



Summary Report 2018

Evidence Synthesis Hackathon











Evidence Synthesis Hackathon Summary Report May 2018

Neal R Haddaway^{1,2} and Martin J Westgate³

¹ Mistra EviEM, Stockholm Environment Institute, Stockholm, Sweden

² Africa Centre for Evidence, University of Johannesburg, Johannesburg, South Africa
³ Fenner School of Environment & Society, Australian National University, Canberra,

Australia

This report summarises a workshop held in Stockholm at Stockholm Environment Institute (SEI), Global Water Partnership (GWP) and Stockholm International Water Institute (SIWI) from the 23rd to the 25th April 2018. The event brought together 25 programmers and coders from across the globe in an attempt to collectively solve some of the biggest issues facing evidence synthesis using technology. The event was generously funded by Mistra EviEM (www.eviem.se/en) and the Fenner School's Environment & Society Synthesis Program (Australian National University).

Further details of the Evidence Synthesis Hackathon can be found on the website: <u>www.evidencesynthesishackathon.com</u>.

This report is published by the Stockholm Environment Institute (SEI).

This report is distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits any non-commercial use, distribution, and reproduction in any medium, provided the original authors and source are credited.

All images in this publication remain the sole property of the photographer and may not be reproduced.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



Please cite this report as:

Neal R Haddaway and Martin J Westgate (2018) Evidence Synthesis Hackathon 2018. Stockholm Environment Institute, Sweden 31 pp. DOI: 10.6084/m9.figshare.6322937

Contents

1.	What is an Evidence Synthesis Hackathon and why is it important?	Page 2
2.	How was the Evidence Synthesis Hackathon 2018 (ESHackathon2018)	
	funded and organised?	Page 5
3.	Who attended ESHackathon2018?	Page 8
4.	What happened during ESHackathon2018?	Page 17
5.	How was ESHackathon2018 portrayed on social media?	Page 21
6.	What were the main outputs from ESHackathon2018?	Page 23
7.	What are the plans for future Evidence Synthesis Hackathons?	Page 29



1. What is an Evidence Synthesis Hackathon and why is it important?

The Importance of Evidence Synthesis

Systematic reviews and maps are carefully planned and conducted forms of literature review that aim to maximise transparency and comprehensiveness, and minimise risk of bias. They do this through the use of 1) peerreviewed a priori planning documents; 2) a step-wise approach to collating evidence from multiple sources (including grey literature), 3) screening records at title, abstract and full text according to predetermined inclusion criteria; 4) critically appraising all evidence on its validity (only compulsory for systematic reviews); and 5) transparently documenting all decisions made and evidence found at all steps. Systematic reviews and maps should be conducted according to the rigorous standards outlined in the guidance produced by several key organisations across many disciplines, including those of the Collaboration for Environmental Evidence (CEE; www.environmentalevidence.org).

Question Formulation

Protocol (p Searching Article Scree Data Extrac Critical App Synthesis Final Revie Protocol (peer-reviewed and published) Article Screening Data Extraction Critical Appraisal Final Review (peer-reviewed and published) Communication

Because of their necessarily robust methods, systematic reviews and maps are particularly resource intensive. Furthermore, the requirements for transparency and repeatability mean that detailed documentation is needed. The large number

of detailed electronic supplementary files associated with a systematic review or map can facilitate the propagation of errors or unintentional missing information that precludes verification or replication. Many tasks within a systematic review are repetitive and, once carefully planned, may not be particularly intellectually demanding.

Evidence synthesis + technology.

A Role for Technology

In recent years, a number of technological solutions (software and 'apps') have been created to improve resource efficiency, transparency and validity of systematic reviews and maps. However, many of these tools are not open source and cannot be further developed for specific needs. Furthermore, many gaps remain where technology could further assist systematic reviewers.

Hackathons: The Solution to our Problems?

A 'hackathon' is defined as creative problem solving using code and programming (https://hackathon.guide/). Hackathons can focus on practical or theoretical solutions to problems, but most revolve around technological solutions.



Evidence Synthesis Hackathon

The Evidence Synthesis Hackathon

This Evidence Synthesis Hackathon (ESHackathon) aimed to bring together experienced coders and programmers from across the world to tackle some of the key challenges hampering efficiency, transparency and validity in systematic review or map conduct and communication.

A guiding principle for all of our activities during and after the ESHackathon was that all of the tools and techniques that we produce or refine should be available and usable by everyone. Therefore, all code was produced Open Source (typically using GitHub repositories; see Section 6), meaning that anyone can see exactly how the programmes were developed and can download and modify them if they want to. This transparency also means that the tools will all be usable by anyone, irrespective of their resources or funding. Making tools that are accessible to all was one of the main reasons for convening the ESHackathon in the first place.

The ESHackathon accepted applications from coders, programmers and evidence synthesis specialists from all over the world, selecting the best applicants to attend a 3-day workshop in Stockholm.

By bringing together experienced coders and programmers, this hackathon aimed to provide technological solutions at various stages of the systematic review/map process.

The overall aim was to help reviewers to improve efficiency, transparency, and validity of their reviews.



2. How was the Evidence Synthesis Hackathon 2018 (ESHackathon2018) funded and organised?

Funding

The ESHackathon 2018 was generously funded by Mistra EviEM (www.eviem.se/en), a project that ran from 2012 to 2018 conducting systematic reviews and maps relating to environmental issues relevant to Sweden. Funding was also provided through the Environment and Society Synthesis Program, an initiative of the Fenner School of Environment & Society at the Australian National University. The ESHackathon was held across three venues in Stockholm: Stockholm Environment Institute, the Global Water Partnership, and the Stockholm International Water Institute. These organisations generously provided working space free-of-charge.

Funding was used primarily for transport and subsistence for the 26 attendees travelling from overseas: 9 people came from the UK, 7 from the USA, 3 from Australia, 2 from Nepal, 2 from Chile, and 1 from each of Singapore, South Africa and the Netherlands. Participants' carbon footprints from international travel was offset using the Vi Skogen scheme (www.viskogen.se).

Participants were provided with accommodation across three hotels in the city centre from Sunday 22nd to Thursday 26th April. Food and drink was provided throughout the event, starting with a social event at SEI on the Sunday for those who had already arrived.

All accommodation was organised by Neal Haddaway booking directly. This enabled costs to be kept to a minimum: approximately 1,000 SEK per person per night for hotel accommodation. Flights were booked via SEI's partner travel agent, Tranås, and in Canberra via Travelmakers. Full details of the budget are provided in Appendix A. Other costs included catering, working out at approximately 1,000 SEK per person for the entire event.

Selection of applicants

Applicants were invited to submit an expression of interest (EoI) to participate in the ESHackathon via a web form at www.evidencesynthesishackathon.com. This Eol form was circulated over social media and via email networks for approximately 5 weeks before the deadline. We received 59 Eols, and these were then assessed by Neal Haddaway and Martin Westgate based on their level of experience, programming skills, knowledge of evidence synthesis and their geographical location. Participants were selected based on merit and whilst aiming for diversity in location, background, skills, and interests. A final set of 22 participants were invited to take part in the ESHackathon. In addition, Marc Lajeunesse and Wolfgang Viechtbauer, experts in meta-analysis and systematic reviews, having written packages for relevant tasks within the R statistical software environment, were invited to attend as keynote speakers.



Three experts in evidence synthesis, Magnus Land, Biljana Macura and Jacqui Eales, were invited to assist in the coordination of the hackathon, acting as mentors and providing subject expertise/reality checking with each of the groups of participants.



3. Who attended ESHackathon2018?

Invited speakers





Marc is an Associate Professor in ecology and evolutionary biology at the University of South Florida. He is an elected member of the Society for Research Synthesis Methodology (SRSM) and has 15 years of experience in research synthesis methods. Marc has developed metaanalysis and systematic review tools in Python (<u>OpenMEE</u>) and R (<u>metagear</u>).



Wolfgang Viechtbauer | Maastricht University

Wolfgang is associate professor of methodology and statistics at Maastricht University in the Netherlands. His research is primarily focused on developing statistical methods and software for metaanalysis and the design and analysis of longitudinal and multilevel studies using appropriate mixed-effects models. He is author of the <u>metafor</u> package for R, which covers a wide variety of standard and advanced meta-analytic methods (e.g., fixed/random/mixed-effects

models, meta-regression, multilevel and multivariate meta-analysis, network meta-analysis, phylogenetic and spatio-temporal models).

Participants

The participants in the inaugural Evidence Synthesis Hackathon are listed below, in alphabetical order by last name.



Laurie Baker | University of Glasgow

Laurie is a final year PhD student in Biology at the University of Glasgow. For her PhD she is using Bayesian state-space models to understand spatial and temporal patterns in disease transmission. During her masters and bachelors in marine biology she studied grey seal movement off the coast of Nova Scotia and the effect of management decisions on the Chilean pink cusk-eel fishery. Her interests are in spatial modeling and she's recently discovered the joys

of text analysis and topic modeling.



Sergio Leonardo Benítez Díaz | Prodigious

Sergio is a system engineer with the emphasis on the design of user interfaces for improved user experiences. His work focussed on improved human-machine interactions and novel tools for data visualisation. His interests include football, tennis and travel.



Panagiotis Bozelos I University of Oxford

Panagiotis is a PhD candidate in Theoretical and Computational Neuroscience, with a major in Molecular Biology and Genetics. He is also employed as a Data Analyst at the Centre for Neural Circuits and Behaviour, University of Oxford. His research interests revolve around hippocampal processing functions of spatial and non-spatial information. He is also interested in the fields of Machine Learning / Artificial Intelligence, and History / Philosophy of Science.



Katie Corker | Grand Valley State University

Katie is a quantitative methodologist and assistant professor of psychology at Grand Valley State University. Her personality and social psychology research seeks to understand motivational differences between individuals. Katie also has active research interests that intersect with the open science and reproducibility in science movements. Together with others in the community (see improvingpsych.org), she builds and supports structures that

improve methods and practices in psychology. She is also interested in meta-analysis and cumulative knowledge building.



Brian Cottrell | AT&T

Brian is a mobile software engineer and hackathon enthusiast from Redondo Beach, California. He received his degree in physics from the University of California, Santa Barbara and currently works on developing mobile and TV applications at AT&T. In his free time, he enjoys building prototype software as a part of hackathons or other challenge events as well as competitive sailboat racing.



Sanita Dhaubanjar | International Water Management Institute

Sanita is a research officer in the Water Futures group at the International Water Management Institute. A water resources engineer by degree, Sanita currently supports watershed modeling research with hydrological and climate data processing, analysis and visualization. Her primary interest is the application of models and data analysis to provide information needed to achieve equitable and transparent water governance and management. In that pursuit, she has worked in

environmental modeling, hydro-meteorological monitoring, hydroeconomics and forecasting. Sanita is passionate about finding novel ways to synthesize and visualize scientific evidence to support better decision making in water management.



Spencer Dixon | UN Environment World Conservation Monitoring Centre

Spencer is a full stack web developer at UN Environment World Conservation Monitoring Centre in Cambridge, UK. He currently works on creating various web based tools

like <u>ProtectedPlanet.net</u> and <u>SpeciesPlus.net</u> that allow organisations and governments to make better decisions around conservation and biodiversity. In his spare time, he is experimenting with machine

learning in conservation, speaking about blockchain, and making music.



Andrew Feierman | Data-Driven Yale

Andrew is a Quantitative Analyst for the Yale Data-Driven Environmental Solutions group. His research and experience is centered around energy efficiency in real estate. Prior to joining Yale, he worked for the Institute for Market Transformation in Washington, DC, where he modeled city-level environmental policies and worked with real estate companies to reduce energy consumption within their buildings. He has a B.A. from American University's School of

International Service, and further education from the New York City Data Science Academy.



Matt Grainger | Newcastle University

Matt is a Research Associate in the Modelling, Evidence & Policy Research Group at Newcastle University. He works with graphical Bayesian decision models to contextualise scientific evidence for the benefit of policy makers in the realm of biodiversity conservation and sustainability. He is an expert in combining quantitative and qualitative data and dealing with high levels of uncertainty. He also has a side-line in the conservation of pheasants, predominately in Southeast Asia.



Charles Gray | La Trobe University

Charles is a proud <u>mathbassador</u> for the Australian Mathematical Sciences Institute's <u>Choose Maths</u> program. Her role as a math-talkingdoing-advocate grrrl is a relatively new career development, after spending almost twenty years working as a classically-trained pianist and music teacher. She lectures at La Trobe University, where she is undertaking a PhD in statistical data science and takes an active role in the Victorian branch of the Statistical Society of Australia. She likes to

think of herself as a data detective who tells stories with data.



Sarah Han | Collide LLC

Sarah is a multi-disciplinary software engineer and UX/UI designer with skills in front end development, 3D web visualizations and 3D modeling. She is passionate about integrating design and technology to develop beautiful, functional and interactive products that enhance people's lives. Sarah is particularly interested in emerging and experimental

technologies. She regularly participates in hackathons to create innovative inventions.





Peter Ma | Clean Water Al

Peter has been a software developer for more than 14 years, and has been involved in many start-ups and projects that pushed the envelope of innovation. He is part of the Intel Software Innovator program, and focusses on emerging technologies like AI and IoT. He is currently working on Clean Water AI, an AI system that detects dangerous bacteria in the water sources.

Andrew Martin | University of Oxford

Andrew is in the final stages of his DPhil in long-term ecology at Oxford University. He uses dendroecological, stable isotope, and mechanistic modelling techniques to address questions surrounding Arctic 'shrubification'. Andrew became interested in evidence-based techniques through creating an evidence-map of the controls on Arctic shrub growth and expansion. With a background as a software developer (in Government, and as an intern at Microsoft Research),

coding is central to all aspects of his research, particularly the application of functional programming and web technologies. Aside from his Arctic research, Andrew is also lead researcher on the Global Pollen Project (globalpollenproject.org), which aims to establish a global pollen taxonomy.



Geoffrey Martin | Yale NUS College

Geoffrey is a third-year undergraduate student from Yale-NUS College majoring in computer science and statistics. He has spent time conducting data driven research on economics, dating, health policy and the environmental sciences, for several research centers at his university and in the data analytics industry. His interests span the intersection of data science and macrosociological issues, and will be spending the coming summer as a machine learning intern at NASA's

Jet Propulsion Laboratory.



Sonia Mitchell | Glasgow University

Sonia is a PhD student from the University of Glasgow and member of the <u>Boyd Orr Centre for Population and Ecosystem Health</u>. My main research interests focus on the measurement and interpretation of biological diversity. I have been involved in projects involving a wide range of applications ranging from an investigation of the spatial and temporal dynamics of forest biodiversity to a comparative phenotypic, genetic, and phylogenetic study of the transmission of antimicrobial

resistance across sympatric human and animal populations. I am the author of the rdiversity package for R (<u>https://cran.r-project.org/web/packages/rdiversity/index.html</u>), which provides a framework for the measurement and partitioning of the (similarity-sensitive) biodiversity of a metacommunity and its constituent subcommunities.



Christopher Penkin | Digital Solution Foundry

Chris has 16 years web and application development experience, working on a wide range of different technologies over the years. His passion is to not just creating great software, but great looking, easy to use software where the user is the central character of any solution.



Daniel Perez | Epistemonikos

Daniel studied Computer Science, and co-developed Expenews.com in 2007, a service to allow mountaineers and adventurers to share their experiences of Antarctica or Everest in realtime. In 2009 he traveled to Silicon Valley to co-found Zappedy (since acquired by Groupon), and then in 2011 co-created Epistemonikos with Gabriel Rada, a non for profit that seeks to provide the best evidence in health. Today Epistemonikos is the world's largest systematic reviews search engine.

He teaches about Blockchain and cryptocurrencies, and programs in Python, Ruby and Node. XP (of Agile development) and Customer development are strong influences on his approach to software development.



Gihan Samarasinghe | University of New South Wales

Gihan is a Research Associate in the School of Biological Earth and Environmental Sciences (BEES) of University of New South Wales (UNSW), Sydney, Australia, and mainly working in the project titled "Developing a Methodology for Systematic Review and Establishing Synthesis Methodology in Built Environment: Towards Evidence-Based Practice and Policy". Gihan did his PhD in Machine Learning and Computer Vision in the School of Computer Science and Engineering of

UNSW and he has research and industrial experience in Machine Learning, Database Systems, and Software Engineering.



Gorm Shackelford | University of Cambridge

Gorm works on sustainable agriculture as part of the Conservation Evidence project at the University of Cambridge (<u>www.conservationevidence.com</u>). He systematically reviews evidence for the effects of agricultural practices on soil, water, biodiversity, crops, and other variables. His academic training was not in computer programming, but he has increasingly been using Python (Django) to develop web apps, such as a crowdsourcing platform for systematic

reviews of global catastrophic risks (<u>www.x-risk.net</u>) and a platform for meta-analysis that is now in development. He has used R extensively for meta-analysis, GIS, and statistical modelling. Before he started working in academia, he was a professional portrait photographer (<u>www.gormshackelford.com</u>).



Mridul Shrestha | Alight Tech

Mridul works on providing technology solutions to companies of diverse domains working in the field of agriculture, medical appliances, internet of things among many others. He is a startup enthusiast and prior to founding Alight (Tech-consulting), he co-founded GoGazzab, an ecommerce startup, for the Nepalese market. Mridul completed his engineering (BSc. & MSc.) from Germany with an Honors degree in Technology management.



Ezgi Tanriver-Ayder I University of Edinburgh

Ezgi is a Biostatistician and a PhD student at the Centre for Clinical Brain Sciences working with the <u>CAMRADES group</u>. Her research focuses on improving statistical approaches for meta analysis in preclinical drug discovery research. Her key interest is on providing solutions to reproducibility issue and enhance ways to translate the obtained outcomes from animal research to human studies. Her project mainly involves reviewing existing statistical methods used in

systematic review and meta analysis of preclinical data as well as developing new techniques to overcome the limitations in the existing tools, including Bayesian approaches.



Juan Vásquez I Epistemonikos

Juan is a software engineer at Epistemonikos Foundation. He has contributed on the development of the Epistemonikos database (www.epistemonikos.org) and some other projects on the evidence synthesis field such as iSoF (Interactive Summary Of Findings), iEtD (Interactive Evidence To Decision) and most recently L·OVE (Living OVerview of the Evidence) which is a platform with evidence organized by PICO question designed to make possible the creation of living

overviews. His interests are data visualization, processing, and mining, and machine learning.

Systematic review experts



Jacqui Eales | Exeter University

Jacqui is an evidence synthesis specialist who has spent the last 9 years working in systematic reviews in the environmental management and human health sectors. Her background is in ecology and conservation, having completed a PhD in genetics of tropical invasive species in 2008. She is PI on a long-term research project characterising the terrestrial and marine biodiversity of a tropical forested Caribbean island, Dominica, in partnership with the NGO

Operation Wallacea and the Dominican Government. Jacqui has also focused on education and capacity building, previously holding a position as Lecturer in Conservation at Bangor University, UK.



Biljana Macura I Stockholm Environment Institute

Biljana is a Research Fellow at SEI, Sweden. She is an environmental social scientist with interdisciplinary background. She holds a PhD in forest policy from University of Padova, Italy and Bangor University, United Kingdom. Biljana is currently working with the MISTRA EviEM project (www.eviem.se) on evidence synthesis in the field of environmental management where she conducts systematic reviews, provides training and works on improvement of systematic review

methods. Biljana is also Editorial Manager of the Collaboration for Environmental Evidence journal – Environmental Evidence.



Magnus Land I Stockholm Environment Institute and Formas

Magnus is an analyst at Formas, a Swedish research funding agency where the conduct of evidence syntheses on environmental management issues currently is being implemented. Magnus is also a research fellow at SEI, Sweden, where he has worked with the MISTRA EviEM project conducting systematic reviews and systematic maps. He is also interested in developing systematic review methods. Magnus has a background in aquatic geochemistry and holds a PhD in applied geology from Luleå University of Technology, Sweden. He has also

been working at California Institute of Technology and Stockholm University, and as an environmental consultant at WSP.

Organisers



Neal Haddaway | Stockholm Environment Institute and Africa Centre for Evidence

Neal is a research fellow at the Stockholm Environment Institute. His main research interests are around the production and use of environmental evidence in decision-making, by improving the transparency, efficiency and reliability of evidence synthesis as a methodology. Neal is the co-creator of ROSES (RepOrting standards for Systematic Evidence Syntheses; www.roses-reporting.com), a set of

rigorous standards for reporting the conduct of systematic reviews and maps in environmental topics.



Martin Westgate | Australian National University

Martin is a research fellow at the Australian National University. His research focusses on how scientific information can be used to mitigate human impacts on the environment, via a combination of empirical ecology and evidence synthesis. Martin is the creator of revtools (<u>https://revtools.net</u>), an R package for interactive visualisation of bibliographic data during evidence synthesis projects.



4. What happened during ESHackathon2018?

Coordination of the event

The ESHackathon started with an icebreaker exercise for participants to better get to know one another: arranging themselves outside based on where they had travelled from, and along a line based on their understanding of evidence synthesis and their level of excitement about the ESHackathon! Next, Neal Haddaway gave an overview presentation introducing systematic reviews and systematic maps, focusing on why they were important and what the key procedural steps and cornerstones are. Wolfgang Viechtbauer and Marc Lajeunesse then introduced their R packages, metaphor and metagear.



Following these important introductions, the hackathon began in earnest. Martin Westgate led a discussion of parts of the review process where technology might be useful. Participants were then invited to create a mind-map on one wall of the main room,

based on steps or key procedures in the conduct of a review. They then broke out into groups to discuss the problem and think of possible solutions that they could contribute to over the course of the hackathon.



The hacking

During the hackathon, groups spent time understanding the problem they wanted to address, before getting to grips with what their solution might look like. Teams started coding on day one, refining their tools over the course of the event.

Due to the non-competitive nature of the ESHackathon, participants were able to move between projects according to where their skills were needed, depending on the languages they coded in or their knowledge of problems and workarounds.

Midway through the hackathon, some new projects started up where participants identified a gap. This flexibility maximised the number and the quality of outputs from the ESHackathon. A half-way check-in on the morning of the second day gave everyone the opportunity to present their projects and discuss some of the key challenges they had faced so far.

Presenting the results

At the end of the third day of the hackathon, all of the participants gathered in the SEI Lounge to give a brief 5 minute presentation of their progress. A total of 10 outputs were presented (see Section 6). Many of these were applications, but some were conceptual outputs in the form of papers.

Following the presentations the group walked across town for a social event in Södermalm.





Exciting hackathon in the pipeline for #systematicreview tools! Looking for potential sponsors, tech startups, etc. @importio maybe? Great chance to sponsor the next major tool in evidence synthesis!



4:09 PM - 24 Nov 2017



Lessearch, CEE, EviEM and 7 others



5. How was ESHackathon2018 portrayed on social media?

Advertising the ESHackathon

Social media was used throughout the planning and execution phases of the ESHackathon. Twitter was used as a means of circulating a link to the Expression of Interest (EoI) form. Tweets relating to the calls for EoIs were widely circulated (Liked and Retweeted) (see the link below for a timeline of feeds relating to the ESHackathon):

https://twitter.com/i/moments/9992180273580 15488

LiveTweeting the event

During the hackathon, the participants were encouraged to liveTweet the event using the #ESHackathon hasthtag, as did the organisers. These tweets reached a large audience across environment and other sectors, generating great interest in the outputs (see some of the replies in the link above).

The response on social media was overwhelmingly positive, with great interest in the tools that were produced.





6. What were the main outputs from ESHackathon2018?

The first main output from the hackathon was a living network of likeminded programmers and researchers with an interest in evidence synthesis. Participants left having made new friends and contacts that may have an influence far beyond their involvement in producing the apps described below. A second key output was capacity sharing. The organisers shared their experience and knowledge of evidence synthesis, and in turn, the participants contributed a rich and diverse set of experiences and expertise related to programming, statistical environments, platforms for data sharing and organisation, existing tools for data analysis and visualisation, and their own networks and contacts. These outputs cannot be understated, and the organisers believed that if these were the only outputs from the ESHackathon2018 then the event would have been successful. Beyond these important outcomes, however, the participants produced a startling number and array of impressive tools, outlined below. Details of these apps can be found on the ESHackathon GitHub page: https://github.com/ESHackathon.

Output 1: EviAtlas

Systematic Maps are, according to the <u>Environmental Evidence Journal</u>, "overviews of the quantity and quality of evidence in relation to a broad (open) question of policy or management relevance." In simple terms, this means that documents are categorized according to the type, location, and publication information available for each work within a particular topic. Systematic maps are often used for environmental research, where it is particularly important to track the location of study sites. The spatial nature of a systematic map, particularly for environmental research, means that academics often use some kind of geographic map to analyze and present their information. Understanding the academic community's familiarity with the R programming language, the team decided to build a webapp using R Shiny that could automate certain parts of creating a systematic map for environmental research.

Using EviAtlas, a researcher will be able to shorten the time needed to generate key plots and maps of the previous work they're analyzing. What might have taken a full days work in the past could now be condensed into about an hour. EviAtlas is still in development, and is open to pull requests: <u>https://github.com/ESHackathon/eviatlas/</u>.

Output 2: ROSES (RepOrting standards for Systematic Evidence Syntheses) website overhaul

The ROSES forms were developed to improve the standards of evidence synthesis reporting and the transparency of the methods used for reviews and maps. The website (<u>www.roses-reporting.com</u>) aims to help adoption of the ROSES forms as well as improve the barrier to entry on using the ROSES forms in the least obtrusive way possible. In this project we've been working to improve the user experience, increasing efficiency and helping reviewers get the most out of the forms and website. We will be releasing the new version of the ROSES website shortly.

Output 3: A function for dynamically generating analysis reports based on model outputs within metafor

The function dynamically generates an analysis report (in html, pdf, or docx format) based on a model object. The report includes information about the model that was fitted, the distribution of the observed outcomes, the estimate of the average outcome based on the fitted model, tests and statistics that are informative about potential (residual) heterogeneity in the outcomes, checks for outliers and/or influential studies, and tests for funnel plot asymmetry. A forest plot and a funnel plot are also provided. References for all methods/analysis steps are also added to the report and cited appropriately. Additional functionality for reports based on meta-regression models will be incorporated soon. The function is already part of the 'devel' version of the metafor package and can be found here: https://github.com/wviechtb/metafor.

Output 4: PDF annotation and data coding/extraction tool

Extraction of content from articles, also known as coding, is an important part of evidence synthesis, especially for meta-analysis that require coding of multiple predefined parameters that are to be extracted from articles. This task is usually tedious therefore multiple people, potentially including external helpers may involve in coding. Software tools that assist efficient content extraction and also enable indexing of extracted context against the field labels, are highly desirable. The most significant barrier against such tools is that the majority of the articles are available in *pdf* format, because contents in *pdf* files are embedded in highly abstract and protected manner. The main contribution of the prototype is accessing contents in *pdf* articles selectively. The tool is built around ReactJS JavaScript framework (https://reactjs.org/), therefore suitable for deploying in a local virtual web-server in a desktop environment or in a centrally hosted web-server, as a web application. This application takes in a *CSV* file with fields to be extracted as headers and loads *pdf* files from a server folder. Then the coding can be performed using a right-click menu that brings up list of fields, and then saves the fields, selected contents / values on in the *pdf* and any user comments back to the *CSV* file as a new raw per single *pdf*.

Output 5: KeywordX (Search strategy support tool)

Defining a good search strategy for systematic reviews can be a particularly challenging task. Some of the problems encountered are: when asking two people for a strategy they will get totally different outputs, the number of hits is prohibitively high, there are missing relevant references because a specific keyword was omitted, few means of validating search strategies exist, it is difficult to adapt the strategy for other databases, errors may be introduced when adapting strategies between databases, etc.

The project focuses on adding keywords not occurring in a strategy (finding the keyword X), so you don't miss a relevant reference.

The project is based on several smaller projects, for reusability. <u>Search counter</u> an API receiving a search strategy return the number of hits in several databases, <u>Search parser</u> an API receiving a search strategy in text format and identifies the parts of the search: Boolean operators, words, search in operator, etc). <u>Keywords extract</u> giving a text (for example a RIS document), will extract the relevant keywords for that text. <u>Search front</u> is the web interface for KeywordX project.

We hope to improve the project and tackle most of the previously described problems. If you want to collaborate, feel free to make pull requests.

Output 6: Tool to extract reference lists from PDFs

Full-text PDFs are almost always the most reliable source of information from academic articles. Even though several resources allow for the extraction of data from full-text documents, most of the time the information is incomplete, inaccurate, or not available. PDFs were created to look great, not to extract data from. So, when you try to copy/paste from PDF you often get unexpected results.

In this first version the project allows users to easily copy text from a PDF and attempts to automatically identify the references. You can try the current version at: <u>http://35.196.139.104/</u> and check the repository here: <u>https://github.com/ESHackathon/pdf-to-text</u>.

Output 7: Paperweight (using natural language processing to improve search queries)

Paperweight, driven by a combination of natural language processing (NLP) algorithms. In the evidence synthesis process, the first steps typically require reviewers to manually build a database of articles and journals they want to summarize. This process entails an exhaustive search of <u>Google Scholar</u> using manually chosen keywords. This approach is vulnerable to bias since the reviewer might be more likely to find certain articles or journals in their review over other ones, depending on the selected search keywords. Tackling this problem, Paperweight seeks to remove the need for a reviewer to manually choose keywords to form their search queries.

In essence, <u>Paperweight</u> takes as input an <u>RIS file</u> of publications (which can be exported from Scopus or Web of Science) that the reviewer is confident should be included in the final evidence synthesis. Then, Paperweight outputs a list of summary keywords and phrases, extracted using the RAKE and TextRank NLP algorithms, that the reviewer can then use for their search query. In this way, the reviewer need only identify several publications that they know will be included in their final review to retrieve a larger list of publications that should also be included in the review. Although Paperweight does not claim to remove all bias, as the reviewer ultimately still needs to decide on an initial collection of publications, the team still believes it can meaningfully reduce early stage bias in evidence synthesis. Paperweight is still under development and is open to pull requests at: https://github.com/ESHackathon/paperweight-python.

Output 8: Thalloo Evidence-Mapping: A Jekyll Theme for Visualisation of Datasets

This project provides an easy-to-use template for web visualisations of environmental evidence maps. *Thalloo* is a combination of map components and a Jekyll theme that enable quick, simple, and customisable deployment of a web-based tool to display evidence maps. The framework has the following features: i) Visual clustering and display of categorical data. Given a display category (e.g. crop, commodity), and a custom colour palette, points are displayed on a map. Depending on the zoom level and extent, points are clustered dynamically for best display. Any cluster can be selected to see the full metadata about the evidence points it contains; ii) Filtering. Data can be filtered by property in real time, using multiple filters within a property, and using multiple properties to filter; iii) Slicing of dimensionality. Given continuous data (e.g. publication year, time, or an effect size), the map allows real time 'slicing' of the dataset along one or many dimensions; iv) Abstract and funding logos. Provide attribution to your funders and partner institutions by including their logos at the top of your map view.

The mapping components are written using D3.js. The website is static, and can be compiled using the Jekyll static site builder. All code is TypeScript, but to create your own *Thalloo* site no coding experience is required. You can host one or many evidence maps using GitHub Pages' free hosting. Visit the GitHub page here: <u>https://github.com/AndrewIOM/thalloo</u>.

Output 9: A tool to fill in missing information from incomplete references

Often, citations downloaded from bibliographic databases and other resources, such as Google Scholar, are missing certain details like abstracts or volume/page details that are important for a variety of reasons, such as screening in systematic reviews or locating full text documents. This functionality is intended to be used for filling in missing information from a set of citation files, including abstracts. More information can be found on Github here: https://github.com/ESHackathon/fill_in_incomplete_refs.

Output 10: An academic paper on using R for evidence synthesis

R is a widely-used, open source programming language and statistical environment. Users are able to contribute add-ons to R functionality in a standardised way by developing new software 'packages'. However, identifying which packages are most useful for a specific task can be challenging, particularly for evidence synthesis (ES) projects which typically include a number of discrete tasks, many using packages that may have been designed for other purposes. Consequently, a valuable tool for future researchers (and hackathons) would be a 'map' of available software packages, showing how those packages apply to ES. This would help guide new users through effective workflows, as well as identifying parts of the evidence synthesis process that are currently well supported in R, or conversely, in need of further software development. This project is currently in the data collection phase, wherein participants systematically search for R packages of potential value to ES projects and catalogue their findings in a structured way. The intended output is an academic article describing our findings, linked to a live database of R packages, the functions they contain, and the specific ES tasks that they each solve.

Output 11: An academic paper on the limitations and biases of commercial bibliographic databases - and a suggested alternative

Reliable evidence synthesis requires access to a comprehensive, unbiased body of literature that can be searched for relevant information. Systematic reviewers typically search multiple (upwards of 10) bibliographic databases to identify sets of search results that might yield relevant results. Access to these databases is often restrictively expensive, hampering efforts to synthesise evidence by smaller organisations and groups from low- and middle- income countries, for example. When reviewers export references from these databases they must typically do so in small batches (this supposedly stops people from replicating commercial databases for profit): for Web of Science this must be done in batches of 500, which can add considerable time to a review with 20,000 search results or more! Finally, databases such as Web of Science exacerbate publication bias by selecting journals and publishers that are perceived to be of 'high impact', for example using citation indices. So, these resources may be expensive, hard to use, and offer a biased selection of evidence. In order to facilitate evidence synthesis and to reduce bias in how information is indexed and found, we call for the production of an Open Source, Open Access on-

stop-shop database that catalogues all known academic research. Since tables of contents are freely available online, technology exists that can produce such an important and useful tool.



7. What are the plans for future Evidence Synthesis Hackathons?

Success of ESHackathon2018

The inaugural ESHackathon was unanimously declared a success by the participants and the organisers. The event produced a large number of applications and papers, connected a diverse and experienced group of experts, and shared capacity across a broad suite of previously disconnected communities. The participants will continue to develop and refine their apps with the support of the organisers. Externally, the hackathon was met with great interest, and many users of the apps produced have already made themselves known to us.

Future hackathons

Having demonstrated a highly successful proof-of-concept, we strongly hope there will be future ESHackathons. Due to the positive comments regarding the non-competitive nature of the event, amongst other things, we see no reason to suggest changes to the format.

The ESHackathon2018 left the scope and target problems that participants could focus on entirely open. This worked well, giving participants a real sense of ownership of the solutions they produced. However, future iterations of the ESHackathon could potentially be more focused, for example dealing only with issues relating to search strategy support, or data extraction. One alternative to combine these approaches would be to invite a larger number of participants and focus on several key steps in the synthesis pathway.

There has been a rapid increase in interest in evidence synthesis technology, and there are now several review management tools that support reviewers throughout the process of conducting a review (e.g. EPPI Reviewer 5, which will be Open Source when released). Recognising this, future hackathons could connect with these tools to avoid possible redundancy and maximise the utility and uptake of the tools produced during the hackathon.

We were lucky enough to have a large sum of funding available for ESHackathon2018. That said, this sum is not substantially more than would be expected for a workshop organised for 30 people travelling internationally (approximately 250,000 SEK). Future hackathons could attempt to cut costs by asking for organisations to contribute to funding travel for their staff as participants. However, this might significantly affect the number and type of applications that are received. An alternative is to seek dedicated funding to cover full costs of participation in the event. Given that the attendees are donating their time free-of-charge, we strongly think this is an important means of ensuring a high calibre and positive mindframe for those participating.

Acknowledgements

The organisers are extremely grateful to Mistra EviEM and ANU for their generous funding of the event. Thanks, too, go to SEI, GWP and SIWI for generously donating meeting spaces throughout the hackathon. Many thanks to Marc Lajeunesse and Wolfgang Viechtbauer for attending the ESHackathon2018 and presenting their software packages. Thanks to the assistant experts for helping to steer the participants down the right track every now and then. Finally, thanks to all the excellent participants, without whom, this would not have been possible.

Image credits

Images were taken by organisers and participants and rights remain with the photographers.