

Why, what, when and how?

Curatorial decisions and the
importance of their documentation

Sabine Schostag
Web curator

IIPC GA WAC 2019

Keywords: web curation, collection building, documentation

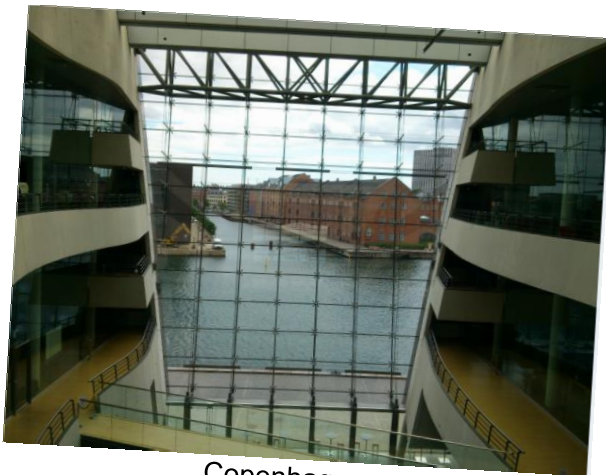


**DET KGL.
BIBLIOTEK**

Royal Danish Library

Netarkivet – The Danish national Web archive

- The Danish Web Sphere is part of the Danish Legal deposit law since 1. july 2005
- The Royal Danish Library (Det Kgl. Bibliotek) runs Netarkivet



Copenhagen

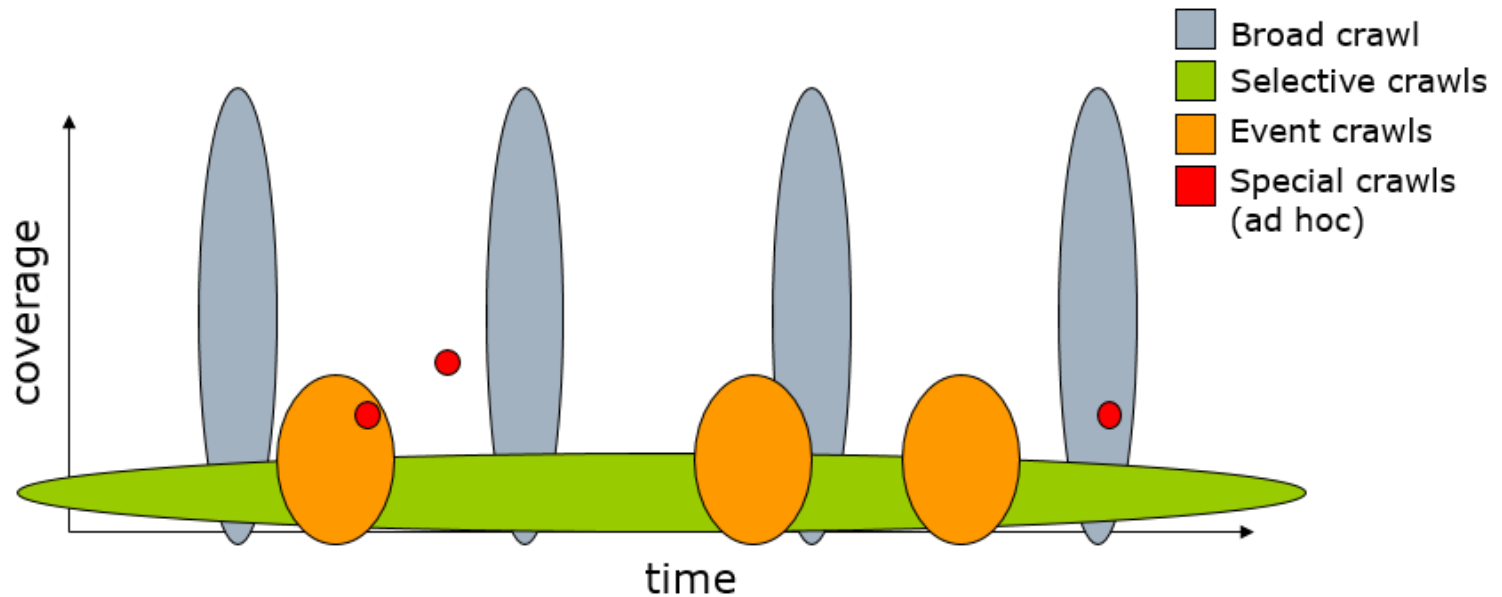


Aarhus



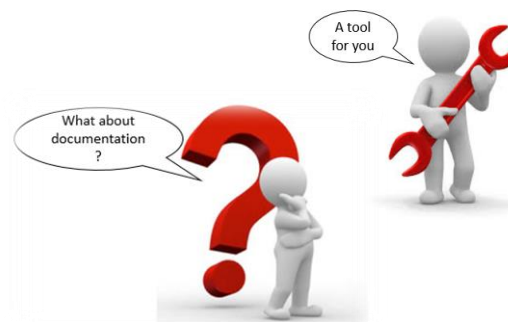
The Beginning – A team, a tool and a strategy

- 6 curators
- An in-house developed curator tool – Netarchive Suite (NAS)
- Collection strategies



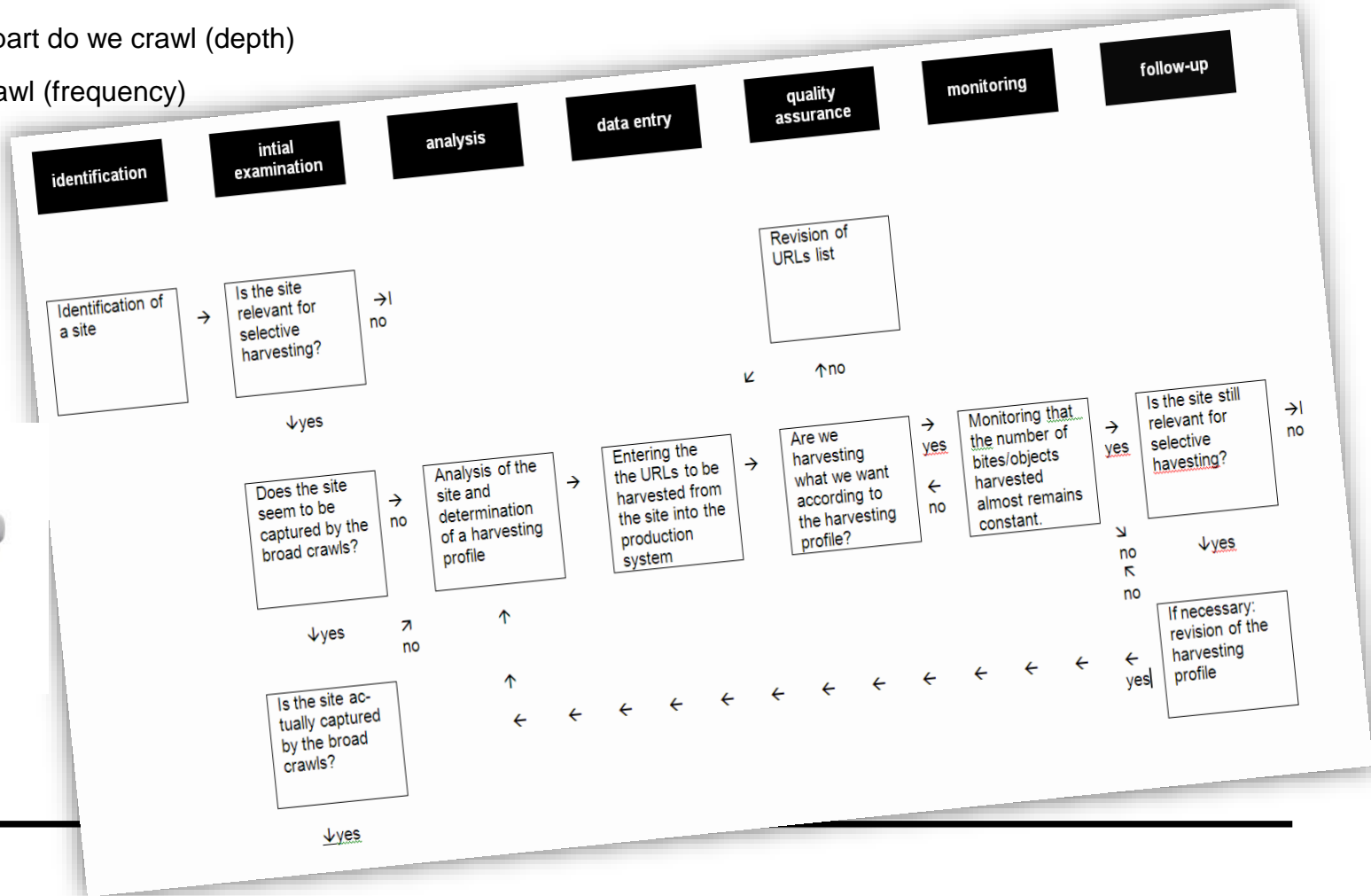
Decisions – and how to keep trace of them

- NAS – not enough space and functionalities for the curators documentation
 - A special need for the selective crawls
- Requirements for a tool:
 - Easy to access for all involved persons
 - Easy to edit, make changes, ad content
 - Offer the options to document
 - Selections and deselections with reasons for decisions
 - Start and end of crawl periods
 - QA observations and follow-ups



The selective workflow— How to keep trace

- Why is a domain in the workflow?
- How and what part do we crawl (depth)
- When do we crawl (frequency)



First documentation Tool – Word documents, folders, path finder



domænenavn

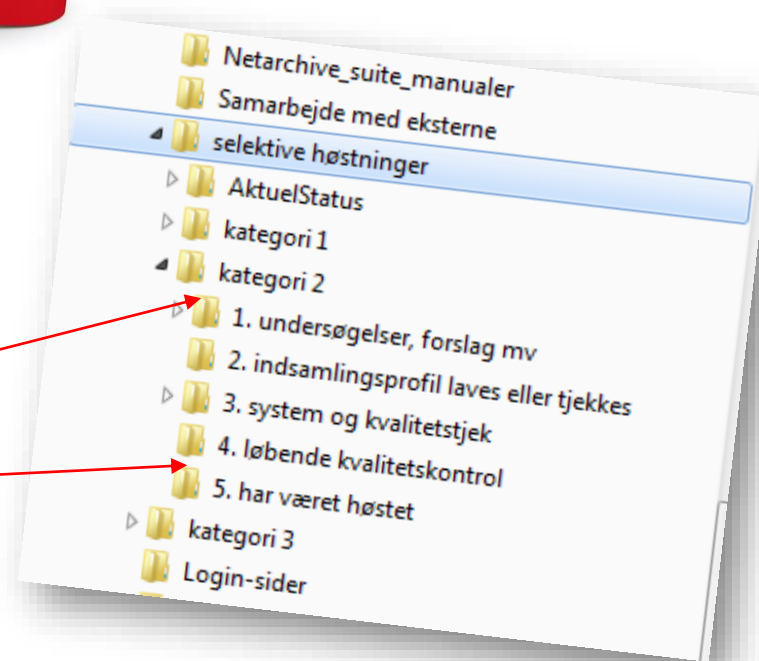
Undersøgelse
Initialer dd-md-åååå:

Login-indhold
Adgang:
Initialer dd-md-åååå:

Initialer dd-md-åååå	URL	Frekvens	Dybde

Kommentarer (fx 'er lagt i systemet', 'høstes med login' etc.)
Initialer dd-md-åååå

Eksisterende [høstingsdefinitioner](#).
Klik her og se hvilke [høstingsdefinitioner](#) der findes i systemet



Word documents

Path Finder

Improving the documentation– introduction of a wiki



Arbejderen.dk

Indholdsfortegnelse (ajud)

- 1. Arbejderen.dk (kategorier)
- 1.1. Indsamlingen
- 1.2. Login-adgang
- 1.3. Indsamlingen
- 1.4. Kommentarer (fx 'er lagt i systemet', 'høstes med login' etc.)

arbejderen.dk (kategori 1)

Undersøgelser

KON 2008-01-21: Side undersøgelse
JAN 2010-04-22: Ingen ændringer
JUL 2012-07-19: Viglig konfigurationsnyhedsnote på venstremenu. Løb blev ryddet op (jul 2012) - kan sikkert tilskrives sommerperioden. Tidligere målinger har været aktive. Særlig træning har forsvundet

Login-adgang

JUL 2008-11-10: Der er ikke login-vedhold

Indsamlingsprofil	Urt	Frekvens	Dybde
Indsaml 3333-mm-04	http://www.arbejderen.dk/	3 x dagligt 1 x dagligt	5 (forude) 2 (forude + 2)
(for 2008-01-21)	http://www.arbejderen.dk/	6 x dagligt 2 (forude)	2 (forude + 2)
KON 2008-01-21	http://www.arbejderen.dk/	1 x dagligt 6 (forude)	6 (forude)
jan 2012-10-18	http://www.arbejderen.dk/	2 x dagligt 1 x dagligt	2 (forude + 2)
otto	3	4	

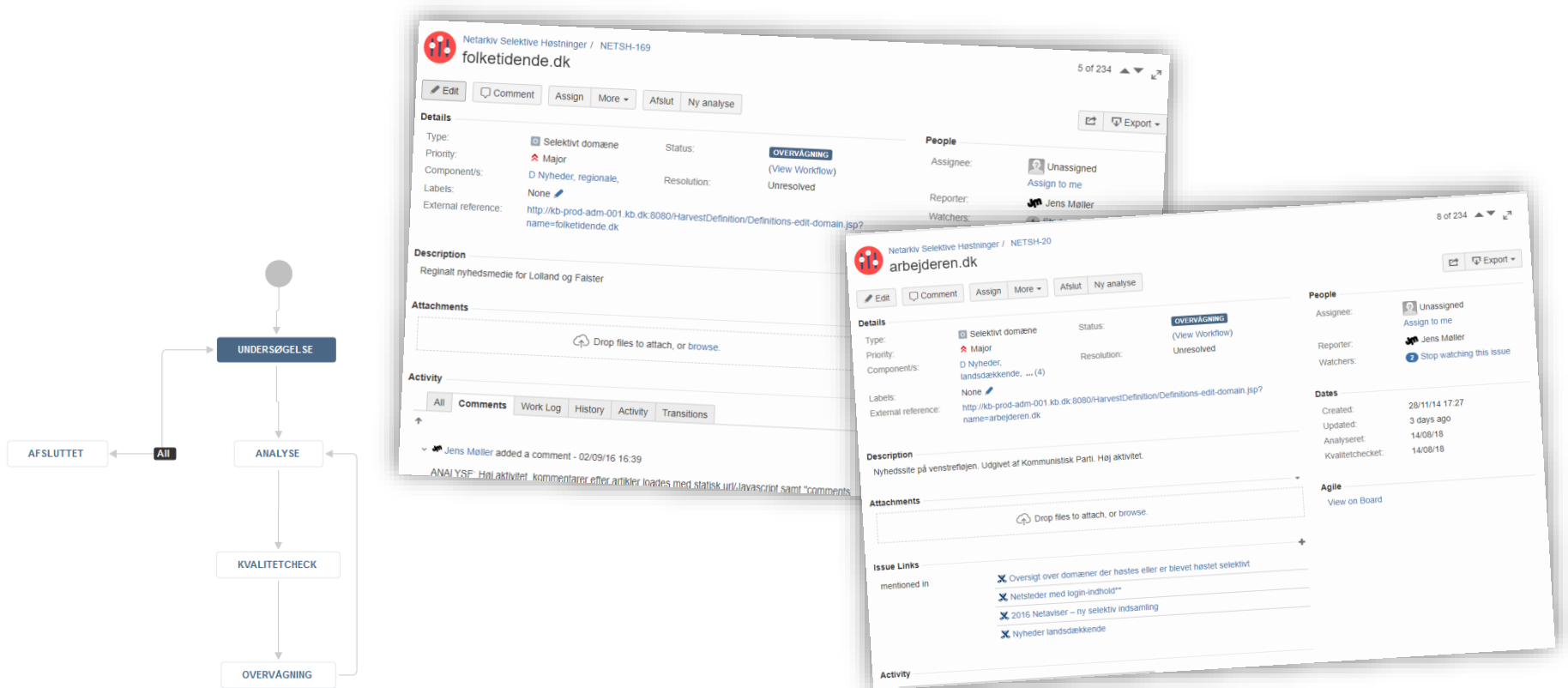
Kommentarer (fx 'er lagt i systemet', 'høstes med login' etc.)

KON 2007-01-08: #Cielnet opsætning
LAN 2007-01-24: Ligger allerede i systemet med den nyeste indsamlingsprofil. Skal kvitteres
JUL 2008-11-10: Indsamlingsprofil fortsat OK
LO 2010-04-22: Kvitteret OK
JUL 2010-11-30: Begge konfigurationer er kvitteret OK
Vær opmærksom på mega-lange tods taler på enkelte blæder, men de ER også høstet og vares ok
JUL 2011-07-13: Domængrense sat på 91 GB for at nedbringe antallet af jobs fra 5183 (indsamlingen) til 5183 (indsamlingen) Karl_Dagbj_2_jul

```
= [[DEC Alpha]]  
= [[AX CRIS]], [[Fujitsu FR]],  
[[M32R]], [[m68k]], [[Imaginat]],  
[[Nios II]], [[OpenRISC]], [[PA-RISC]],  
[[Tilera Technologies|TILE64]], [[Un]]  
= [[monolithic Kernel]]  
= GPLv2 and andere  
ges  
= "linux" (dt. {{{IPA|'li:nɔks}}}) oder  
in der Regel [[Freie Software]
```



Another migration - Atlassian Confluence and Jira



Atlassian jira and confluence Pros

- Visual editor
- Different overview displays, search and filtering is feasible
- Simultaneous editing without "destroying" each others edits
- From manual to automated registration of tokens, dates, activities...
- Individual access management for every single page
 - Documentation along with the collections for researchers possible
 - Internal informations (such as sensitive data, contact informations to website owners) can be restricted to the curator team



Challenge: a kind of standard for formulations

- We want to share the documentation with our researchers
- No insider expressions: sharing requires formulations understandable by persons outside from the curator team.



Questions?

- Sabine Schostag sas@kb.dk
- Netarkivet kontakt@kb.dk

