

Inhoud

Inleiding 7

DEEL 1: DE BASISBEGINSELEN VAN KANBAN 9

Hoofdstuk 1. Kanban in productie 11

- 1.1 De oorsprong van Kanban 11
- 1.2 Het geheim van Kanban: pull 13
- 1.3 Pull leidt tot flow 17
- 1.4 Waarom Kanban zo goed werkt 20
- 1.5 Stop met starten, start met afmaken 23

Hoofdstuk 2. Kanban in dienstverlening 25

- 2.1 Verschillen met productie 26
- 2.2 Overeenkomsten met productie 29
- 2.3 Principes en werkwijzen van Kanban 29

Hoofdstuk 3. Flow bereiken bij verstoringen 39

- 3.1 Waarom pull de sleutel is tot flow 39
- 3.2 Flow herkennen en beïnvloeden 42

Hoofdstuk 4. De flow verbeteren 45

- 4.1 Meten geeft inzicht 46
- 4.2 Elk Kanban-systeem is uniek 50

DEEL 2: KANBAN IN DE PRAKTIJK 53

Hoofdstuk 5. Persoonlijke Kanban 55

- 5.1 Waarom Persoonlijke Kanban? 55
- 5.2 Wat is Persoonlijke Kanban 57
- 5.3 Persoonlijke Kanban in de praktijk 63

Hoofdstuk 6. Kanban voor groepen of teams 83

- 6.1 Succes met Kanban voor teams 83
- 6.2 Starten met Team Kanban 86
- 6.3 Team Kanban in de praktijk 91

Hoofdstuk 7. Tevreden klanten met Kanban 109

- 7.1 Flow realiseren in de keten 110
- 7.2 Starten met Kanban in de keten 111
- 7.3 Klantfocus in de praktijk 112

BIJLAGEN 137

- Leeslijst 139
- Register 141

Inleiding

Ik hoorde voor het eerst over Kanban op een conferentie in 2008 en was meteen verkocht. Destijds dacht ik dat dat kwam door de onderliggende wiskundige schoonheid – een model dat je in staat stelt om beter te begrijpen hoe het werk wordt uitgevoerd. De mogelijkheid om te meten en op basis daarvan verbeteringen aan te brengen, had een grote aantrekkingskracht op mij.

Maar later ontdekte ik een tweede reden waarom ik zo enthousiast was over Kanban: niet alleen verschaft het een model van de werkelijkheid, maar Kanban heeft ook oog voor het menselijke aspect. Juist die combinatie spreekt mij zo aan. Dankzij Kanban krijg je inzicht in de huidige situatie, wat jou én anderen in staat stelt passende oplossingen te bedenken. Vooral die samenwerking is essentieel, omdat de omstandigheden waaronder het werk wordt gedaan zo veranderlijk en complex zijn, dat de kennis voor de beste oplossing zelden bij één persoon ligt.

Kanban gaat over het visualiseren van het werk en legt de nadruk op het afmaken van het werk door de voortgang erin te houden zonder de werkdruk te hoog op te voeren.

In dit boek is niet de theorie maar de praktijk leidend. Hoe hebben anderen Kanban toegepast in hun werk? Waar liepen zij tegenaan en voor welke oplossingen hebben zij gekozen? Hoe hebben zij hun aanpak steeds verbeterd? Juist die praktische insteek zal jou helpen om ook succesvol te zijn met Kanban. Want met Kanban krijg je meer gedaan in minder tijd, zodat je tijd overhoudt voor andere dingen.

DEEL 1

De basisbeginselen van Kanban

Hoofdstuk 1

Kanban in productie

1.1 De oorsprong van Kanban

Voor het ontstaan van Kanban moeten we terug naar het eind van de jaren 40 in Japan. Na de Tweede Wereldoorlog stond het land voor de enorme opgave van wederopbouw. Dat gold ook voor autofabrikant Toyota. Voor inspiratie keek Toyota naar de Amerikaanse auto-industrie en met name naar Ford. Om zo kostenefficiënt mogelijk te kunnen produceren maakte Ford slechts enkele modellen, allemaal zwart, maar wel in grote aantallen.

Daarvoor was het belangrijk om de machines continu te laten draaien – fabrieksarbeiders moesten er vooral voor zorgen dat de machines niet stil kwamen te staan. Om zo kostenefficiënt mogelijk te produceren, werden halffabricaten en onderdelen in grote aantallen gemaakt in specialistische werkplaatsen. Vervolgens werden deze opgeslagen in het magazijn om op een later moment verplaatst en verder verwerkt te worden in een andere werkplaats. Deze tijdelijke opslag betekent dat er tijd zit tussen het maken en het verwerken van de onderdelen. Het gevolg is dat eventuele productiefouten pas later worden opgemerkt.

Deze manier van produceren is bovendien kapitaalintensief – er is veel materiaal en veel opslagruimte nodig en er ontstaan grote tussenvoorraden. Dat kost veel geld en dat had Toyota niet. Wat Toyota nodig had, was een

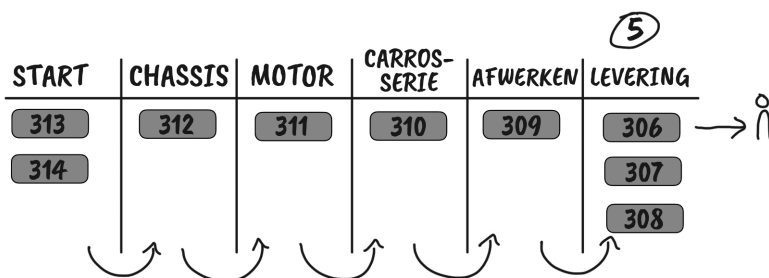
snelle cashflow, van order tot betaling. Anders dan Ford opereerde Toyota in een veel kleinere afzetmarkt, waarvoor een groter aantal modellen met lage aantallen vereist was. Toyota moest dus met andere ogen naar het productieproces kijken.

Een belangrijk principe dat Toyota overnam van Ford was de *continue materiaalstroom*. Hierin streeft men ernaar dat het materiaal, zoals halffabricaten, na productie direct verwerkt wordt zonder dat tussentijdse opslag nodig is. Zo is het materiaal steeds ‘in beweging’ en ontstaat er *flow* oftewel doorstroom. Toyota vulde dit aan met een zogenoemd *pull-systeem*. In de Verenigde Staten werd dit pull-systeem al toegepast in supermarkten om voorraden op het juiste moment weer aan te vullen. Denk maar aan het vullen van de schappen in de lokale supermarkt. Een leeg schap is een teken dat het product moet worden aangevuld. De hoeveelheid waarmee het schap wordt aangevuld is gelijk aan wat er die dag verkocht is. Simpel gezegd, in een pull-systeem bepaalt het aantal producten dat is afgenomen hoeveel producten worden aangevuld (en dus geproduceerd). Bij een lagere afname van het product, wordt de productie teruggedroefd, terwijl men bij een hogere afname de productie opvoert. De afname ‘trekt’ als het ware de productie aan, vandaar de term pull.

Hieruit volgen twee uitgangspunten die de basis vormden voor Kanban:

- >> Maak alleen een nieuw product als er ook een product is afgenomen (pull).
- >> Maak alleen een nieuw product als er een signaal is ontvangen dat door een afname is veroorzaakt.

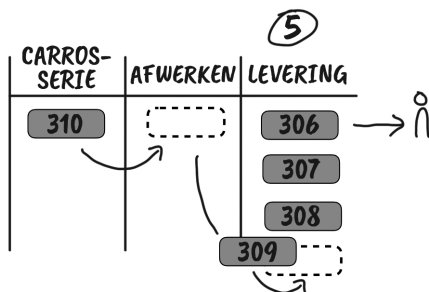
Het signaal om te gaan produceren is een *kanban*. Dat is een Japans woord dat letterlijk vertaald *visueel* (kan) *teken* (ban) betekent, een zichtbaar signaal dat bijvoorbeeld aangeeft dat het tijd is om de voorraad aan te vullen.



Afbeelding 2. Visualisatie van de werkstations in de vorm van een Kanban-bord, inclusief de levering aan klanten

werkstation. In dit geval is dat voor elk werkstation exact 1, omdat elk station telkens aan slechts één auto tegelijk werkt. Het cijfer 5 boven de laatste kolom geeft het maximumaantal auto's aan dat in voorraad gehouden mag worden – deze auto's zijn klaar voor levering. Dit maximum is afgestemd op de daadwerkelijke afname. Wordt het maximum bereikt, dan zal de productie stoppen, totdat de voorraad slinkt omdat er weer auto's zijn afgenomen.

Telkens wanneer het werk in een werkstation klaar is, gaat de auto in wording door naar het volgende werkstation. In afbeelding 3 zie je hoe dat in z'n werk gaat. Als het werkstation Afwerken klaar is met order 309,



Afbeelding 3. De werking van pull op basis van de afname van een product

Hoofdstuk 2

Kanban in dienstverlening

Een systeem dat is gebaseerd op pull en kanbans heeft Toyota geholpen om auto's te produceren met een continue materiaalstroom en minimale voorraden. Toyota opereerde in een markt met lage volumes en was toch in staat verschillende modellen aan te bieden. Er zijn meer situaties te bedenken waarin een laag volume en veel maatwerk voor een beperkt aantal klanten een rol spelen. Met name veel maatwerk maakt zo'n omgeving al snel complex. De dienstverlenende sector is hiervan een goed voorbeeld. Kan Kanban ook hier succesvol worden toegepast?

Tot dusverre hebben we het in de uitleg van Kanban vooral gehad over de productie van auto's – dat voorbeeld laat zich gemakkelijk vertalen naar andere tastbare producten. Maar hoe moeten concepten als pull, kanbans en doorstroom toegepast worden als de producten minder tastbaar zijn, zoals bij dienstverlening?

Daarvoor kijken we eerst naar de mate waarin orders variëren. Bij auto's kan een klant kiezen uit verschillende kleuren, velgen of opties, maar dat verandert de order en het productieproces niet wezenlijk. De werkstations worden in dezelfde volgorde afgewerkt, de werktijd zal voor elk werkstation nagenoeg hetzelfde zijn en de productie vraagt geen andere kennis of vaardigheden van werknemers. Kortom, ondanks de keuzemogelijkheden varieert het werk nauwelijks.

Variatie belemmert de doorstroom (one-piece flow). Daarom is het bij de productie extra belangrijk om zo veel mogelijk variatie weg te nemen. Maar hoe werkt dat als er door maatwerk sprake is van veel variatie, zoals bij dienstverlening?

Neem bijvoorbeeld het oversluiten van een hypotheek. Stel, je hebt een tijd geleden een huis gekocht en een hypotheek afgesloten met een hoge rente. Als de huidige rentestand lager is dan de hypotheekrente die je nu betaalt, dan kan het interessant zijn om je hypotheek over te sluiten. Een hypotheekadviseur bekijkt de rente en voorwaarden van andere geldschieters, de belastingregels en jouw persoonlijke financiële situatie om tot een advies te komen. Als jouw financiële situatie complex is, zal hij of zij wellicht de specifieke kennis van een collega nodig hebben om tot een goed advies te komen. Enkele factoren, zoals de belastingregels, zijn voor iedereen hetzelfde, maar veel factoren verschillen van persoon tot persoon, zoals het leenbedrag, het inkomen en de risicobereidheid. Hypotheekadvies is dus maatwerk en afgestemd op de specifieke situatie van de klant.

2.1 Verschillen met productie

Dienstverlening verschilt op drie punten wezenlijk van productie: er is sprake van meer variatie, het product is niet tastbaar en communicatie is een essentieel onderdeel van het werk.

Wat betreft de grotere variatie is het voorbeeld van het oversluiten van een hypotheek typerend voor dienstverlening. De omstandigheden van elke klant zijn uniek waardoor de afhandeling van elk oversluitverzoek anders is. Hoewel het verzoek in de basis vaak hetzelfde is, zorgt de persoonlijke financiële situatie van de klant voor veel variatie!

Soms is er specifieke kennis nodig die maar bij een beperkt aantal werknemers beschikbaar is. Juist die specialisten zijn vaak niet per direct beschikbaar waardoor wachttijden ontstaan en de doorstroom stukt.

Deze principes ondersteunen aanpassingen in complexe omgevingen als dienstverlening. ‘Start met nu’ zorgt er bijvoorbeeld voor dat je eerst probeert te begrijpen hoe het proces nu werkt. Het evolutionaire karakter benadrukt dat het met name bij dienstverlening onzeker is hoe aanpassingen uitpakken. Wat werkt behoud je, en wat niet werkt pas je aan. Bovendien is leiderschap nodig om aanpassingen voor te stellen en daadwerkelijk uit te voeren.

Overigens, houdt evolutionair hier niet in dat veranderingen langzaam gaan, zoals bij de ontwikkeling (evolutie) van organismen. Met evolutionair wordt hier bedoeld dat je alleen die veranderingen behoudt die ook daadwerkelijk een verbetering zijn en andere veranderingen verworpt.

Kanban: zes algemene werkwijzen

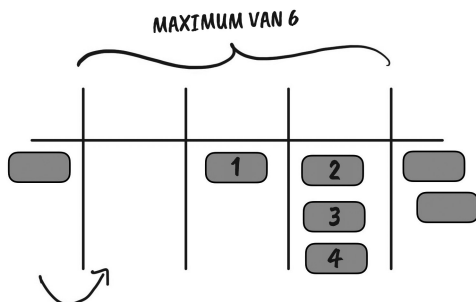
In de praktijk heeft de Kanban-methode zich ruimschoots bewezen door het structureel toepassen van zes algemene werkwijzen:

1. Visualiseren van het werk
2. Limiteren van onderhanden werk
3. Zorgen voor doorstroom
4. Expliciet maken van afspraken
5. Zorgen voor feedbackmechanismen
6. Samen verbeteren & evolueren door te experimenteren en gebruik te maken van data en wetenschappelijke methoden

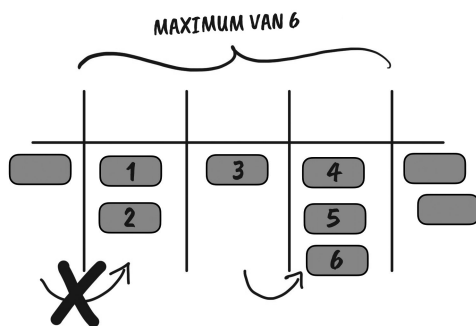
1. Visualiseren van het werk

Bij dienstverlening is het resultaat van het werk niet zo zicht- en tastbaar als bijvoorbeeld bij de productie van een auto. Daarom is een belangrijke eerste stap bij dienstverlening om het werk zichtbaar te maken.

In de praktijk werkt een drempelwaarde als volgt. Stel dat er een drempelwaarde is ingesteld van 6 voor de hoeveelheid onderhanden werk (afbeelding 10). Zolang deze waarde niet is bereikt, is het starten van nieuw werk toegestaan. Starten kan totdat dit maximum is bereikt. Vanaf dat moment stop je met het starten van nieuw werk (afbeelding 11). Pas als iets is afgemaakt (en afgenomen door de klant) kun je weer met iets nieuws starten.



Afbeelding 10. Illustratie van een Kanban-bord met een WiP-limiet voor de middelste drie kolommen – er is ruimte om nog 2x met werk te starten



Afbeelding 11. Er zijn zes werkeenheden onderhanden, het maximum is bereikt – er is nu dus geen ruimte om extra werk te starten

Hoofdstuk 4

De flow verbeteren

Kenmerkend voor dienstverlenend werk zijn de lage flow-efficiënties, variërend van 5 tot 20%. Een percentage van 10% betekent dat bij een doorlooptijd van 10 dagen, 1 dag wordt besteed aan het feitelijke werk zelf en 9 dagen opgaan aan wachttijd en vertragingen. De gewerkte tijd is dus slechts een fractie van de totale doorlooptijd.

Stel dat we een manier vinden om het werk in slechts de helft van de tijd te doen. De totale doorlooptijd wordt dan 9,5 dag, namelijk een halve dag plus de 9 dagen wachttijd. De totale doorlooptijd verandert dus nauwelijks – en de flow-efficiëntie evenmin.

Maar wat gebeurt er als we de wachttijd en vertragingen, waardoor het werk stilligt, weten te halveren? De totale doorlooptijd gaat nu van 10 naar 5,5 dagen: 1 dag voor het werk zelf en 4,5 dag dat er niet aan gewerkt wordt. Dat is een flinke verbetering van bijna 50% en bijna een verdubbeling van de flow-efficiëntie!

Bij lage flow-efficiënties levert sneller of harder werken dus relatief weinig op. Het grootste effect bereik je door vertragingen te verkorten of zelfs helemaal weg te nemen. Voorbeelden van vertragingen zijn het wachten op leveranciers, het wachten op beslissingen of het wachten op toestemmingen – bijvoorbeeld omdat de beslissers maar één keer per maand bij elkaar komen.

4.1 Meten geeft inzicht

De meeste verbeteringen richten zich op het reduceren of wegnemen van vertragingen. Het effect op de doorstroom – bepaald door de doorlooptijd, productiviteit, flow-efficiëntie en WiP – wordt zichtbaar door deze factoren te meten. Metingen helpen ook om achteraf te kunnen bepalen of een verbetering ook echt gunstig uitpakt. Zonder te meten, zonder feitelijke gegevens, weet je dat niet zeker.

Bij een lage flow-efficiëntie is het reduceren of wegnemen van vertragingen het allerbelangrijkste.

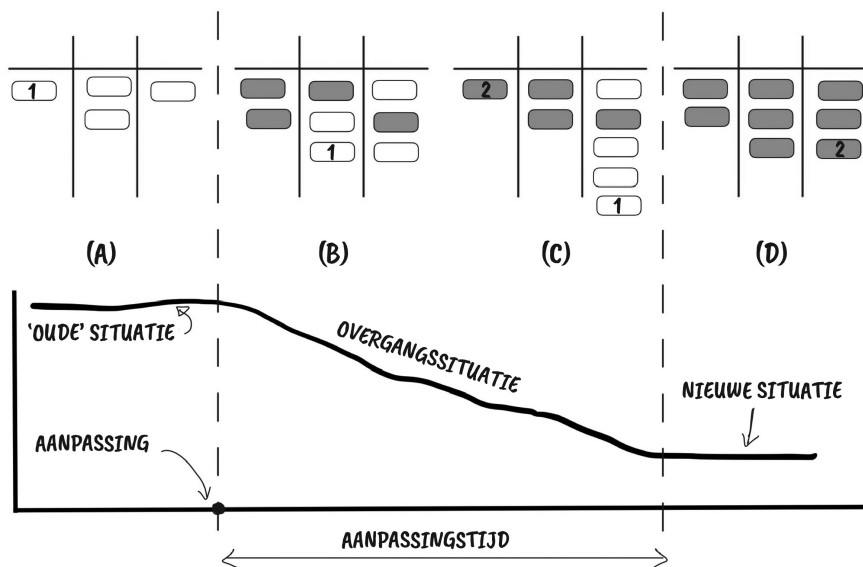
Flow in cijfers

Door te meten ontdek je wat er qua productiviteit mogelijk is. Als je weet hoelang het duurt om een verzoek af te handelen, weet je ook wat je aan klanten kunt beloven. En als je weet hoeveel producten je kunt leveren, weet je wat je aan werk kunt oppakken.

Maar wat moet je meten? Omdat verbeteringen zich vooral concentreren op de doorstroom is het in elk geval een goed idee om de hoofdkenmerken van doorstroom te meten (zie ook afbeelding 13):

- » **Doorlooptijd.** De totale tijd van start tot afname door de klant, inclusief wachttijd.
- » **Productiviteit.** Het aantal producten geleverd in een bepaalde periode.
- » **Work-in-Progress.** Hoeveelheid onderhanden werk.
- » **Flow-efficiëntie.** De tijd dat er daadwerkelijk is gewerkt, gedeeld door de doorlooptijd.

Voor klanten is juist de doorlooptijd belangrijk. Je bestelt je product natuurlijk liever bij een leverancier die je snel helpt, dan bij een leverancier waar je er lang op moet wachten.



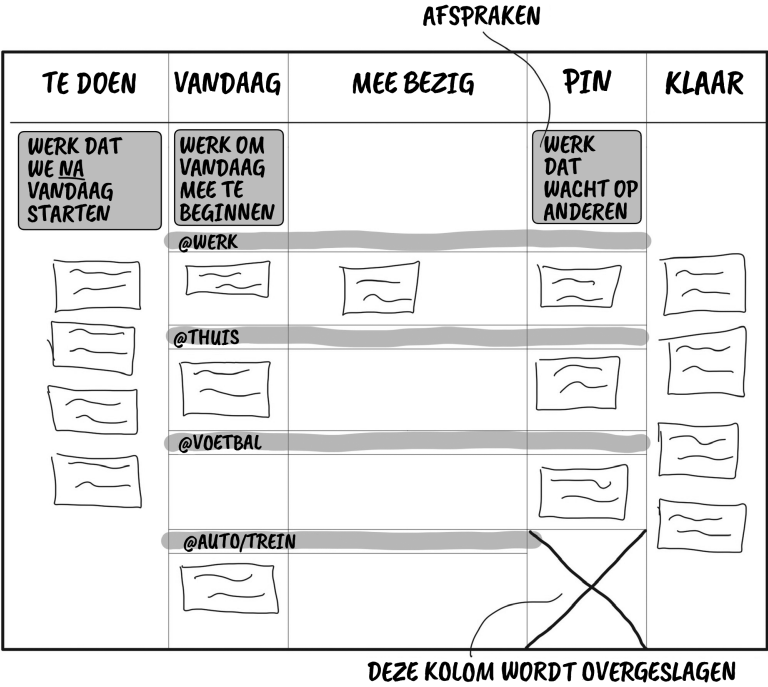
Afbeelding 14. Het effect van de verandering. Bord (A) laat de beginsituatie zien.

Na de verbetering is er een mix van werk van voor en na de verandering (B).

In (C) is al het werk van vóór de aanpassing klaar en (D) laat alleen werk zien dat na (C) op het bord is gekomen.

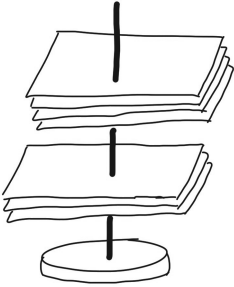
Net na de aanpassing (B) zal het bord een mix bevatten van oud (witte kaarten) en nieuw werk (grijze kaarten): de overgangssituatie. Na een tijdje is al het werk van vóór de aanpassing afgehandeld (C) en bevat het bord alleen nog nieuw werk dat gestart is na de aanpassing. Dat duurt precies zo lang als de tijdsduur voor kaart 1 om afgehandeld te worden: ongeveer de doorlooptijd.

De situatie bij C heeft nog nieuw te starten werk (kaart 2) vanuit B. We verwachten dat de doorlooptijden voor deze kaarten nog beïnvloed zijn door de oude situatie waarvan de witte kaarten (oud werk) een zichtbaar overblijfsel zijn. Als kaart 2 is afgehandeld (D) weten we zeker dat geen enkel kaartje nog is beïnvloed door het oude werk en dat de metingen



Afbeelding 21. Het persoonlijke Kanban-bord met twee extra kolommen

De naam ‘Pin’ is letterlijk afgeleid van een metalen pin (zie afbeelding 22) waar briefjes op gestoken worden om ze te bewaren tot later.



Afbeelding 22. Een stalen pin om briefjes op te steken die je later nog nodig hebt

Feedbackloops ondersteunen de besluitvorming

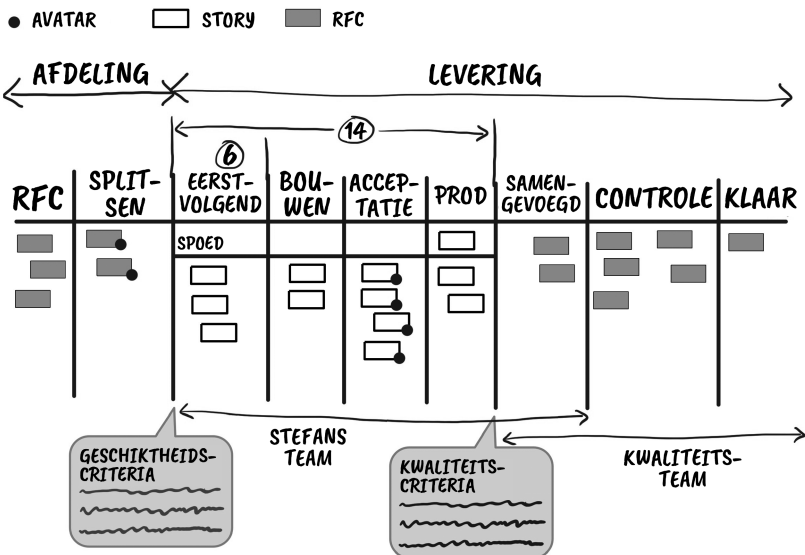
Het Kanban-bord en de bijbehorende afspraken vormen tezamen het Kanban-systeem, afgestemd op de besluitvorming die plaatsvindt tijdens de Kanban-ritmieken. Maar in de praktijk laat je het Kanban-systeem ook de bestaande overleggen ondersteunen. Om het systeem en het overleg goed op elkaar te laten aansluiten, stel je jezelf de volgende vragen:

- » Welke besluiten nemen we tijdens het overleg?
- » Welke informatie is nodig voor deze besluiten?
- » Wie moeten aanwezig zijn om deze beslissingen te kunnen nemen?
- » Hoe vaak is elk overleg nodig?
- » Wat is de agenda?
- » Hoe communiceren we de resultaten van het overleg?

Net als bij Persoonlijke Kanban maakt het team voor nieuw op te pakken werk onderscheid tussen twee categorieën: enkele werkitems die eerst opgepakt moeten worden en de rest. Hiervoor staan op het bord respectievelijk de kolommen Eerstvolgend en Opties (zie afbeelding 34). Voor nieuwe fouten bepaalt het team ook nog welke serviceklasse van toepassing is: Spoed of Gangbaar. Door aan een deel van de fouten de serviceklasse Gangbaar toe te kennen zullen die fouten niet meer storend werken.

Voor de Kanban-ritmieken besluit het team om het huidige voortgangs-overleg als basis te nemen en geen andere overleggen te introduceren. Wel gaat het team een wijziging doorvoeren. Namelijk dat ze tijdens het overleg het Kanban-bord gebruiken om te bekijken welk werk verder naar rechts – dus meer richting Gereed – gebracht gaat worden. Er wordt eerst alles aan gedaan om blokkades die de voortgang belemmeren op te heffen, voordat het team aan iets nieuws begint. Daardoor komt er meer focus op het afmaken van het werk.

Hiermee is het huidige voortgangsoverleg een combinatie geworden van de replenishment meeting en de Kanban-meeting.



Afbeelding 43. Een tweelagig Kanban-bord waarbij de eerste vier kolommen onder Levering overeenkomen met het Team Kanban-bord uit het vorige hoofdstuk

Effectieve metingen bij een hiërarchie

Het visualiseren van de werkstroom alleen verklaart nog niet waarom klanten vinden dat het te lang duurt. Om hier meer inzicht in te krijgen moet je data te verzamelen over de doorlooptijd van de RFC van begin tot eind. Door te meten wil het team in de komende weken de volgende data verzamelen:

- » De hoeveelheid onderhanden werk per dag
- » De hoeveelheid onderhanden werk voor de fasen Afdeling en Levering
- » De doorlooptijd voor de fase Levering

Deze meetgegevens zijn voor de RFC's en niet te vergelijken met de metingen die het team in het vorige hoofdstuk heeft gedaan voor de story's.

Voorspelbaar zijn bij variatie

Wie herkent het niet? Een klant doet een verzoek, dat je accepteert, en vraagt vervolgens wanneer het klaar is. Een logische vraag want de klant wil natuurlijk weten wanneer hij of zij het resultaat kan verwachten. Het gevaar is dat wanneer je – als uitvoerder van het verzoek – een datum noemt, je je verplicht om ook op of voor die datum te leveren.

Dat hoeft op zich geen probleem te zijn, maar het wordt lastiger wanneer je dat voor elk verzoek doet, waardoor al het onderhanden werk op een bepaalde datum klaar moet zijn. Dan is er weinig tot geen ruimte en flexibiliteit om met variaties en tegenslagen om te gaan. Tegelijkertijd wil je ook voorspelbaar zijn richting de klant. Je kunt overwerken om een leverdatum te halen, maar dat wil je natuurlijk liever vermijden.

Doel: weten wat we kunnen toezeggen

Ondanks de recente verbeteringen vinden de klanten van het team nog steeds dat het te lang duurt. Ze willen graag weten waar ze aan toe zijn en hebben behoefte aan duidelijkheid over de levertijd. Af en toe besluit een klant om een verzoek in te trekken – het heeft dan te lang geduurd en intussen is er een RFC gekomen die voor de klant urgenter is en voorrang krijgt. Voor het team is dit storend en frustrerend omdat zij zodoende niet kunnen afmaken waar ze aan begonnen zijn! Ze willen weten op welke manier ze de geschapen verwachtingen ook daadwerkelijk kunnen waarmaken. Wat kunnen ze beloven en welke afspraken zijn er nodig?

Verwachtingen scheppen die je kunt waarmaken

Het team moet in feite twee vragen beantwoorden:

1. Wat is een goed moment om de opdracht aan te nemen?

Op welk moment is er voldoende duidelijk om de opdracht aan te nemen? Hebben we de capaciteit om met de opdracht te starten? Kunnen we op tijd afspraken met leveranciers maken? Is wat de klant vraagt wel realistisch?