



Triptych of Pompeius Occo
and his Wife Gerbrich Claesdr.
Jacob Cornelisz. van
Oostanen - KMSKA

IIIF image endpoint

Stand van zaken 18/06/21



We hebben gewerkt aan

1. De testomgeving (infra)
2. De image processing workflow (dev)
3. Analyse afmetingen ikv herschalen (analyse)

Testomgeving

Installatie en configuratie is quasi voltooid:

- VM in meemoo datacenter in Oostkamp
- Debian 10 (Buster)
- Licensed Kakadu
- Python3
- 10 + 100GB diskpace (flexibel uitbreidbaar)

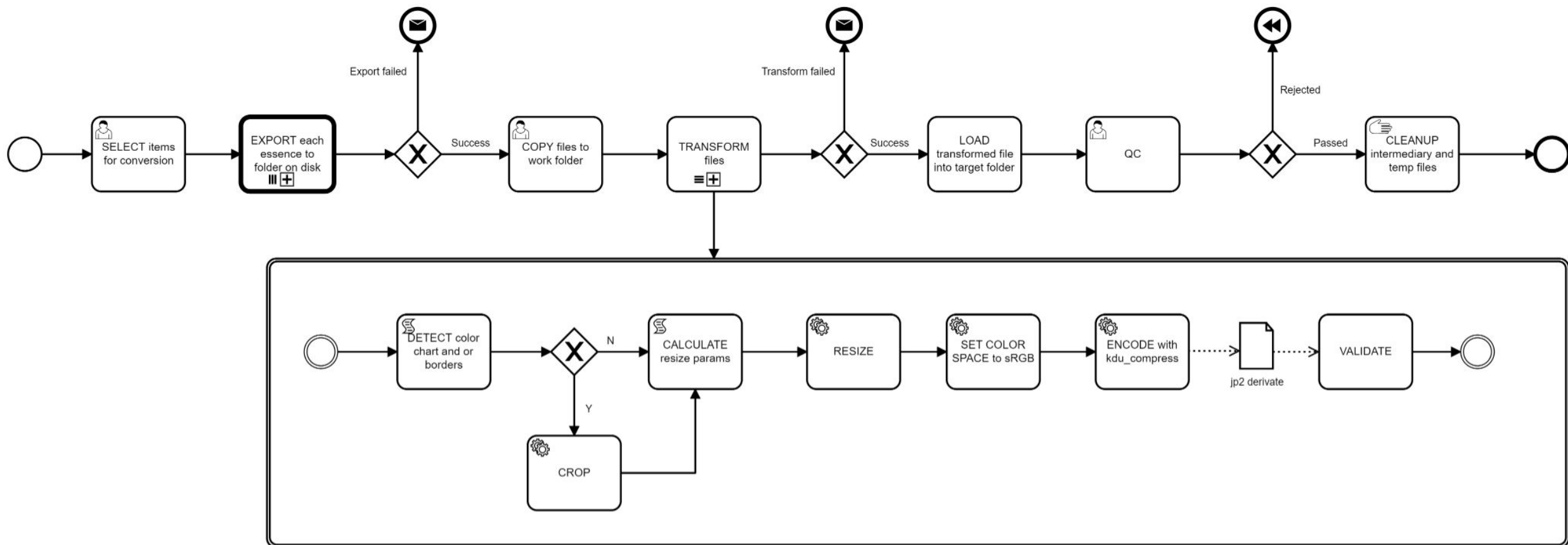


Testomgeving

Nog work in progress

- configuratie omgeving in Puppet
- installatie image server
- publieke toegang via images-tst.meemoo.be

Image processing workflow



Zie <https://github.com/viaacode/iiif-deliverables/blob/main/d4-appB-workflow.adoc>

Image processing workflow

Proof-of-concept staat op testomgeving.

- volledige flow minus herschalen:
 - export uit archief
 - verwijderen kleurenkaart (indien aanwezig)
 - omzetten naar sRGB (indien ander profiel)
 - compress en encode als jp2 met Kakadu
 - opslaan in lokatie voor image server
- nu nog manuele trigger, in productie zal dit met watch folder werken
- herschalen nog even on hold in afwachting van beslissing
- nog geen validatie van afgeleid bestand



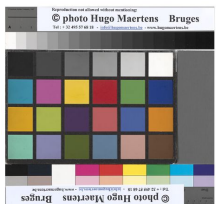
Image processing workflow

Issues met de crop-tool

- crop blijkt geen crop maar een overlay
- verlies van kleur bij omzetting naar sRGB
- metadata gaat verloren

Todo's

- herschalen
- volledige validatie eindproduct:
 - 300dpi
 - metadata in bestand
 - sRGB icc
 - valide jp2



Triptych of Pompeius Occo and his Wife Gerbrich Claesdr.
Jacob Cornelisz. van Oostanen - KMSKA

<https://arthub.vlaamsekunstcollectie.be/nl/catalog/kmska:cxas9b>



Workflow

Git repository in meemoo Github:

<https://github.com/viaacode/iiif-image-processing>



Analyse herschalen

Waarom herschalen?

Archiefbestanden kunnen groot zijn. Voor de representatie zijn vaak kleinere afmetingen afdoende.

Om zowel diskpace als bandbreedte en laadtijd (performantiecriteriën) te beperken worden de afgeleide bestanden herschaald tot een afmeting die zowel voldoende detail kan bevatten als rekening houdt met voormelde performantiecriteriën.

Analyse herschalen

Dynamisch vs vast herschalen

Bij vaste herschaling wordt de langste of kortste zijde herleid tot een vooraf bepaalde waarde. Waarbij het resulterend bestand voldoende detail bevat en tegelijk de performantie waarborgt.

In een eerdere fase van het IIF-project werd dit bepaald als (max.) 5000 pixels voor de kortste zijde.

Voor bestanden die merkelijk groter zijn dan het gemiddelde betekent dit echter een veel ingrijpender herschaling met kwaliteitsverlies.

Analyse herschalen

Dynamisch vs vast herschalen

Bij vaste herschaling wordt de langste of kortste zijde herleid tot een vooraf bepaalde waarde. Hierbij wordt echter louter gekeken naar de resolutie van het digitale bestand uitgedrukt in pixels en niet naar de fysieke afmeting van een werk.

In D4 van dit project is daarom besloten om na te gaan of er een correlatie bestaat tussen de fysieke grootte van de werken en de resolutie in pixels van de beelden. Een of aan de hand van deze correlatie een dynamische herschaling mogelijk is die een betere balans tussen performantie en kwaliteit waarborgt.

Analyse herschalen

Eerste test met selectie van 77 testbestanden (vkc fase 1)

Resultaat:

- gemiddelde breedte = 90 cm
- gemiddeld aantal pixels = 6169px
- geen correlatie tussen fysieke breedte en breedte in pixels

maar kleine set en slechte p-value

```
Sample: 77
Mean width in cm: 90
Mean width in px: 6169
cor (pearson): 0.09
```

```
data: cm and px
t = 0.80326, df = 75, p-value = 0.4244
alternative hypothesis: true correlation is not equal
to 0
95 percent confidence interval:
-0.1344031 0.3099238
sample estimates:
cor
0.09235595
```


Analyse herschalen

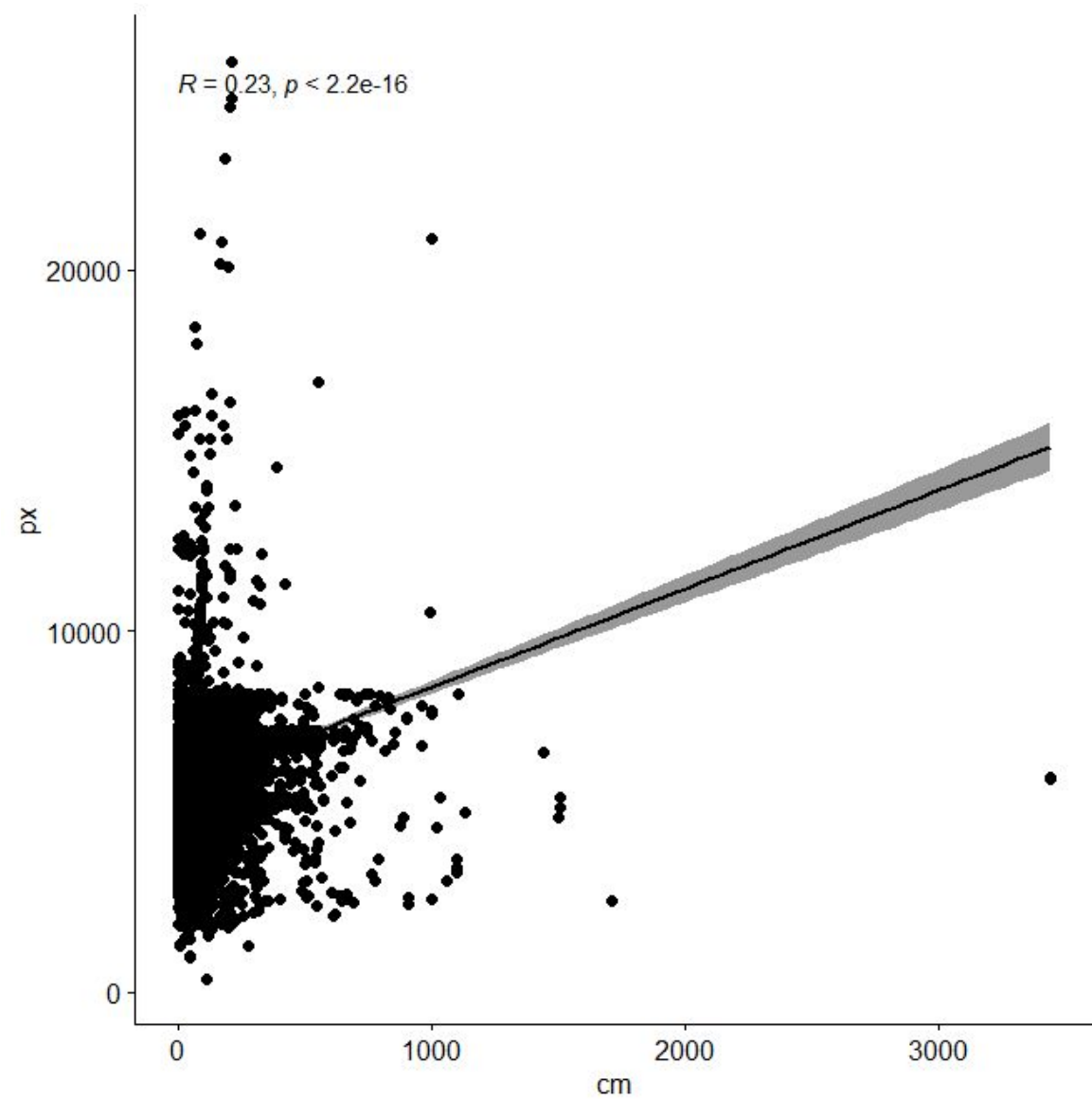
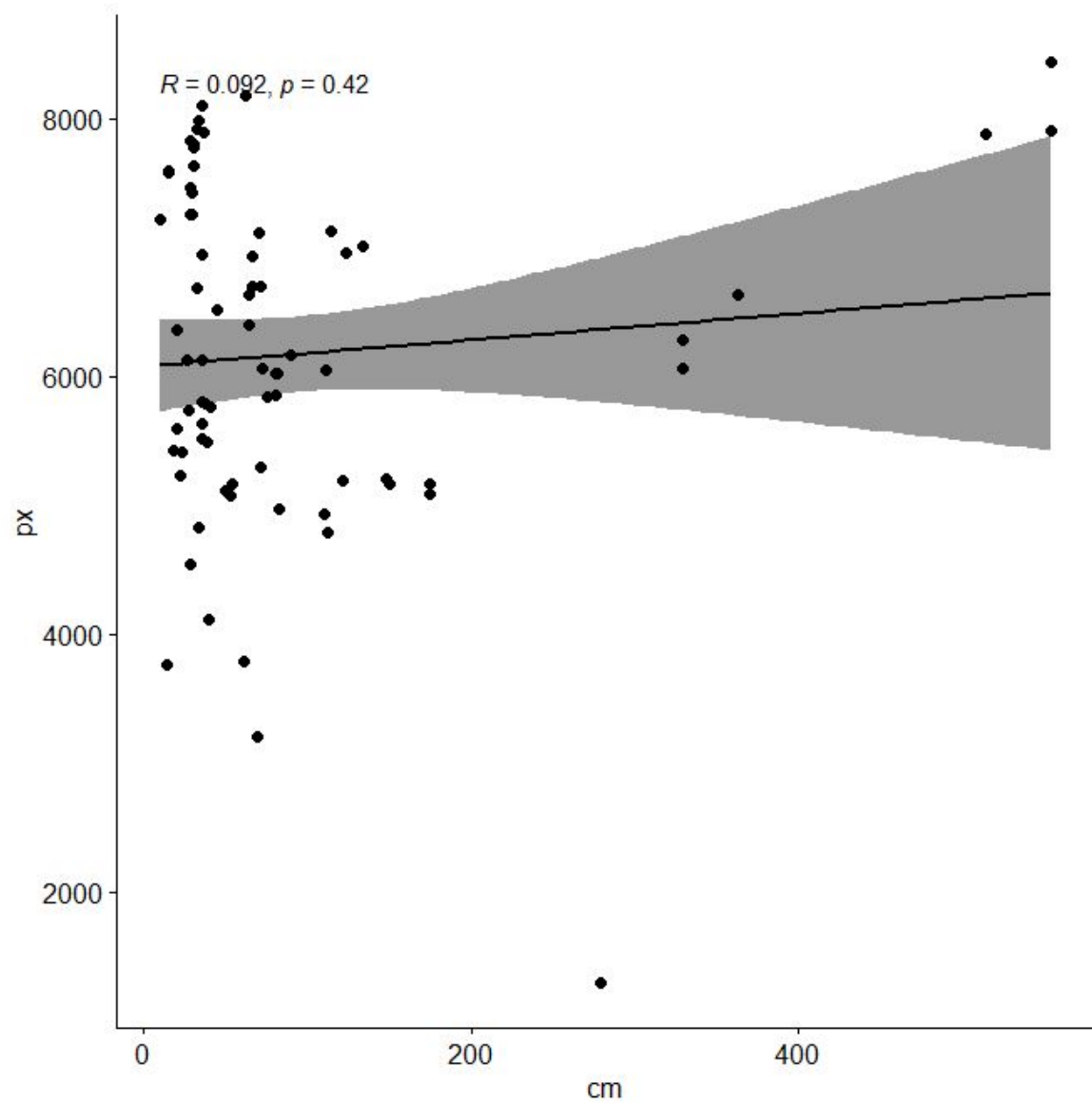
Tweede test met alle beelden uit doorsnede VKC/AIF

Resultaat:

- gemiddelde breedte = 98 cm
- gemiddeld aantal pixels = 6024px
- weinig tot geen correlatie tussen fysieke breedte en breedte in pixels

```
Sample: 12639
Mean width in cm: 98
Mean width in px: 6024
cor (pearson): 0.09
```

```
data: cm and px
t = 26.744, df = 12639, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal
to 0
95 percent confidence interval:
 0.2148652 0.2478637
sample estimates:
      cor
0.231431
```





Analyse herschalen

Uitkomst

Er is geen waarneembare correlatie tussen de fysieke afmeting van beelden en de resolutie in pixels.

De resolutie clustert rond een mediaan van 6050 pixels, met het gros tussen 4000 en 8000 pixels, afnemend in aantal boven 10.000px tot enkele zeldzame pieken boven 20.000px (max. 25k).

Analyse herschalen

Voorstel getrappt herschalen

- Geen min.
- Afbeeldingen tot 5.000px breedte ongemoeid laten
- Van 5.001-10.000px 50% herschalen
- Max. breedte 10.000px

Voorbeelden

4.325px => 4.325px

7.600px => 5.000px + 1.300px = 6.300px

11.000px => 5.000px + 3.000px = 8.000px

18.000px => 5.000px + 9.000px = 14.000px => 10.000px

**Zijn er nog
vragen?**