

Gemeenschappelijke Metadata Harvester

nieuwe 'WISH' 28 oktober 2014

Aanleiding

- KB harvester moeite met variatie materiaal
 - beperkte zichtbaarheid e-Depot
- NARCIS wel succesvol
 - sterke zichtbaarheid NARCIS

Normalisatie

Normalisatie een probleem voor minstens drie partijen:

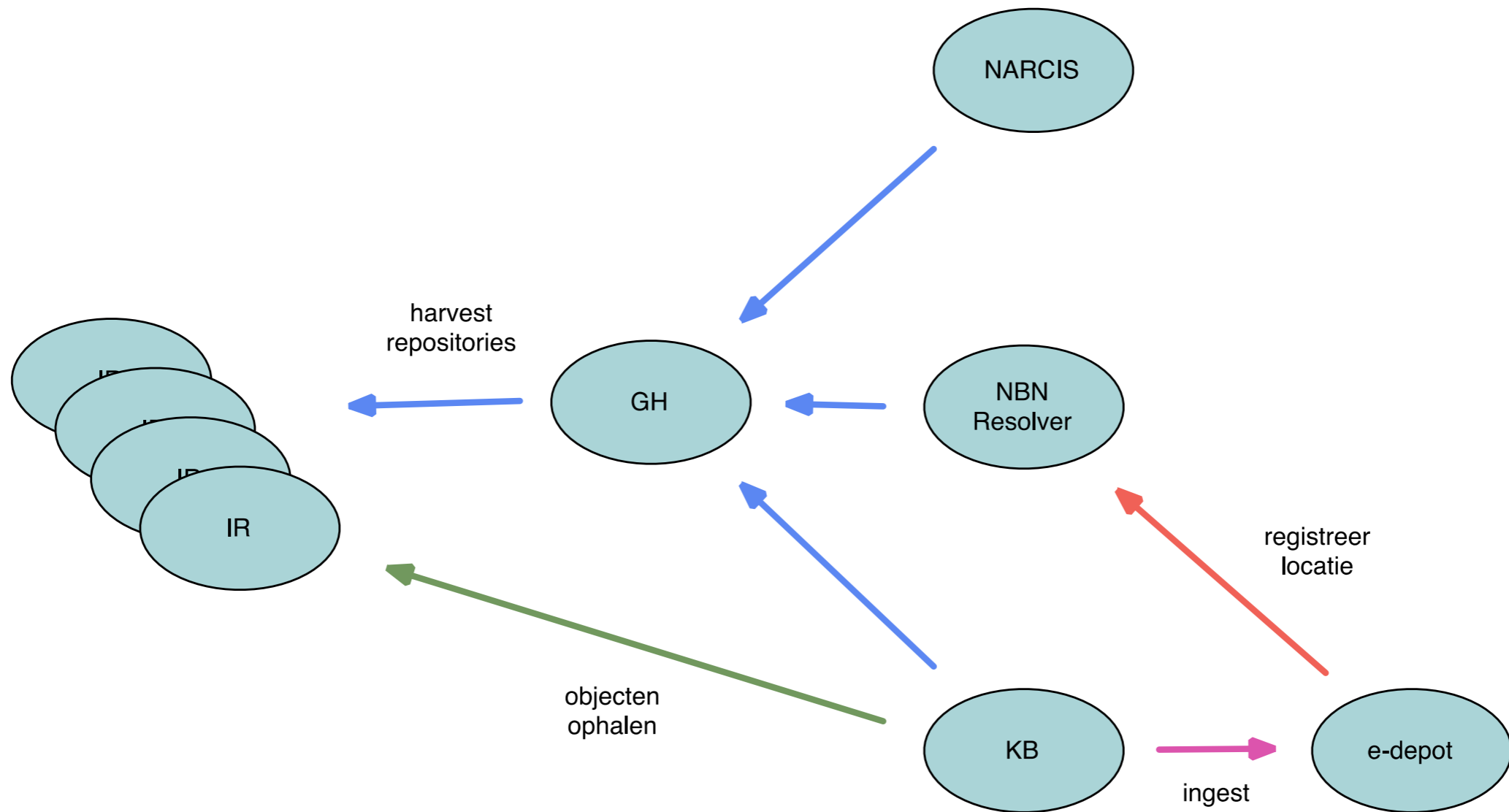
- KB e-Depot
- NARCIS
- Nationale Resolver

Oplossing

Oplossing: gezamenlijke harvesting & normalisatie van metadata

- Efficiënter: normalisatie op één plek.
- Eenduidiger: één aggregator van de metadata, betere synchronisatie.

Model



Werking GMH

- Harvester gebaseerd op techniek NARCIS harvester.
- Validatie mbv schema's.
- Normalisatie
- Opslag originele én genormaliseerde records
- Uitlevering via OAI-PMH (origineel, genormaliseerd of combinatie)

Stand van Zaken

- Afgelopen periode finetuning van o.a. normalisatie en logging.
- Nu afstemming met repositories over normalisatieproblemen
 - afspraken niet duidelijk, strijdig of achterhaald
 - behoefte aan een 'WISH' platform.

Wat gaat er zoals mis?

- MODS: Geen valide MODS genre
- MODS: Verplichte originInfo/dateIssued ontbreekt
- DIDL: No accessRights found for objectfile
- DIDL: Resource location niet gevonden (urn:nbn url)
- MODS: titleInfo/title has empty value
- DIDL: ongeldige URI (niet toegestane tekens, bijvoorbeeld []).
- MODS: ongeldig xs:ID (bv _123)
- DIDL: Geen resource URL in human startpage
- MODS: Metadata niet geldig, child ontbreekt.
- MODS: Ongeldige taalcode

Toekomst

- KB nu in overgangsfase directe harvesting naar harvesting via GMH
- Overgang NARCIS en Resolver in planning
- Todo: Aanpak voor één loket
- Aansluiten op ontwikkelingen toepassingsprofielen