

# Beyond the Dataset: Integrating Public Voices in Data Science

## AUTHORS

Ana-Paula Rubio<sup>1</sup>, Janette Dunn<sup>1</sup>, Janice Borthwick<sup>1</sup>,  
Sheila Saunders<sup>1</sup>, Claire Tochel<sup>1</sup>, Heather Anderson<sup>2</sup>,  
Lorna Coull<sup>2</sup>, Malihe Javidi<sup>1</sup>, Heather Cloete<sup>2</sup>, Alice  
McTrusty<sup>2</sup>, Jonathan Penny<sup>1</sup>, Tom MacGillivray<sup>2</sup>, Andrew  
J Tatham<sup>2,3</sup>, Tracy Jackson<sup>1</sup>, Carol Porteous, Niall  
Strang<sup>4</sup>, Baljean Dhillon<sup>3</sup>, Miguel O. Bernabeu<sup>1</sup>

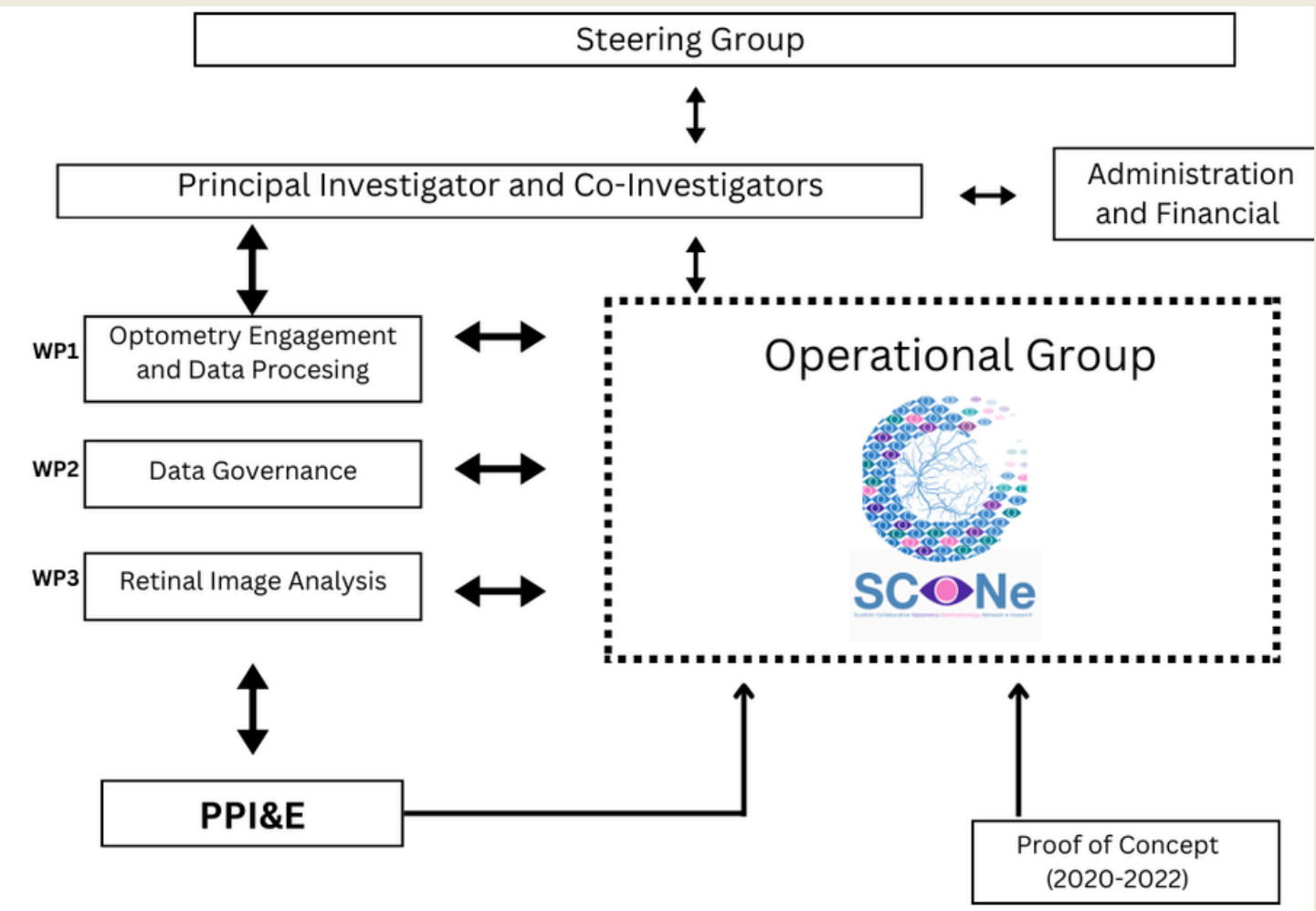
## AFFILIATIONS

1. Centre for Medical Informatics, University of Edinburgh, Edinburgh, UK
2. Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
3. Princess Alexandra Eye Pavilion, NHS Lothian, Edinburgh, UK
4. Glasgow Caledonian University, Glasgow, UK

*Big data and AI offer powerful opportunities for early disease detection, but their success depends on more than just technology — they rely on public trust, transparency, and meaningful involvement. This project shows how Patient and Public Involvement and Engagement (PPI&E) can strengthen the future of data-driven health research.*

## Introduction

SCONe is a national initiative collecting retinal images from Scottish optometry practices and linking them to health data to support early disease detection and research. While technological infrastructure is central, public trust and involvement are essential to its success. This work presents the co-development of SCONe's PPI&E strategy, created with public partners to ensure the project reflects societal values, builds trust, and offers a model for meaningful involvement in data-intensive health research.



## Objective

To design and implement a meaningful PPI&E strategy within the SCONe project, ensuring that public voices shape the development, governance, and communication of a large-scale retinal image repository aimed at supporting early detection across a range of health conditions.

## SCONe's PPIE approach

- **Strategic Early Involvement:** The SCONe PPI group was formed at the project's outset, shifting from tokenistic inclusion to genuine co-production in data science research.
- **Three-Tiered Framework:** The PPI&E strategy distinguishes between involvement, engagement, and consultation, aligning with the IAP2 Public Participation Spectrum to clarify roles and expectations.
- **Evolving and Inclusive Approach:** The strategy embraces flexibility, enabling responsive and context-specific participation that builds public trust and shapes impactful research outcomes.

## Challenges in Data-Intensive Health Research: Lessons from SCONe

### • Scale and Complexity → Inclusion

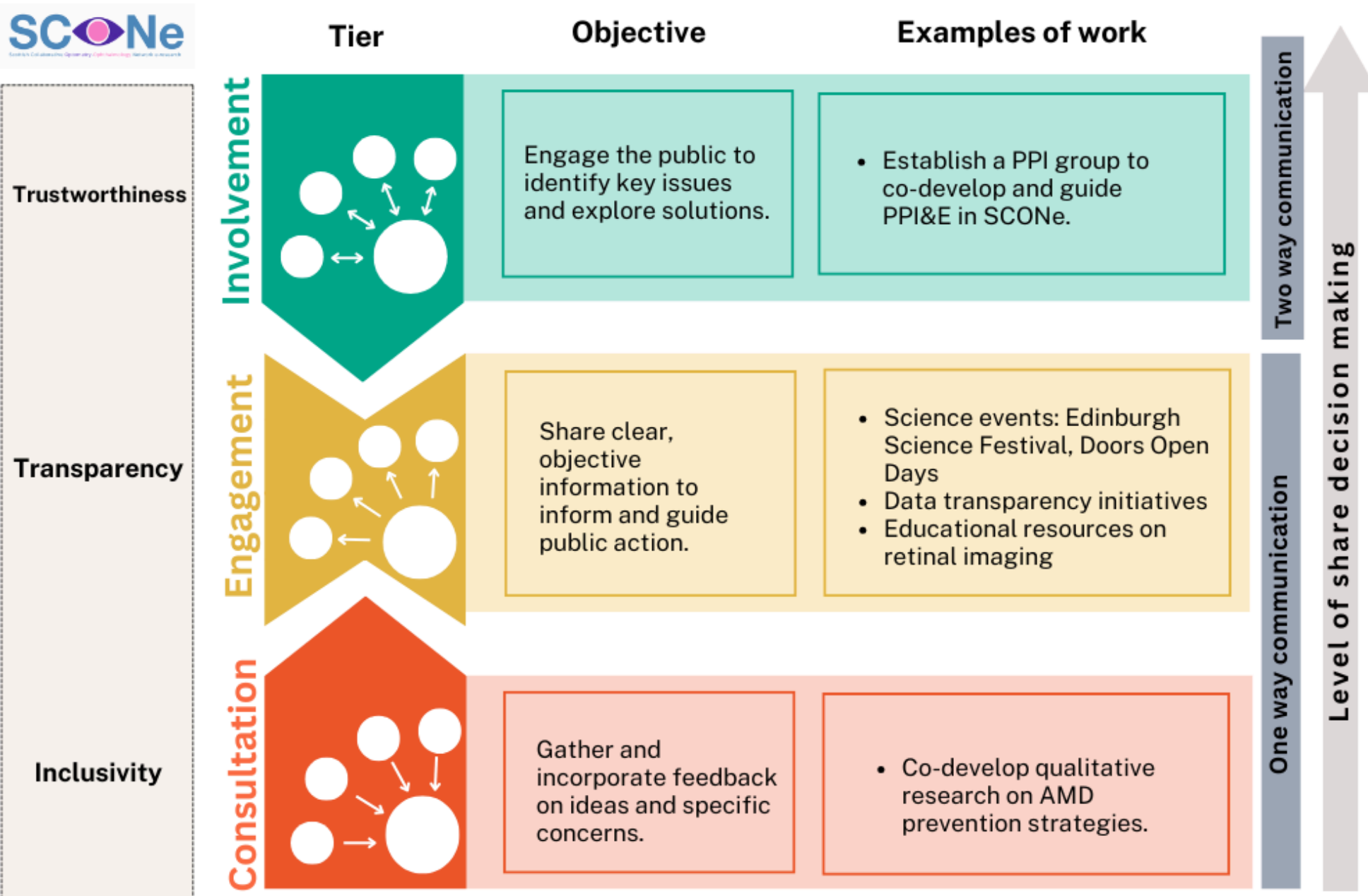
Data science research often risks excluding diverse public voices because of the sheer scale and complexity of datasets and methods. SCONe addressed this by embedding inclusion at the foundation of its PPIE strategy, ensuring that a wide range of experiences and perspectives are represented throughout the research process.

### • Governance, Control, and the Black Box → Transparency

Health data governance is difficult to navigate for both researchers and the public. At the same time, AI models often behave as "black boxes," making it hard to understand how decisions are made. SCONe tackled these issues by fostering transparency through bidirectional communication, building mutual understanding about governance structures and the ethical challenges of AI.

### • Trust and Public Support → Trustworthy Systems

In data science, trust cannot simply be requested—it must be earned through genuine interaction and openness. SCONe built trustworthy systems by prioritizing meaningful public involvement and ongoing engagement, creating relationships that support and sustain trust over time.



## Conclusion

PPI&E marks a shift from researcher-led studies to collaborative research shaped with and by the public. In complex data science projects like SCONe, it helps build transparency, trust, and relevance.

While there's no one-size-fits-all model, this flexibility is a strength. SCONe shows that meaningful involvement must be context-driven, evolving over time to reflect the needs of both the public and the research. True impact comes from sustained, adaptable collaboration.