Guide for *Kanban - the Flow Strategy*™

September 2019

Developed and sustained by:

- John Coleman on behalf of Orderly Disruption Limited
- Daniel Vacanti on behalf of Daniel S. Vacanti, Inc.

Purpose

This guide contains the minimum set of rules for Kanban - the Flow Strategy™ (KFS). For simplicity, in the rest of this document when we say *Kanban*, we mean *Kanban - the Flow Strategy™ (KFS)*.

By reducing Kanban to its essential components, the hope is that this guide will be a unifying reference for the community. By building upon Kanban fundamentals, the strategy presented here can accommodate the full spectrum of development and organizational challenges. Followers can layer in additional practices as needed as long as the minimum requirements in this guide are adhered to.

Using Kanban

The flow-based perspective of Kanban can enhance and complement your approach to knowledge work and help your team move towards a shared purpose/direction. This guide provides the basic structure for Kanban, which the reader can augment using models such as Evidence Based Management (Scrum.org, 2019), Cynefin Sense Making (Snowden, 2019), Theory of Constraints (Goldratt, 1999), and scaling (to name a few). Consider this guide as a reference model.

Adopting Kanban comes with committing to optimize the flow of value. To apply Kanban, teams must be willing to:

1. actively manage Work in Progress (WIP),
2. use Workflow Policies to support flow, and
3. avoid local optimization.

Kanban has a hyper-focus on transparency, visualization, learning, and flow to deliver optimal value. Tightening the feedback loop with Kanban is a proven strategy for empirically improving a process.

There is an often-heard adage that “Kanban starts with what you do now” meaning that a team can evolve its practices incrementally. While that can be true, the statement can be misleading because without adopting the fundamental components, it’s not Kanban (regardless of where you start). Adopting all of Kanban’s practices immediately may require wholesale change to existing processes--which is perfectly fine. Because without active management of WIP and compliance with Workflow Policies to support flow, all that is left is visual management, which is useful but inadequate for optimizing the flow of value.

Depending on the context, professional knowledge workers such as those working in the areas of marketing, human resources, finance, legal, software, hardware, telecoms, energy, supply chain management, pharma/healthcare, manufacturing, automotive, and innovation can benefit from the application of Kanban along with what they do now.

Definition of Kanban

Kanban is a strategy for optimizing the flow of value through a process that uses a visual, work-in-progress optimized pull system.

Central to the definition of Kanban is the concept of "flow." Flow is the movement of customer value throughout the product development system. Kanban aims to optimize flow by improving the overall effectiveness, efficiency, and predictability of a process.

Workflow

Workflow is a fundamental concept of Kanban, and that cannot be understated. All other elements of this guide depend heavily on how a team specifies its definition of Workflow. The definition should change as the team empirically discovers better ways of flowing work. Consistently defining Workflow is necessary when applying other elements in this guide (e.g., flow measures).

Optimizing flow requires defining what flow means in your context. Each team must create its definition of Workflow containing the following elements:

- Defined points at which the team considers work to have started and to have finished.
- A definition of the individual units of customer value (Work Items) that are flowing through the team's system.
- A definition of the Workflow states that the Work Items flow through from start to finish (of which there must be at least one active state). This includes:
    - Identifying Work Items that are not yet in an active state as *not started*.
    - Identifying Work Items entering the active state (*started*) as *Work in Progress* (WIP).
    - Identifying Work Items that have passed through all of the active states planned as *finished* or *Done*.
- Workflow Policies about how work can flow through each state (which may include items from a definition of Done or *ready* for each state of the Workflow, and pull policies).
- A definition of how WIP will be optimized

- A set Service Level Expectation (SLE) that communicates a forecast of how long it should take to complete Work Items.

In summary, Workflow includes a shared understanding within the team of how it defines work (Work Items), the start state of the process, the active states for the Work Items, and the finished state of the process.

Note that the states in the definition of Workflow might go upstream of the team. The states might also go downstream of the team. Workflow can go as far as the team's sphere of influence can stretch. Finally, a particular Work Item might not flow through all of the active states, and might not even flow sequentially through the active states.

If the three Kanban practices are adhered to then the Workflow can be referred to as the "Kanban system".

Service Level Expectation

A Service Level Expectation (SLE) forecasts how long it should take a given item to flow from start to finish within your Workflow. The SLE itself has two parts: a period of elapsed time and a probability associated with that period (e.g., "85% of Work Items will be finished in eight days or less"). The team should base its SLE on historical cycle time, and once calculated, should post it on the Kanban Board. If historical cycle time data doesn't exist, the team should make a best guess until there is enough historical data for a proper SLE calculation.

Rightsizing

When the team begins to look at its next work priority, it may consider rightsizing if the Work Item does not fit the SLE. If the next Work Item is too big, it may have to be revised or decomposed (broken down). Rightsizing should be done just-in-time as there is no reward for having a Backlog full of Work Items that fit within the SLE; quite the opposite, it would be a waste to refine items we might never work on.

Optional rightsizing, refinement, and decomposition help with getting the team's next priority to fit the SLE. Rightsizing, in particular, refers to whether Work Items:

- fit the SLE,
- are too big for the SLE and therefore require refinement and decomposition or proof-of-concept-type-research

Kanban - the Flow Strategy™ Practices

Teams achieve flow optimization by using the following three practices:

- Defining and visualizing the Workflow
- Actively managing WIP
- Reviewing and improving the definition of Workflow

Visualization of the Workflow - the Kanban Board

Visualization using the Kanban Board is the way the team makes Workflow transparent. The Kanban Board must prompt the right conversations at the right time so that one can proactively suggest opportunities for improvement.

Active Management of Work Items in Progress

Optimizing WIP is a necessary component to achieve flow, but it alone is not sufficient. The second practice to establish flow is the active management of WIP. Active management can take several forms, including but not limited to, the following:

- Optimizing WIP
- Responding quickly to relatively-aged/blocked Work Items
- Making sure that Work Items are only pulled into the Workflow at about the same rate that they leave the Workflow (this assumes the absence of a stable and ridiculously expensive constraint)
- Ensuring Work Items aren't left to age unnecessarily and are completed according to an established SLE
- Unblocking work that piles up in a column or columns

It is the team's responsibility to ensure the continuous proactive, active, and reactive management of WIP.

Optimizing WIP

WIP refers to the Work Items the team has started but has not yet finished. Teams using Kanban must explicitly control these in-progress Work Items from the time they consider them *selected*/*committed* until the time they consider them *finished*. That control is usually represented as numbers or (preferably) slots/tokens on a Kanban Board and they are called *WIP limits*. A WIP limit can include Work Items in a single column, several grouped

columns/lanes/areas, or a whole board. Once the team has established a WIP limit, it refrains from pulling/selecting more than that number of Work Items into a given part of the Workflow.

The team (regardless of organizational level) controls what the limits are and how it will apply them.

A side effect of optimizing WIP is that it creates a pull system. It is called a pull system because the team starts work on an item (pulls) only when there is a clear signal that it is time to do so. A push system, on the other hand, demands the team starts work on a Work Item whenever someone requests it. When the WIP drops below a limit that the team has defined, that is a signal to start/select/pull new work. Note, though, that instead of putting a Work Item on the next state's queue (if it's to another person it's effectively a push), it is preferable to ask, "I'm done with this Work Item; who can take it from me?" (Wovchko, 2019).

Work gets replenished when capacity allows, or whenever the team chooses. Optimization of WIP not only helps Workflow but often also improves the team's focus, commitment, and collaboration.

Review and Improve the Workflow for Continuous Improvement

Workflow Policies are the rules the team decides to use for its Workflow. The team makes these policies explicit, revising them when necessary. Small changes to those Workflow Policies can have a material impact on how the team performs overall. Workflow Policies should include but are not limited to a review of the exceptions and signals, including relative Work Item aging.

Kanban Measures and Analytics

The application of Kanban requires the collection and analysis of a minimum set of flow measures. These measures are necessary for the active management of WIP. They also make the flow transparent and enable flow-oriented inspection and adaptation. These measures are a reflection of the current health and performance of the team's approach. They will also point to interventions that can improve the team and the value it delivers.

The Basic Measures of Flow for Kanban

The four mandatory measures of flow that teams need to track are as follows:

- WIP: The number of Work Items started but not finished (according to the definition of Workflow).

- Throughput: The number of Work Items finished per unit of time. Note the measurement of throughput is the exact count of Work Items.
- Work Item Age: The amount of elapsed time between when a Work Item started and the current time.
- Cycle Time: The amount of elapsed time between when a Work Item starts and when a Work Item finishes (it is ideally in use) plus one day.

Calculating cycle times and Work Item age requires the team to (at a minimum) track the start date and finish date of each Work Item. You must monitor these measures continually. There are other flow measures that the team may want to examine, but these are the minimum required.

History and Acknowledgments

The set of Kanban practices commonly referred to as Kanban mostly originated on a team at Corbis in 2006. Those practices quickly spread to encompass a large and diverse international community that over the years, continued to enhance and evolve the approach.

Various books can aid those adopting Kanban. Many are listed as references at the end of this guide. However, the following are three we recommend as a foundation:

- *The Principles of Product Development Flow: Second Generation Lean Product Development*. - Reinertsen, D. (2009)
- *When Will It Be Done?* - Vacanti, D. (2019)
- *Practical Kanban: From Team Focus to Creating Value,* 1st ed. - Leopold, K. (2017)

Thank you to these and the other practitioners who have contributed to making Kanban a viable and successful strategy.

A special note of thanks to:

- Daniel Doiron for permission to include a discussion about combining Kanban - The Flow Strategy™ with Tameflow in one of this guide's optional addendums.
- Dave Snowden for permission to include an optional addendum about combining Kanban with the Cynefin Framework for deeply complex work
- Deborah Zanke for editing
- Emily Coleman, 15 years of age at the time of writing, for the inspiration to consider societal value in this guide
- Vladimirs Ivanovs for reviewing

Optional Addendums

Addendums have been included with this guide to outline some ways to enhance specific Kanban practices, which are useful in some contexts. Unlike this guide, the addendums represent optional practices that teams may find useful when implementing Kanban. The addendums will evolve, like this guide. We are learning.

Coleman, J. (2019). *Addendum to Kanban the Flow Strategy™ - implementation options & details*. 1st ed. [ebook] London, UK. Available at: https://kanbanguides.org..

Coleman, J. (2019). *Guide for Kanban for Complexity™ (Kanplexity™) - an addendum to the Guide for Kanban the Flow Strategy™*. [ebook] London, UK. Available at: https://kanbanguides.org..

License

Please send specific and clear reports of mistakes/suggestions in this guide and suggestions to info@ace.works. Thank you.

Definitions

Backlog

A Backlog is a collection of Work Items. Some teams may use similar terms *input queue* or *pool of ideas.* Kanban isn't just for product development---the Backlog can contain a range of Work Item types. Twenty-first-century Backlogs include Work Items for initiatives such as problem statements, hypotheses, research, experiments, product development, service delivery, failure demand reduction, technical debt (Alliance, Letouzey and Whelan, 2019) reduction, jobs to be done, or a set of projects/programs. Regardless of the number of teams involved, there should be one Backlog only. The team prunes the Backlog continually, and Work Items get rightsized (if deemed necessary) on a just-in-time basis. Teams must be able to see the Backlog via visualization that can also reflect any hidden or unplanned work that emerges. The Backlog is not ordered. Pull/select from the Backlog does not mean pull/select from the top of the Backlog.

Cycle Time

Cycle Time is the interval between a commitment point and delivery to an endpoint plus one day (to avoid zero cycle times). There can be several Cycle Times on a Workflow. Cycle Time in this guide refers to elapsed days regardless of weekends/public holidays. You might differentiate types of Cycle Time, such as service time, flow time, lead time, time-in-progress, time-in-process.

Flow

Flow is the speed and smoothness of the delivery of Work Items through the Workflow. Achieving flow is a balancing act. You must optimize value delivery and actively manage WIP, while at the same time support the assumptions embedded in queue management theory, such as Little's Law (Little, 2019).

Good flow feels like a well-oiled machine. The team enjoys a sense of calm and gets more work done with minimal stress. Bad flow is working in a chaotic environment where Work Items are rotting in WIP, delivery is random, and the team is stressed and overburdened.

A kanban

A *kanban* (literally meaning *visual signal* in Japanese) is a visual cue that triggers you to replenish/select/start or pull/move a Work Item. For example, the appearance of an open WIP slot on a Kanban Board is a signal to pull in another work item. Cards represent Work Items, and blank slots are the signal (kanban) to replenish/select/start or pull/move.

Kanban Board

A Kanban Board is a visual representation of a team's Workflow and includes Workflow Policies and WIP limits. The value of the Kanban Board is that it makes Workflow transparent and alerts teams when hidden or unplanned work emerges and when Workflow is blocked. It should be accessible to the team at all times.

The Kanban Board is for use at the team level, multi-team/team-of-teams level, multi-team-of-teams level, organizational level, or any other level. Work Items on different levels tend to have natural degrees of granularity.

Job to Be Done

Jobs to be done is an approach to identifying customer needs advanced by Clayton Christensen (Christensen and Raynor, 2003). Its premise is that customers frequently do not know what they want, even when they tell us they do. Often, customers discover what they want by seeing it. Ignoring this results in developing and marketing products and services that underperform.

Asking what job a particular product/service/feature does for the customer helps to identify the underlying need. For example, a taxi service with a centralized dispatch allows a customer to get from point A to point B, but might not meet the entire set of customer needs related to this transaction. What if that service can be faster, more convenient, and safer? By discovering these unmet needs, ride-hailing services such as Uber and Lyft developed peer-reviewed, transparent, friendly, traceable, cashless, and timely transportation options and disrupted the taxi industry.

Throughput

Throughput is the delivery rate of Work Items to a specific state in the Workflow, usually to the done state, but it could be to any endpoint such as *ready for customer acceptance testing*, *deployed*, or *customer impact confirmed*. Throughput is calculated by the number of Work Items to arrive at the delivery point per time period (e.g., per hour, per day, per week). Just as there can be more than one Cycle Time per Workflow, there can be more than one Throughput per Workflow.

Value

"The value of something such as a quality, attitude, or method is its importance or usefulness. If you place a particular value on something, that is the importance or usefulness you think it has." (Collinsdictionary.com, 2019, 2)

In Kanban, we consider customer value (providing something useful) and knowledge value (learning), which has the benefit of reducing risk. One might also consider how to make the world a better place, known as societal value.

Customer and End-user Value

Customers and end-users attempt to forecast and later confirm a positive impact on their worlds as value. At a minimum, outcomes should arise, positive or negative. Activities and outputs are insufficient. Frequency of impact is what teams should aim for.

Societal Value

Value can also be societal value, leaving the world in a better place socio-economically, environmentally, and from a people-safety perspective. Generation of societal value requires will, purposeful passion, focus, mental energy, and a fresh approach to communications.

Knowledge Value

Value may consider the management of risk, such as the cost of doing nothing, opportunity cost, cost of delay (Reinertsen, 2009), cost of discovery, cost of delivery, erosion of failure demand (Seddon, 2019), or the erosion of technical debt (Alliance, Letouzey and Whelan, 2019).

Learning/acquiring knowledge about a risk can lower it, or at the least, make it visible. A proof of concept for a solution may not actually work. Let's assume a Work item is completed according to our Workflow Policies, and it is released. Discovering whether the Work Item we released works or not is knowledge value because we've learned something.

More often than not, knowledge value/learning can be seen as the opposite of risk. How one includes the management of risk does not matter, as long as it is sufficiently addressed in a timely and proactive manner. Organizational agility is not simply about adaptiveness for delivering customer and end-user value, it's also about avoiding waste.

WIP

In this guide, WIP stands for Work in Progress. WIP refers to Work Items that have passed a commitment point and have not yet arrived at a delivery point. As there may be multiple commitment and delivery points in the Workflow, there may be multiple versions of WIP for the Workflow.

Blocking can happen for many reasons. It could be the result of waiting for someone to solve a problem, deliver a dependency, or provide information, for example. A blocked Work Item could also result from moving to work on something else (production downtime, change in priority), or of waiting for something to happen, such as a law to pass, or office space to be ready. It's important to have clear signals and conversations when Work Items are not active.

Some argue WIP should stand for *work in process*. Proponents in this camp contend that some work in progress is actually blocked, queueing or otherwise waiting, so the word *progress* is misleading. The same, though, could be said of *work in process*. We're confident that teams can recognize when a Work Item is stalled. Also, the word *progress* is a better fit with the concept of flow.

Workflow Policy

*"A **policy** is a set of ideas or plans that is used as a basis for making decisions, especially in politics, economics, or business."* (Collinsdictionary.com, 2019, 1)

A routine Workflow Policy outlines how the team defines the moment when Work Items progress from unselected/uncommitted to selected/committed, and from one in-progress state to another, or from in-progress to finished/Done/delivered. The team should create as many Workflow Policies as it deems necessary.

Teams using Kanban write down or in some way visualize their Workflow Policies so that everyone understands them.

The Kanban Board should display all relevant Workflow Policies or direct team members to where they can find them. A minimum set of Workflow Policies is necessary to optimize flow, which usually means supporting the assumptions made in Little's Law for delivery (Vacanti, D., 2019, 2). There are times when the team can (and sometimes should) breach Workflow Policies by exception.

Work Item

A Work Item holds the potential for value. To realize that value, the Work Item must get released. For example, a release could set you on the path to understanding customer value or you could derive knowledge value from the release of a proof of concept. People use words like *theme*, *epic*, *feature*, *story*, and *product backlog item* to capture the different levels of granularity of a Work Item. This guide uses the single term *Work Item.* Whatever one calls it is fine as long as it is delivering customer/end-user/knowledge/societal value.

Only Work Items should be WIP limited. If our Work Items do not deliver value, they are fake. Limiting WIP for fake Work Items reduces the collaborative effort to deliver potential value; people focus less on the common objective and deliver stuff, not value. The result of using fake Work Items is that the team delivers less value less often and it is more likely to be negative value such as for example, failure demand (Seddon, 2019), technical debt (Alliance, Letouzey and Whelan, 2019) or reverse impact to what was intended.

Work Item can also be referred to as "batch" or "item".

References

Alliance, A., Letouzey, J. and Whelan, D. (2019). *Introduction to the Technical Debt Concept | Agile Alliance*. [online] Agile Alliance. Available at: https://www.agilealliance.org/introduction-to-the-technical-debt-concept/ [Accessed 9 Sep. 2019].

Collinsdictionary.com. (2019, 1). *Policy definition and meaning | Collins English Dictionary*. [online] Available at: https://www.collinsdictionary.com/dictionary/english/policy [Accessed 4 Aug. 2019].

Collinsdictionary.com. (2019, 2). *Value definition and meaning | Collins English Dictionary*. [online] Available at: https://www.collinsdictionary.com/dictionary/english/value [Accessed 3 Aug. 2019].

Christensen, C. and Raynor, M. (2003). *The innovator's solution*. [Concordville, Pa.]: Soundview Executive Book Summaries.

Goldratt, E. (1999). *Theory of constraints*. Great Barrington, MA: North River Press.

Little, John D.C. (2019). [online] Available at:https://people.cs.umass.edu/~emery/classes/cmpsci691st/readings/OS/Littles-Law-50-Years-Later.pdf [Accessed 9 Sep. 2019].

The Principles of Product Development Flow: Second Generation Lean Product Development. - Reinertsen, D. (2009)

Scrum.org. (2019). Evidence Based Management (EBM). [online] Available at: https://www.scrum.org/resources/evidence-based-management [Accessed 22 Mar. 2019].

Seddon, J. (2019). Failure demand | Vanguard. [online] Vanguard-method.net. Available at: https://vanguard-method.net/library/systems-principles/failure-demand/ [Accessed 22 Mar. 2019].

Snowden, D. (2019). A Leader's Framework for Decision Making. [online] Harvard Business Review. Available at: https://hbr.org/2007/11/a-leaders-framework-for-decision-making [Accessed 22 Mar. 2019].

Ulwick, T. (2019). The Core Tenets of Jobs-to-be-Done Theory. [online] Jobs-to-be-Done + Outcome-Driven Innovation. Available at: https://jobs-to-be-done.com/the-5-tenets-of-jobs-to-be-done-theory-ba58c3a093c1 [Accessed 28 Mar. 2019].

Vacanti, D. (2019, 2). LKCE12: Daniel Vacanti - Little's Flaw. [online] Vimeo. Available at: https://vimeo.com/52683659 [Accessed 24 Mar. 2019].

Wovchko, B. (2019). Starting Kanban. [online] Hugeio.com. Available at: https://www.hugeio.com/startingkanban/ [Accessed 22 Mar. 2019].