. The financial costs of accessing commercial databases and the complexity of the required search Strategies entrench inequity in a fundamental step in evidence synthesis. We will create a **research discovery** service that feeds a living, open **data repository** based on OpenAlex. This repository will have a conventional text index and a vector index from a large language model (LLM) fine-tuned for retrieval tasks in climate and health [29]. We will compile gold standard datasets to evaluate the coverage of the repository (adding additional sources including the web and full text documents when needed) and the efficiency of the semantic vector search [30] to find relevant research for evidence synthesis [31].

Second, huge resources are currently devoted to screening records, critical appraisal and data extraction [32]. We will facilitate a step-change in efficiency by building **data enhancement** services into the repository. These will continually and automatically classify new research using supervised, unsupervised and generative approaches (including topic modelling). Retrieval Augmented Generation (RAG) workflows [33] will be built to extract qualitative and quantitative data, and undertake critical appraisal. Research priorities will include selecting and fine-tuning an LLM for accuracy within this domain.

Third, where data are ambiguous, non-standardised or semantically imprecise, we cannot rely on automation alone to deliver the level of accuracy and interpretation needed for synthesis. We will therefore develop **data curation** tools and human-in-the-loop workflows. These will maximise the efficiency of human-machine collaboration for sense-making and data harmonisation [34]. The data enhancement and curation workflows will prepare data for synthesis and include quantitative estimates of uncertainty in machine predictions.

Fourth, evidence and communication tools are often not well adapted to the needs of decision makers. We will therefore build and customise **decision-making support** tools. These will support evidence mapping and exploration, translation, re-analysis, and re-contextualisation by decision makers and their advisors.

Technology development will be in collaboration with decision-makers and affected communities facilitated by **WP1**, integrated with evaluation work led by **WP4** and engagement and support capacity building led by **WP5**.

Deliverables / Milestones

- **Data repository** of climate and health research. A comprehensive collection of research outputs, including FAIR data extracted and structured using the services described below.
- **Research discovery** workflows that use OpenAlex and other sources identified through ongoing evaluation
- **Data enhancement** using LLMs and RAG workflows for specific tasks
- **Data curation** using human-in-the-loop tools and workflows to ensure data is of sufficient quality for evidence synthesis
- **Decision-making support** tools to enable decision-makers and others to engage with and use living evidence.

WP3: Evaluation

Institutions: PIK (*UCL, ASCEND, FEF, Campbell, Cochrane*)

Aims/activities

High quality evaluation is central to this project. Building on previous work [35–38], **WP3** delivers research that delineates the (moving) frontier of what evidence synthesis tasks can be responsibly automated with which digital technologies. Through evaluation research, we explore the design choices involved in deploying LLMs in next-generation DESTs [39], informing the development of the tools we build. **WP3** involves evaluating how well the tools work, as well as how researchers work with tools. If DESTs are to be used safely and responsibly without eroding methodological standards [40], human-in-the-loop processes to evaluate AI systems are vital. This work package designs human-in-the-loop processes to effectively and responsibly manage the division of labour between humans and machines while optimising performance and maintaining trust through effective validation and the quantification of uncertainty.

We use data generated in the impact cases of **WP4**, and from outside of the project (e.g. the Wellcome-funded Pathfinder Initiative). The insights we gain inform the design of tools built in **WP2**, and through engagement and capacity building co-ordinated by **WP5**, this work package informs changes in practice in the wider evidence synthesis community, mainstreaming innovative validation processes that are vital for the responsible use of machine learning in evidence synthesis.

Deliverables / Milestones

- **Research delineating the frontier of automation through the use of LLMs.** Delivers robust insights on the capabilities, limitations, and optimal design choices of LLM-assisted ES across the whole pipeline from the identification of studies to data extraction and synthesis.
- **Research on robust validation processes to ensure the responsible division of labour between humans and machines in human-in-the-loop processes.** Methods papers on validation procedures. Research on methods to optimise the division of labour between humans and machines.
- **Evaluation of the user-tool interface.** Research on user behaviour where DESTs are involved in evidence synthesis processes.
- **Recommendations for the responsible use of DESTs.** Informed by the research in deliverables 1–3, we publish recommendations addressing how DESTs can be used in evidence synthesis projects without eroding methodological standards.

WP4: Impact cases

Institutions: LSHTM (*ASCEND, PIK, eBASE, ACRES*)

Aims/activities

In **WP4** we will safely and responsibly apply DESTs in six Alive communities of practice (**WP1**) for living evidence to showcase that DESTs are fit-for-purpose in climate and health, work for diverse users, and deliver real-world impacts (see Figure 4). We can draw on extensive experience of systematic maps and reviews in climate and health [20, 41–46]. LSHTM, eBASE, ACRES and ASCEND provide strong connections and long-standing relationships to regional and local decision-makers, combined with experience in brokering actionable evidence for impactful decision making.

Figure 4: Overview of impact cases with evidence gaps and constituents

WP4 IMPACT CASES Showcasing the transformational power of Digital Evidence Synthesis Tools in six communities of practice for the delivery of rigorous and living evidence that matters to evidence users



Our cases are carefully selected to: fill critical evidence synthesis gaps identified by decision makers; cover a broad spectrum of evidence synthesis methods required to address heterogeneous evidence; involve a range of evidence users with different pathways to impact, governance scales, and intervention points at the science–policy interface; address different geographies from high to low availability of resources and evidence.

The first two impact cases (1 & 2) will co-create needs-driven living evidence mapping methodologies to improve the effectiveness of the overall evidence ecosystem for knowledge users operating at local to global scales.

Case 1 will provide the first living, multi-purpose map of the entire evidence-based climate and health literature by combining AI-facilitated content synthesis, data enhancement strategies, and bibliometrics.

Case 2 will create a living evidence gap map and thematic synthesis of climate-related health impacts and co-benefits of actions in cities by advancing automation of traditional mapping methodologies and transferring evidence to evidence-poor areas.

Next, impact cases 3–6 will focus on co-producing rigorous living systematic reviews that enable fast evidence-based decision making for reaping health co-benefits of climate actions.

Case 3 will demonstrate a living quantitative synthesis addressing the lack of comparative evidence on impacts on human mortality and morbidity of various climate change mitigation and adaptation responses, using proven methods of data harvesting [47].

Case 4 will complement this by demonstrating how to **advance automation and scaling in mixed methods synthesis**, where human behaviour and uptake of solutions play a key role in success. We will use climate and health actions for sustainable food systems as our core example.

In case 5 (extension of case 2) we will work with local decision-makers on **effective evidence transfers**, focusing on transferring knowledge on heat-related mortality and morbidity in cities to evidence-poor settings by combining empirical and modelling evidence.

In case 6 we will work with the Global SDG Synthesis Coalition to **accelerate progress towards climate and health-related SDGs**. We will explore radical automation strategies to strengthen large-scale UN-style science assessments, and learn how to work with less standardised, grey literature sources in synthesis.

Deliverables / Milestones

- **Research articles:** a set of peer-reviewed living systematic maps and reviews as well as papers advancing evidence synthesis methodology in the space of climate and health
- **Customised Interactive, living evidence platforms:** co-produced and designed with the members of our communities of practice
- **Evidence engagement and impact creation workshops:** a continuous flow of workshops in coordination with **WP1** that guides evidence production as well as tool development and use novel story-telling methods for creating impact
- **A diverse, global community of practice** for climate and health DESTs across research, practice, and policy.

WP5: Engagement and capacity building

Institutions: Campbell Collaboration , Cochrane , ASCEND, eBASE, ACRES (*all partners*)

Aims/activities

Our goal is to mainstream DEST best practice among users, producers and funders of high quality evidence synthesis for sustainable impact.

We recognize the urgent need to accelerate technology adoption in the field of evidence synthesis. This work package will use the unique networks, skills, and facilitation power in the consortium to bridge between DEST developers and current and future DEST users.

Deliverables / Milestones

- **Joint roadmap for AI in evidence synthesis:** published by Campbell and Cochrane
- **FAIR open data standards** - defined and implemented by partners and shared with the wider community
- **Best practices, responsible AI, and methods guidelines:** to ensure quality, endorsed by Campbell and Cochrane
- **Support for institutions shifting to DEST best practice:** validated by a global group of early adopters
- **Tool sharing and community engagement:** enhance DEST development, evaluation and uptake by engaging with other tool developers through annual hackathons, tool sharing, open source projects, and other collaboration opportunities
- **Knowledge and skills sharing:** reap network effects in the provision of training (train-the-trainer) and capitalise on extensive training networks, Campbell and Cochrane's leading sector conferences, and ASCEND's support programme for African synthesis teams

WP6: Project management, monitoring and evaluation, and sustainability

Institutions: PIK (*all partners*)

Aims/activities

WP6 will ensure that the consortium stays focused on our planned goals, makes adjustments where necessary, and shares learning to continue to have an impact beyond the funding period through rigorous project coordination and steering, global and regional advisory bodies, regular monitoring and independent, external project evaluation.

Deliverables / Milestones

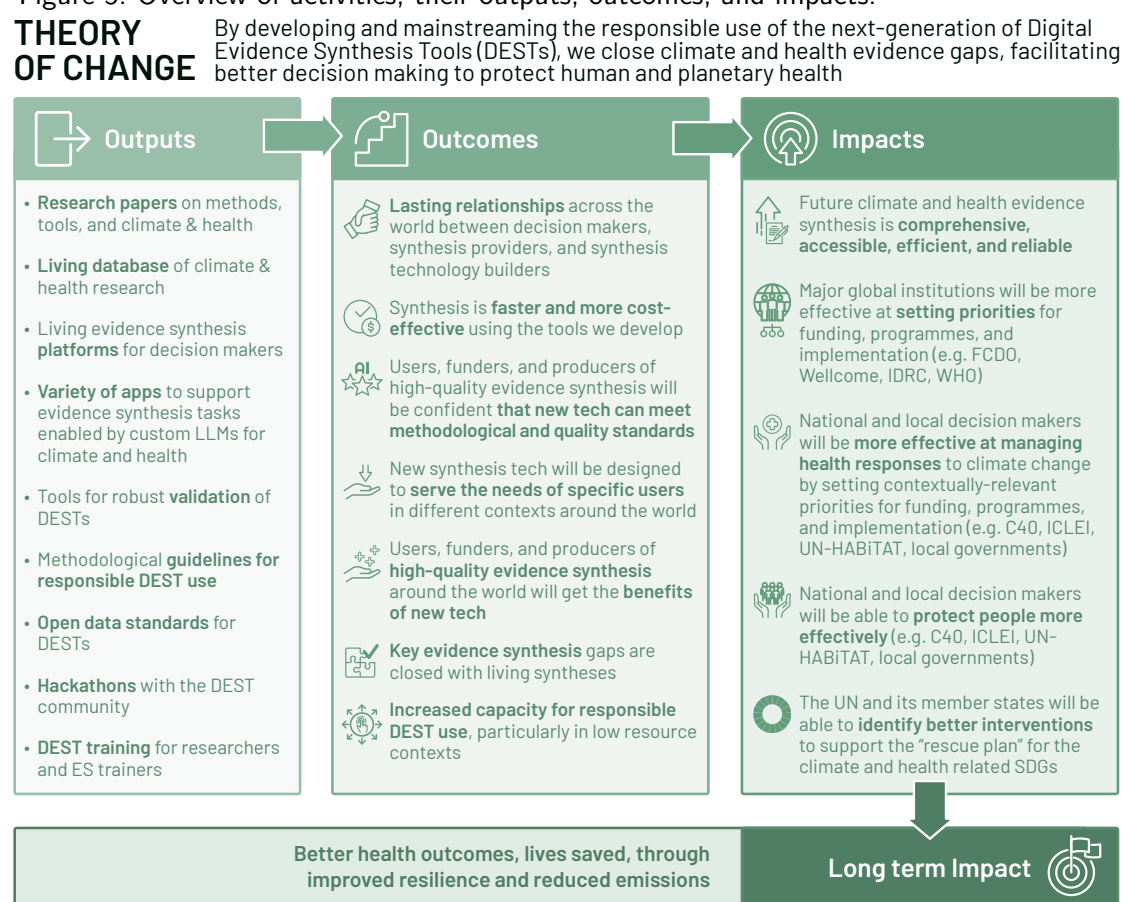
- **Communications:** Project website and DEST knowledge hub; coordination across press offices
- **Project coordination & steering:** Two full consortium meetings per year (3 days; one in-person; one virtual); bi-annual meetings of global steering committee and regional public advisory boards, monthly project executive team meetings
- **Learning:** DESTinY seminar series
- **Monitoring and evaluation:** Annual progress updates; final independent project evaluation

4 Theory of Change

Our Theory of Change diagram (Figure 5) outlines the pathways which lead to the expected impacts from the planned outputs of the work packages. This framework offers us an opportunity to evaluate—and where necessary adapt—our work throughout the project and make sure we fill important gaps in the existing evidence support system.

We recognise that the process of realising project impact will be more complex than can be captured in the diagram below. Our pathways to impact depend on co-producing work that is relevant, accessible, and useful to the different groups of decision-makers we will serve. This detailed and nuanced understanding of decision-maker needs and contexts will be informed by our experienced team using the Alive model to put co-production at the heart of everything we do.

Figure 5: Overview of activities, their outputs, outcomes, and impacts.



5 Equity, diversity, and inclusion

We've designed this project with a shared commitment to shift power that has historically been concentrated inequitably.

This shapes our

- shared responsibility and equal opportunity to host and attend project meetings in Africa and Europe,
- planned project budget allocating funds for external participants from LMIC to attend our hackathons, conferences, and trainings,
- strategy, of focusing on impact cases that define what work needs to be done and how it should be done to understand and meet the expressed needs of affected resource-scarce communities and decision-makers
- co-production, using the proven user-led Alive model
- cooperation, every organisation in our team has commitment to and experience working in a truly global way, and we are a team of organisations with structural advantages of working in high income countries, and organisations created and led from LMICs
- wider network, so the consortium includes both global and regional networks and we allocate resources to ensure the participation of stakeholders from LMICs in our work and events
- technology, where our team has experience developing and rolling out AI collaboratively in many countries and languages for tasks with extremely sensitive EDI concerns
- monitoring and evaluation, because we recognise that no approach is perfect and that we have all made mistakes, we commit to regular reflective reviews to ensure we are addressing and not exacerbating power imbalances through the design and delivery of our project.

Evidence synthesis done well can help to provide a more equitable information environment for everyone and create more equitable outcomes in the face of climate and health risks. We are committed to making sure our work lives up to that opportunity as an ongoing process—evaluating and learning as we go.

References

1. Cissé, G., McLeman, R., Adams, H., Aldunce, P., Bowen, K., Campbell-Lendrum, D. & Clayton, S. 2022: health, wellbeing, and the changing structure of communities, 1041–1170 (2022).
2. Romanello, M., Napoli, C. D., Green, C., Kennard, H., Lampard, P., Scamman, D. & **Jan Minx**. The 2023 report of the Lancet Countdown on health and climate change: the imperative for a health-centred response in a world facing irreversible harms. *The Lancet* **402**, 2346–2394 (2023).
3. Hess, J. J., Ranadive, N., Boyer, C., Aleksandrowicz, L., Anenberg, S. C., Aunan, K., Belesova, K., Bell, M. L., Bickersteth, S., Bowen, K., *et al.* Guidelines for modeling and reporting health effects of climate change mitigation actions. *Environmental Health Perspectives* **128**, 115001 (2020).
4. Whitmee, S., Green, R., Belesova, K., Hassan, S., Cuevas, S., Murage, P. & Picetti, R. Pathways to a healthy net-zero future: report of the Lancet Pathfinder Commission. *The Lancet* **403**, 67–110 (2024).
5. Creutzig, F., Niamir, L., Bai, X., **Max Callaghan**, Cullen, J., Díaz-José, J. & *et al.*, M. F. Demand-side solutions to climate change mitigation consistent with high levels of well-being. *Nature Climate Change* **12**, 36–46 (2022).
6. **Julian H. Elliott**, Turner, T., Clavisi, O., **James Thomas**, Higgins, J. P., Mavergames, C. & Gruen, R. L. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS medicine* **11**, e1001603 (2014).
7. **Julian H. Elliott**, Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S. & Salanti, G. Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of clinical epidemiology* **91**, 23–30 (2017).
8. **Julian Elliott**, Lawrence, R., Minx, J. C., Oladapo, O. T., Ravaud, P., Jeppesen, B. T., **James Thomas**, Turner, T., Vandvik, P. O. & Grimshaw, J. M. Decision makers need constantly updated evidence synthesis. *Nature* **600**, 383–385 (2021).
9. **Jan C. Minx**, **Max Callaghan**, Lamb, W. F., Garard, J. & Edenhofer, O. Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy* **77**, 252–259 (2017).
10. O'Mara-Eves, A., **James Thomas**, McNaught, J., Miwa, M. & Ananiadou, S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews* **4**, 1–22 (2015).
11. **Hugh Waddington**, Masset, E. & Jimenez, E. What have we learned after ten years of systematic reviews in international development? *Journal of Development Effectiveness* **18**, 1–16. <https://doi.org/10.1080/19439342.2018.1441166> (2018).
12. Marshall, I. J., Noel-Storr, A., Kuiper, J., **James Thomas** & Wallace, B. C. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. *Research synthesis methods* **9**, 602–614 (2018).
13. **Max W. Callaghan** & Müller-Hansen, F. Statistical stopping criteria for automated screening in systematic reviews. *Systematic Reviews* **9**, 1–14 (2020).

14. Qureshi, R., Shaughnessy, D., Gill, K. A., Robinson, K. A., Li, T. & Agai, E. Are ChatGPT and large language models “the answer” to bringing us closer to systematic review automation? *Systematic Reviews* **12**, 72 (2023).
15. Schmidt, L., Hair, K., Graziozi, S., Campbell, F., Kapp, C., Khanteymoori, A., Craig, D., Engelbert, M. & **James Thomas**. Exploring the use of a Large Language Model for data extraction in systematic reviews: a rapid feasibility study. *arXiv preprint arXiv:2405.14445* (2024).
16. Kapoor, S. & Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **4** (2023).
17. Devane, D., Burke, N. N., Treweek, S., Clarke, M., **James Thomas**, Booth, A., Tricco, A. C. & Saif-Ur-Rahman, K. M. Study within a review (SWAR). *Journal of Evidence-Based Medicine* **15**, 328 (2022).
18. **Repke, Tim & Callaghan, Max**. NACSOS-nexus: NLP Assisted Classification, Synthesis and Online Screening with New and EXTended Usage Scenarios. *arXiv preprint arXiv:2405.04621* (2024).
19. Schmidt, L., Mutlu, A. N. F., Elmore, R., Olorisade, B. K., **Thomas, James** & Higgins, J. P. Data extraction methods for systematic review (semi) automation: Update of a living systematic review. *F1000Research* **10** (2021).
20. Berrang-Ford, L., Sietsma, A. J., **Callaghan, Max, Minx, Jan C, Scheelbeek, Pauline FD**, Haddaway, N. R., **Haines, Andy** & Dangour, A. D. Systematic mapping of global research on climate and health: a machine learning review. *The Lancet Planetary Health* **5**, e514–e525 (2021).
21. **Callaghan, Max**, Schleussner, C.-F., Nath, S., Lejeune, Q., Knutson, T. R., Reichstein, M., Hansen, G., Theokritoff, E., Andrijevic, M., Brecha, R. J., *et al.* Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature climate change* **11**, 966–972 (2021).
22. **Thomas, James**, McDonald, S., Noel-Storr, A., Shemilt, I., **Elliott, Julian**, Mavergames, C. & Marshall, I. J. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. *Journal of Clinical Epidemiology* **133**, 140–151 (2021).
23. Arno, A., **Elliott, Julian**, Wallace, B., Turner, T. & **Thomas, James**. The views of health guideline developers on the use of automation in health evidence synthesis. *Systematic Reviews* **10**, 1–10 (2021).
24. Schmidt, L., Hair, K., Graziozi, S., Campbell, F., Kapp, C., Khanteymoori, A., Craig, D., Engelbert, M. & **Thomas, James**. Exploring the use of a Large Language Model for data extraction in systematic reviews: a rapid feasibility study. *arXiv preprint arXiv:2405.14445* (2024).
25. Shemilt, I., Noel-Storr, A., **Thomas, James**, Featherstone, R. & Mavergames, C. Machine learning reduced workload for the Cochrane COVID-19 Study Register: development and evaluation of the Cochrane COVID-19 Study Classifier. *Systematic Reviews* **11**, 15 (2022).
26. Singh, G., **Thomas, James**, Marshall, I. J., Shawe-Taylor, J. & Wallace, B. C. *Structured multi-label biomedical text tagging via attentive neural tree decoding in 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018* (2020), 2837–2842.

27. Beller, E., Clark, J., Tsafnat, G., Adams, C., Diehl, H., Lund, H., Ouzzani, M., Thayer, K., **Thomas, James**, Turner, T., *et al.* Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). *Systematic reviews* **7**, 1–7 (2018).
28. Cramond, F., O'Mara-Eves, A., Doran-Constant, L., Rice, A. S., Macleod, M. & Thomas, J. The development and evaluation of an online application to assist in the extraction of data from graphs for use in systematic reviews. *Wellcome Open Research* **3** (2018).
29. Howard, J. & Ruder, S. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146* (2018).
30. Xiong, C., Power, R. & Callan, J. *Explicit semantic ranking for academic search via knowledge graph embedding* in *Proceedings of the 26th international conference on world wide web* (2017), 1271–1279.
31. Shemilt, I., Arno, A., **James Thomas**, Lorenc, T., Khouja, C., Raine, G., Sutcliffe, K., Kwan, I., Wright, K. & Sowden, A. Cost-effectiveness of Microsoft Academic Graph with machine learning for automated study identification in a living map of coronavirus disease 2019 (COVID-19) research. *Wellcome Open Research* **6** (2021).
32. Nussbaumer-Streit, B., Ellen, M., Klerings, I., Sfetcu, R., Riva, N., Mahmić-Kaknjó, M., Poulentzas, G., Martinez, P., Baladia, E., Ziganshina, L. E., *et al.* Resource use during systematic review production varies widely: a scoping review. *Journal of clinical epidemiology* **139**, 287–296 (2021).
33. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., *et al.* Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* **33**, 9459–9474 (2020).
34. **James Thomas**, Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C. & Glasziou, P. Living systematic reviews: 2. Combining human and machine effort. *Journal of clinical epidemiology* **91**, 31–37 (2017).
35. Arno, A., **Thomas, James**, Wallace, B., Marshall, I. J., McKenzie, J. E. & **Elliott, Julian H.** Accuracy and efficiency of machine learning–assisted risk-of-bias assessments in “real-world” systematic reviews: a noninferiority randomized controlled trial. *Annals of internal medicine* **175**, 1001–1009 (2022).
36. Marshall, I. J., Noel-Storr, A., Kuiper, J., **Thomas, James** & Wallace, B. C. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. *Research synthesis methods* **9**, 602–614 (2018).
37. **Callaghan, Max W** & Müller-Hansen, F. Statistical stopping criteria for automated screening in systematic reviews. *Systematic Reviews* **9**, 1–14 (2020).
38. Noel-Storr, A., Dooley, G., **Elliott, Julian**, Steele, E., Shemilt, I., Mavergames, C., Wisniewski, S., McDonald, S., Murano, M., Glanville, J., *et al.* An evaluation of Cochrane Crowd found that crowdsourcing produced accurate results in identifying randomized trials. *Journal of Clinical Epidemiology* **133**, 130–139 (2021).
39. Wang, S., Scells, H., Zhuang, S., Potthast, M., Koopman, B. & Zucco, G. Zero-shot Generative Large Language Models for Systematic Review Screening Automation. *arXiv preprint arXiv:2401.06320* (2024).

40. Spijker, R., Dinnes, J., Glanville, J. & Eisinga, A. Searching for and selecting studies. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*, 97–129 (2023).
41. **Scheelbeek, Pauline FD**, Dangour, A. D., Jarmul, S., Turner, G., Sietsma, A. J., **Minx, Jan C**, **Callaghan, Max**, Ajibade, I., Austin, S. E., Biesbroek, R., *et al.* The effects on public health of climate change adaptation responses: a systematic review of evidence from low-and middle-income countries. *Environmental Research Letters* **16**, 073001 (2021).
42. Green, R., Cornelsen, L., Dangour, A. D., Turner, R., Shankar, B., Mazzocchi, M. & Smith, R. D. The effect of rising food prices on food consumption: systematic review with meta-regression. *Bmj* **346** (2013).
43. Aleksandrowicz, L., Green, R., Joy, E. J., Smith, P. & Haines, A. The impacts of dietary change on greenhouse gas emissions, land use, water use, and health: a systematic review. *PloS one* **11**, e0165797 (2016).
44. Nájera Espinosa, S., Hadida, G., Jelmar Sietsma, A., Alae-Carew, C., Turner, G., Green, R., Pastorino, S., Picetti, R. & **Scheelbeek, Pauline**. Mapping the evidence of novel plant-based foods: a systematic review of nutritional, health, and environmental impacts in high-income countries. *Nutrition Reviews*, nuae031 (2024).
45. Carr, T. W., Mkuhlani, S., Segnon, A. C., Ali, Z., Zougmore, R., Dangour, A. D., Green, R. & **Scheelbeek, Pauline**. Climate change impacts and adaptation strategies for crops in West Africa: a systematic review. *Environmental research letters* **17**, 053001 (2022).
46. Whitmee, S., Green, R., Belesova, K., Hassan, S., Cuevas, S., Murage, P., Picetti, R., Clercq-Roques, R., Murray, K., Falconer, J., *et al.* Pathways to a healthy net-zero future: report of the Lancet Pathfinder Commission. *The Lancet* **403**, 67–110 (2024).
47. Sharma Waddington, H., Masset, E., Bick, S. & Cairncross, S. Impact on childhood mortality of interventions to improve drinking water, sanitation, and hygiene (WASH) to households: Systematic review and meta-analysis. *PLoS medicine* **20**, e1004215 (2023).

Table 1: List of collaboration partners and their extended networks.

| Short name | Full name | Partner networks |
|-------------------|---|---|
| ASCEND | African Synthesis Centre for Climate Change, Environment and Development at the University of Cape Town | African Group of Negotiators Expert Support (AGNES), African Adaptation Initiative, Climate and Development Knowledge Network (CDKN); Hosts up to 450 visiting researchers and policymakers from across Africa per year to conduct evidence synthesis; |
| C40 CDP EH!WOZA | C40 Cities Climate Leadership Group CDP—Disclosure, Insight and Action Eh!woza | Global network of nearly 100 mayors of the world's leading cities In 2023, over 24,000 organisations around the world reported data through CDP. Engages people living in townships and informal settlements in high impact research around health challenges |
| GSSC | Global SDG Synthesis Coalition | 45 UN partners along with key synthesis and evidence networks such as the Global Commission on Evidence to Address Societal Challenges. UN member states are involved at various levels, including a high-level political steering committee with representatives from Spain, Ireland, Malawi, Panama, the Philippines, Denmark, Norway, Lithuania, and the United Kingdom. |
| ICLEI | ICLEI - Local Governments for Sustainability e.V. | Local Governments for Sustainability is a global network of more than 2,500 local and regional governments committed to sustainable urban development |
| iCoDe Abakwa IDRC | Intentional Coding for Development International Development and Research Centre | City of Bamenda and similar cities across Africa BAOBAB project, Campbell, FCDO for the Climate Adaptation and Resilience (CLARE) program, Ministry of Foreign Affairs of the Netherlands, Step Change initiative |
| Lancet NWRA | Lancet Countdown Regional Assembly of the North West Region of Cameroon | 300 researchers from around the world 34 city councils in the North West region of Cameroon |
| PAICE | Policy and Implementation for Climate & Health Equity project at the University College London | Carbon disclosure project (CDP), Greater London Authority (GLA), Organisation for Economic Co-operation and Development (OECD), UK Climate change committee (UK-CCC) |
| Pathfinder | Wellcome-funded Pathfinder Initiative at the University of London | C40 cities, Carbon disclosure project (CDP), Sustainable Development Solutions Network (SDSN, which in turn has links with the EU 100 Cities), and OECD |
| NWB | Regional delegation for the north-west Bali health district service | |
| SECTION27 | SECTION27—catalyst for social justice | Engaged with South African government for health system adaptation to climate change impacts |
| UK-EA WHO | Environment Agency of the United Kingdom WHO Collaborating Centre on Climate Change, Health and Sustainable Development | Policy-makers across England in the field of climate and health and in general WHO Headquarters in Geneva, WHO Regional Office for Europe, European Centre for Environment and Health in Bonn |
| WRC | Water Research Commission of South Africa | |