

KANTTEKENINGEN BIJ EEN DATABANK

De afgelopen 2 jaar besteedde BAM veel aandacht aan het in kaart brengen van Mediakunst, Beeldende Kunst en Audiovisuele Kunst in Vlaanderen. De gegevens werden door medewerkers van BAM met zorg verzameld en ingevoerd in een databank die via <http://bamart.be> publiek te raadplegen is. Met name de sectie Mediakunst leverde een rijke hoeveelheid data op, en al bladerend door de website beginnen we ons af te vragen hoe die gegevens beter te begrijpen.

Het wordt al snel duidelijk dat er heel veel informatie over het veld is bijeengebracht, maar ook dat er soms gaten vallen. Creaties zijn bijvoorbeeld maar sporadisch ingevoerd en ook missen we enkele organisaties en personen die zeker van belang zijn voor het veld.



In de hieropvolgende pagina's doen we verslag van een serie experimenten die zowel een beeld geven van de technologie van de databank, als van de gegevens die er in zijn opgeslagen. Ze tonen dus niet 'hoe Mediakunst is', maar laten zien hoe je de BAM-databank op verschillende manieren kunt bekijken.

// Met onder andere:
Nicolas: programmeur en data-activist (OSP)
Femke: ontwerper, kunstenaar en rapporteur (OSP)
Pierre: ontwerper, typograaf, cartograaf (OSP)
Harrison: ontwerper, typograaf (OSP)
Michael: computer wetenschapper gespecialiseerd in interactieve lees- en redactiesystemen
Thomas: ontwerper, betrokken bij de denktank over de toekomst van de BAM-databank
Didier: programmeur, ontwikkelt de BAM databank en website (Système D)

.....

KAART 1: FLOWCHARTS

Het is verleidelijk om direkt met grafieken, kaarten en netwerkvisualisaties te beginnen, maar ons eerste werk bestaat er uit inzicht te krijgen in de manier waarop de gegevens in de BAM-databank zijn opgeslagen. Dat is niet

alleen uit technisch oogpunt noodzakelijk, maar het geeft ons ook een beter beeld van de 'data-praktijk', de manier waarop de gegevens in werkelijkheid gehanteerd worden.



didier.sql
10/10/10

Didier heeft ons een SQL-Dump gestuurd, een letterlijke kopie van de BAM-gegevens die we ieder op onze eigen computers in een locale MYSQL-databank importeren. Nu kunnen we zelf complexe zoekopdrachten uitvoeren en ook de verschillende scripts die we programmeren, direct toegang geven tot de benodigde gegevens.

// Overzichtstabel in PHP-MyAdmin



Een databank bestaat uit verschillende tabellen met gegevens die je met elkaar kunt verbinden. Zo hoeven bijvoorbeeld adresgegevens maar 1 keer te worden ingevoerd, en kun je toch terugvinden in welke plaats het keyword geluid het meest wordt gekozen, maar daarover meer in KAART 4.

Maar hoe is deze databank eigenlijk precies gestructureerd? Nicolas print alle verschillende tabellen uit om te zien hoe ze zijn georganiseerd.



Welke gegevens verwijzen naar elkaar, welke tabellen zijn eigenlijk in gebruik?

In de BAM-databank zijn directe relaties tussen tabellen, zelf opgeslagen in tabellen met het voorvoegsel relations_:



Het valt op dat sommige tabellen die in eerste instantie belangrijk lijken, in feite leeg zijn: [relations_persons_projects](#) heeft bijvoorbeeld 0 rows, en [ag_item_categories](#), waarvan we niet meteen weten waar ze voor dient, bevat juist heel veel data.

[Michael](#) schrijft een klein script in Python, dat een lijst genereert van alle tabellen in de databank, georganiseerd volgens het aantal rows dat in gebruik is.

(...)

```
c2.execute("select
count(*) from
"+tablename)
count =
int(c2.fetchone()[0])
c2.close()

# print
tablename, (50 -
len(tablename)) * " ",
count

tablerows[tablename] =
count
```

(...)

Het volgende overzicht is het resultaat:

```
relations_projects_
works 0
mailing_templates 0
mailing_sending 0
relations_projects_
organisations 0
relations_projects_pa
ges 0
relations_projects_
keywords 0
relations_persons_
projects 0
relations_events_
events 0
relations_persons_
dossiers 1
relations_projects_
files 1
relations_dossiers_
works 1
banners 1
events_reeksen 2
relations_events_
works 2
```

(...)

```
relations_persons_
categories 1096
relations_
organisations_
categories 1156
mailing_cdata 1611
relations_
newscategories_
pages 1646
pages 1793
subsidies 1828
relations_persons_
works 2044
works 2080
relations_works_
images 2141
images 2284
ag_items 4370
ag_item_categories_
5027
mailing_users 8836
mailing_confirmation
9326
```

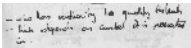
Een paar dagen later e-mailt [Didier](#) een schematische voorstelling van de verschillende tabellen en hun verbindingen.



De *flowchart* diende oorspronkelijk als ontwerp voor de databank, en helpt nu de opzet van het systeem te communiceren.

We transponeren de gegevens over het feitelijk gebruik van de databank naar de *flowchart* om te begrijpen hoe deze functioneert. Nu is de databank pas echt tot leven gekomen.

// aantekening Thomas
werksessie 2:
- who has authority to qualify the link
- link depends on context it is presented in



KAART 2: TAGCLOUDS

Een *tagcloud* is eigenlijk een lijst waarvan elk element automatisch naar waarde wordt geschaald; in één oogopslag wordt duidelijk welke elementen er toe doen en welke weinig prioriteit hebben. Het is een eenvoudige maar doeltreffende manier om gegevens uit een databank te visualiseren.



Als *testquery* selecteert Michaël alle voornamen van personen in de databank en organiseert ze op populariteit:

```
SELECT first_name,  
count( * )  
FROM persons  
GROUP BY first_name  
ORDER BY 'count( * )'  
DESC
```

first_name	count(*)
Jan	19
Eric	13
Dolores	12
Karen	8
Dick	8
Eric	8
Philip	8
Wesley	7
Bill	7
Harv	7
Hans	7
Lucas	7
Tim	6
Kris	6
Earl	6
Lucas	1

De populairste naam met voorsprong is Jan (19). Terwijl in de BAM-databank géén expliciete informatie is opgeslagen over de *gender* van de mediakunstenaar, valt uit deze selectie op te maken dat de eerste 8 resultaten mannennamen zijn; de eerste

vrouwennaam op de lijst is [Els](#) (7). Of hebben vrouwelijke mediakunstenaars meer verschillende voornamen dan mannen?

We kunnen geen manier verzinnen om mannennamen automatisch van vrouwennamen te onderscheiden, dus dat doen we handmatig. Ook dan is dat nog best lastig: is Ibe een mannen- of een vrouwennaam? en Goele? En hoe filter je een vrouw die zich Rob noemt er uit?

.....

KAART 3: RELATIES

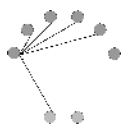
Om de publicatie te structureren, koos [Liesbeth](#) 14 redactionele keywords, die 2 aan 2 als hoofdstuktitel dienen:

[alledaags](#)
[alomtegenwoordig](#)
[curating](#)
[databases](#)
[delen](#)
[geheugen](#)
[identiteit](#)
[immersie](#)
[interactie](#)
[narrativiteit](#)
[participatie](#)
[presentatie](#)
[profielen](#)
[samenwerking](#)

Hoe verhouden die termen zich eigenlijk tot de keywords in de database? Hoe populair zijn ze? En bij wie?

[Nicolas](#) begint met te bepalen wat de populairste 14 termen zijn. Daarvoor moeten we op verschillende plaatsen kijken; verschillende tabellen maken gebruik van de tabel keywords.

// Flowchart van alle tabellen die met de tabel keywords zijn verbonden



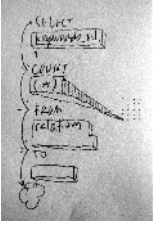
We besluiten dat de keywords gerelateerd aan organisaties, plus die aan personen er in eerste instantie het meest toe doen:

```
SELECT keyword_nl,
COUNT( * )
FROM relations_persons_
keywords, keywords
WHERE keywords.id =
keyword_id
GROUP BY person_id
ORDER BY 'COUNT( * )'
DESC
```

en

```
SELECT keyword_nl,
COUNT( * )
FROM
relations_organisations_
keywords, keywords
WHERE keywords.id =
keyword_id
GROUP BY organisation_id
ORDER BY 'COUNT( * )'
```

DESC



Er zijn duidelijk verschillen tussen de keywords die favoriet zijn bij personen, en die bij organisaties:

// KEYWORDS VAN PERSONEN

```
activisme 50
body 39
antropologie 28
antropologie 19
activisme 18
artificiële
intelligentie 17
commons 17
apparaat 16
cartografie / mapping
14
artificiële
intelligentie 13
body 13
artificiële
intelligentie 11
antropologie 10
archief 8
```

// KEYWORDS VAN ORGANISATIES

```
antropologie 42
activisme 39
bewaking 35
bewaking 34
archief 31
activisme 17
activisme 16
virtuele gemeenschap
(communitiy) 14
cinema 12
archief 11
apparaat 10
activisme 10
body 8
cinema 6
```

Maar in beide gevallen staat activisme bovenaan. Kan het waar zijn dat Vlaamse mediakunstenaars activisme verkiezen boven artificiële intelligentie? En waarom staat het eigenlijk 4 keer op de lijst?

We beginnen opnieuw:

```
SELECT keyword_nl,
COUNT( * ) as cnt
FROM relations_persons_
keywords, keywords
WHERE keywords.id =
relations_persons_
keywords.keyword_id
GROUP BY keyword_id
ORDER by cnt
DESC
```

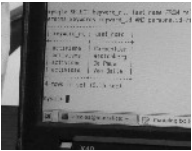
En dan ziet het resultaat er al heel wat overtuigender uit:

```
kunst en wetenschap
10
ruimte 9
geluid 9
media en communicatie
8
cultuurtheorie 8
filosofie 7
globalisering 7
geheugen 7
netwerkcultuur 7
cartografie / mapping
7
politiek 7
```

virtuele gemeenschap
(community) 7
[interactie](#) 7
software 7

Nadat we er de keywords die verbonden zijn met organisaties, files, dossiers en pagina's bij hebben opgeteld, kunnen we eindelijk aan de vergelijking met redactionele sleutelwoorden beginnen. [KAART 3.1]

Wie hebben eigenlijk wél [activisme](#) als keyword gekozen?



De bescheiden afbeelding maakt gebruik van *connectors* in **Inkscape** (een *open source* programma waarmee je vectors kunt tekenen in SVG formaat). Met behulp van een PHP-script haalt [Nicolas](#) data op uit de databank, creëert hij een vector bestand, plaatst de resultaten op de pagina en verbindt gerelateerde gegevens met behulp van een elastische lijn, de connector.

```
<?php
include('ibk_functions.php');
include("dbinfo.inc.php");
mysql_connect($localhost,$username,$password);
@mysql_select_db($database) or die("Unable to select database");
//x=largeur
//y=hauteur
$fn='dyn.svg';
$font=14;
$font_persons=10;
$char_width=6.5;

(...)

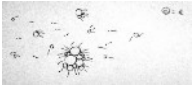
foreach($ar_connectors
as $k => $v){
    $str_out.=svg_connecto
r('c',$k,'left_'.
    $v['org'],
    $ar_orgs[$v['org']]
    ['x'],
    $ar_orgs[$v['org']]
    ['y'],'right_'.
    $v['person'],
    $ar_persons[$v['person']]
    ['x'],
    $ar_persons[$v['person']]
    ['y']);
}
```

Datzelfde script, maar nu met een andere zoekopdracht, gebruiken we om meer voor de hand liggende relaties te visualiseren: die tussen organisaties en personen bijvoorbeeld, of die tussen (...) en (...) [kaart 3.1 + 3.2]

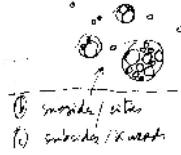
.....

KAART 4: STEDEN

Hoe ziet Vlaanderen er uit door de ogen van de BAM-databank?



Kunnen we iets laten zien van de verspreiding of clustering van organisaties? En hoe zit het met de verdeling van subsidies?



(...)