Wide-Field Ethnography: Studying Software Engineering in 2025 and Beyond

David Socha¹, Robin Adams², Kelly Franznick², Wolff-Michael Roth⁴, Kevin Sullivan⁵, Josh Tenenberg⁶, and Skip Walter⁷
¹University of Washington Bothell, USA, socha@uw.edu. ²Purdue University, USA, rsadams@purdue.edu. ³Blink UX, USA, kelly@blinkux.com. ⁴University of Victoria, Canada, mroth@uvic.ca. ⁵University of Virginia, USA, sullivan@cs.virginia.edu. ⁶University of Washington Tacoma, USA, jtenenbg@uw.edu. ⁷Factor, Inc., USA, skip.walter@factor10x.com.

ABSTRACT
This paper presents a vision of how the Internet of Things will impact the study of software engineering by 2025 and beyond. This inquiry is guided by the following questions. What will it mean to be able to deploy hundreds of sensors and data collectors running concurrently over months to gather very large and rich datasets of the physical, digital, and social aspects of software engineering organizations and the products and services those organizations create? How might such datasets change the types of research questions that can be addressed? What sort of tools will be needed to allow interdisciplinary communities of researchers to collaboratively analyse such datasets? How might such datasets help us understand the principles governing the interplay of physical, cyber, and social aspects of software engineering and its products, and automate aspects of such systems?

Categories and Subject Descriptors
D.2.0 [Software Engineering]: General.

General Terms

Keywords
Wide-field ethnography, Empirical Software Engineering, Sensors.

1. INTRODUCTION
This paper argues that the combination of increasingly rich sensing technologies, highly scalable computing, and advanced analytics presents radical new possibilities for observing, recording, analyzing, understanding, and improving complex physical-cyber-social systems (PCSSs) such as software engineering teams and organizations and their products and services. We can capture high-resolution, multi-perspective, large-scale data of both physical phenomena (e.g., people’s speech and physical actions), and digital phenomena produced by human-machine interactions (e.g., software commits to version-control repositories) enabling rich inquiry into individual and team work in developing PCSSs. Philosophers, psychologists, and social scientists have argued convincingly for the distributed, social, and embodied nature of cognition [1]–[5], highlighting how people make public and available to one another (in their speech, writing, gestures, gaze, body orientation, and actions) the resources necessary for sustaining the visible social order of their joint activity. In the very act of working in PCSSs to create PCSSs, software engineers make visible to one another the interplay of the physical, digital, and social aspects of the systems they create. This affords social science researchers to use sophisticated research-oriented PCSSs to capture this activity for later analysis. It also allows computer scientists to create and use automatic approaches, such as generation and analysis of transcripts from voice recordings with attribution of utterances to individuals in the subject community, complex event detection, and sentiment analysis, to provide additional dimensions of meta-data for analysis by investigators, or to automatically control aspects of the PCSSs being observed. These capacities could have far-reaching implications for how we study software engineering in particular, and the science of complex social systems in general.

We use the term wide-field ethnography (WFE) to refer to this approach of gathering and collaboratively analyzing large, multi-modal, multi-stream datasets of PCSSs in action; WFE uses a wide variety of sensors and data collectors and ethnographically informed observations to gather data from a wide set of data fields (space, time, modalities) across a wide expanse to enable collaborative analysis by a wide set of disciplinary fields.

2. WFE DATASETS
This section examines some existing WFE-like datasets in order to ground the salient features of WFE presented in Section 3.

Over an 11-day period in February 2014 Socha used nine GoPro cameras, six high-quality Zoom H2n audio recorders, screen capture software, and a hand-held camera – all running concurrently – to gather six terabytes of video (380 hours; 981 files), audio, full-room time-lapse imagery (every 5 seconds), screen capture (292 hours), photographs (thousands), and field notes of the BeamCoffer software developers collaborating in situ [6], [7]. BeamCoffer is a pseudonym for a Seattle area software development organization founded in 2004 whose software product helps friends and family share information. By 2013 BeamCoffer employed about 50 people working in a large open...
gathered hundreds of hours of videos of entrepreneurs presenting
shared data through a diverse set of theoretical lenses. Walter
in which an inter-
critiques for the 2014 Design Thinking Research Symposium
In 2014
higher profits.
added analysis, do better work, have more fun, and
amount of
between
understanding in respo
streams in real time. This allows them to evolve their
distributed teams of researchers, who chat and annotate the
processors, and other data collectors used in user experience
sessions. Cloud services translate the streams into multiple
capture, time-lapse photography, photos from a hand-held camera,
field notes, and interviews. It is multi-stream, containing parallel
recordings (e.g., audio, video) streamed into thousands of files.
These datasets are at the core of WFE, and their characteristics
lead to a number of opportunities.

Physical-cyber-social systems (PCSSs) are the engines and
products of work. WFE is designed to make visible these three key
aspects: the physical world in which we live and through which all of our interactions flow; the cyber systems that are increasingly
foundational to much of today’s work and are changing the nature
of work and what it means to be human; and the social systems in
and through which our work is done that define and create our	norms and values, and that give meaning to our lives.

These three dimensions are essential to almost all systems that
humans relate to and inhabit: the teams and organizations we live
and work in; the products and services that we create and use; and
the communities of users, influencers, and purchasers of the
products and services we create. A team of software engineers is a
PCSS; and so is a team of healthcare workers caring for infants in
a neonatal intensive care unit, or a community of researchers
collectively analyzing a WFE dataset. To study software engineering is to study PCSSs.
Gathering widely across modalities, systems, space, and time with multiple data recorders running concurrently helps researchers embrace emergence:

- It allows researchers to follow work over space and time. We followed a single design discussion as it moved around the office space from the “huddle” area to couches to a conference room, even as it changed participants [6].
- It enables researchers to pursue a wide range of research questions using a variety of units of analysis, such as an individual, task, type of work (e.g., triage), location in the room, time of day, topic (e.g., power structures), or practice (e.g., standup meetings).
- It supports inter-disciplinary analysis by communities of researchers using a diversity of theoretical and analytical frameworks.
- It allows researchers to widen the observational field beyond what is possible with human observation. For example, multiple cameras covering the BeamCoffer developer stations allowed us (a) to describe and quantify the nature and amount of interactions between one pair of programmers and other people around them and (b) to question the frequent focus on the “pair” as the analytic unit for pair programming research [10].
- It helps mitigate against researchers’ preconceptions of what type of data to collect. For BeamCoffer, we assumed design would occur primarily in “design” meetings, but having so many cameras in so many places allowed us to see how design was happening everywhere [6].
- It supports longitudinal analysis. Having 21 days of data over the full 17 month case study allowed us to see the diversity of the ways in which the developers use different sections of the whiteboard wall near the huddle area to mediate their work [7].
- It helps reduce the set of hypotheses that can be evidenced by providing sufficient data to disprove hypotheses [11].

WFE enables collaborative and interdisciplinary analysis. Having such large amounts of data covering so many different types of material supports collaborative analysis by communities of researchers who employ a wide range of diverse perspectives [12], analytical tools, and theoretical stances to produce a richer understanding of the system under study. Highly complex and multi-faceted systems like software engineering go beyond the capacity of single researchers or even a group of researchers operating under a single paradigm or research agenda.

Real-time analysis of WFE datasets provides high value for product development and collaborative analysis. The real-time analysis provided by BlinkUX’s Feedback Panel’s allowed researchers and product teams to change the product on the fly, making the best use of the time and resources available. It created an exciting environment and increased collaboration.

Off-line analysis affords research using analytical techniques that require substantial off-line efforts, such as interaction analysis [13] and data mining. WFE datasets could be used for shared analysis by researcher workshops, such as was done in the NSF-sponsored Studying Professional Software Design workshop [14].

Multi-modal analysis. Datasets of qualitatively rich unstructured data (e.g., video) support a type of analysis called multimodality: “in contrast to frameworks that analytically sequester communicative modes like speech and gesture both from each other and from the material world, the multimodal approach instead assumes semiotic complexity as the prerequisite, irreducible condition for communicative social action” [15, p. 1966]. Most research analyzes only one or a few modes, and it is common to focus on speech when only audio recordings are available. Increasingly, researchers are showing interest in studying speech along with the accompanying gesture, perceptual structures (such as gaze and body orientation), body movements, and different parameters associated with the physical production of speech such as speech intensity, pitch and pitch contour, speech rate, and higher order frequencies all of which have psychological and sociological correlates (e.g., [16]). By 2025, we can expect datasets with many more modes of data from many more types of sensors.

4. WFE tools
WFE datasets are too large and complex for a single researcher to easily understand, navigate, browse, filter, annotate, and analyze without tool support. What principles, practices, and digital tools are needed to afford a community of researchers to collaboratively analyze a WFE dataset and iteratively benefit from each other’s work during the process of analysis? How can tools allow a researcher who coded gestures in some videos to upload those coded sections so that they are available for other researchers to use in their analyses? How might such a system provide a multiplicative benefit to a community of researchers?

We conjecture that these tools will be organized around the concept of collaborating streams (see Figure 2). A stream is a time-based sequence of data from one particular data source, such as an audio recorder or a code repository. A stream’s data is organized into a sequence of events (moments in time and segments (contiguous portions of time). An audio stream, for instance, contains at least one segment beginning when the audio recorder started recording and ending when the recorder stopped. The stream for a still-photo camera is a set of events (photos). A stream may contain a set of streams, such as a video camera stream recording video footage, audio, and GPS streams. Researchers can create subsets of existing streams.

The stream abstraction is a continuous sequence of data; it hides the implementation details of files storing data. Streams from different data collectors are loosely coupled, making it easier to handle multiple streams and add new types of streams, such as biometric streams, or virtual-reality streams.

Each stream is tagged with meta-data related to its source, such as data quality (e.g., sensor noise and calibration), and provenance information (e.g., tool-chains and methods used to produce the data). Streams can represent original data, such as video data, or derived data, such as transcripts, or annotations.

Streams collaborate when events or segments map to the same

![Figure 2. A set of collaborating streams (e.g., video, audio, screen capture, photo, biometric, documents, annotations). Vertical lines represent segments. Dots represent events.](image-url)
time, or to the same concept. A concept corresponds to a tag (e.g., person, location, task, type of activity, topic), and can be associated with a stream, event, or segment. Tagging can be done manually or via data mining algorithms. In this way, streams can collaborate along an arbitrary set of dimensions.

Figure 3 shows a possible high-level design of this tool system. Multiple sensors gather information about the PCSS under study. Raw data files (e.g., video files) are uploaded to and stored in a cloud-based file system, which automatically extracts stream metadata from raw files, and automatically transforms or analysing aspects of the raw files. Metadata and tags are stored in a “collaborating stream” database, providing a multi-dimensional relational structure for streams to help researchers navigate and explore the space, time, and modality of PCSS collaboration.

This design supports both researchers using manually intensive analysis techniques like interaction analysis and researchers doing data mining research. Interaction analysis’ slow reading of unstructured qualitative data such as videos from ethnographically informed studies is particularly good for generating insights about PCSSs. Techniques like machine learning from streams of structured quantitative data are particularly good for learning PCSS whose cyber aspects automatically process the WFE data to provide feedback on and control aspects of PCSSs [17]. Insights generated by slow reading inform automation, and automating mundane aspects of slow reading and results from data mining give more time and material for slow reading analysis. Supporting both communities provides opportunities for integrating insights gained by these different analytic approaches.

5. WFE AND SE IN 2025

WFE datasets are coming. By 2025 they will be huge and provide unprecedented levels of visibility into the multi-modal nature of collaboration and cognition in software engineering organizations. The software engineering research community is uniquely situated to enable the creation and use of WFE datasets by reflexively using the very topic we study, software engineering, to create the necessary tools. Creating wide-field ethnography tools not only helps us better understand and automate important aspects of software engineering organizations and the PCSSs they create but also benefits a wide spectrum of researchers who study PCSSs.

6. REFERENCES