**Hoeveel “unique identifiers” heb ik nodig voor Dutch Ships and Sailors?**

Beschrijving voor DSS (Clarin IV)

Conceptversie

Jurjen R. Leinenga

03-05-2013

In het artikel “If you build it, will we come? Large scale digital infrastructures as a dead end for digital humanities” van Jos van Zundert worden drie hoofdrequirements genoemd die ook voor ons project van belang zijn: data uniquely identified, discoverable and usable.

1. *Data uniquely identified*

In de simpelste vorm moet in het project een persoon – zeeman, soldaat, ambachtsman, ambtenaar of anders – gekoppeld kunnen worden aan een schip. De kardinale vraag in dit document is: wanneer heb ik voldoende zekerheid over het bestaan van een unieke persoon en van een uniek schip om ze te kunnen koppelen? Hieronder volgt een opsomming van mogelijke identifiers, opmerkingen over eventuele problemen daarbij en met tenslotte de vraag welke combinaties hiervan voor het vaststellen van het bestaan van iemand of van een schip noodzakelijk zijn.

Persoon

1. Koppeling met naam. Verschillende systemen van achter- en voornamen bekend, soms tegelijkertijd in dezelfde bron. Zeer arbeidsintensief om te standaardiseren. Wat is voldoende als work around? Wat Generale Zeemonsterrollen betreft: soms alleen de voornaam vermeld, soms alleen de achternaam (bij voormannen o.a.) of alleen algemene aanduidingen als “Generale Moorse matroosen uijt Bengalen”.
2. Koppeling met plaats. VOC-Opvarenden geeft alleen de geboorteplaats onder de noemer “herkomstplaats”, 19e eeuwse monsterrollen vermelden de woonplaats, eind 19e eeuwse-20e eeuwse monsterrollen geboorteplaats en/of woonplaats. Misschien zijn in VOC-Opvarenden ook koppelingen met bv. plaats van overlijden mogelijk. Wat Generale Zeemonsterrollen betreft: herkomstplaats soms heel globaal. GeoNames geeft op dit moment alleen de moderne naamvariant van een plaatsnaam, maar wel meertalig (bv. Revall, Tallinn etc.). Ze hanteert daarbij de volgende richtlijnen, volgens <http://www.geonames.org/manual.html> (geraadpleegd op 03-05-2013):
   * the main name for the toponym should be a widely accepted international or English name. Local languages are entered as alternate names.
   * The language code for the alternate names are normally the 2-character [ISO 639 language codes](http://www.loc.gov/standards/iso639-2/php/English_list.php), for more exotic languages that do not have a 2-character ISO code the 3-character code is used instead. Some pseudo language codes are used to mark other entries like postal codes, abbreviations or links (see below).
   * the flag 'isPreferred' helps distinguish between several alternate names in the same language. It marks the most commonly used name.
   * proper casing should be used. Upper case only for the first character of a term or for abbreviations.
   * avoid redundancy. Do not add the same name with 'minus' and blanks between terms, only add the more often use spelling variant in this case. (Ex: "La Colle sur Loup" for "La Colle-sur-Loup", it is sufficient to add "La Colle-sur-Loup" as the geonames search engine can handle searches with or without minus.)
   * use full spelling for 'Sankt' (German), 'Saint' and 'Sainte' (French). The respective abbreviations St. and Ste. are automatically handled by the search engine.
3. Koppeling met tijd: leeftijd, datum van aan- en/of afmonstering, van promotie of andere gebeurtenis, geboorte of overlijden.
4. Koppeling met eigenschap, bv. rang, bemanningslid van een bepaald schip in een bepaald jaar.
5. Koppeling met afbeelding. Aangezien het meestal om hele oude afbeeldingen zal gaan, is de betrouwbaarheid van de koppeling aan een naam niet eenvoudig na te gaan, behalve in het geval van officiële foto’s zoals voor een paspoort.
6. Koppeling met document. Voor koppeling aan bemanningslijsten of monsterrollen: verschillende manieren van annotatie. Een volgorde kan zijn: plaats van bewaarinstelling, naam ervan, naam van het archief, inventarisnummer, folionummer, vermelding ontstaansdatum of registratienummer. Kan dat allemaal in één veld of is het beter dit te splitsen? Voor koppeling aan geboorte-, trouw- en/of sterfteregisters: zie [www.wiewaswie.nl](http://www.wiewaswie.nl) , maar hoe betrouwbaar is die site? Die vraag geldt ook voor alle genealogische sites.
7. Koppeling met gebeurtenis of event. Voor koppeling aan gebeurtenissen: notariële archieven en oline krantenarchief van bv. de KB.

**Om voldoende zekerheid over het bestaan van een unieke persoon te krijgen, zijn tenminste Naam en Tijd belangrijk, aangevuld met òf Plaats òf Eigenschap: A en C, met B of D erbij .**

Schip

1. Koppeling met naam. Registratie van scheepsnamen is bij “onze” vier databases niet identiek. Vreemde talen kunnen registratie bemoeilijken. Is standaardisatie nodig? Zo ja, kunnen we dat erbij nemen? Wat te doen met afkortingen? Zijn de namen van schepen voor de inter-Aziatische vaart vóór 1800 bekend?
2. Koppeling met scheepstype. Bepaling scheepstypen internationaal redelijk unaniem. Wat te doen met verkleinwoorden? Hebben we wat aan hele algemene aanduidingen, zoals “vissersschip”?
3. Koppeling met scheepsgrootte. Welke standaardmaat gaan we hanteren, of houden we ons aan wat er in de bron staat? Mogelijkheden: (diverse) lasten, tonnen, kubieke meters. Sinds 1969 staan deze scheepsgegevens in een meetbrief. De voorloper va de meetbrief is de bijlbrief, waarin ook informatie over de maten van een schip zijn opgenomen.
4. Koppeling met tijd. Voor de tijd na 1800 kan het scheepsregister of scheepskadaster en ook de website [www.marhisdata.nl](http://www.marhisdata.nl) belangrijk zijn. In het algemeen geldt dat een vermelding in het scheepskadaster leidend moet zijn, daarna documenten als bijlbrieven, zeebrieven en aan- of verkoopbescheiden. In havens die in de toenmalige tijd onder Russisch bestuur vielen, zal de Juliaanse kalender zijn gehanteerd in plaats van de Gregoriaanse. Dit kan van belang zijn bij bijmonsterringen in deze havens.
5. Koppeling met gezagvoerder. Komt in DAS en in vrijwel alle monsterrollen standaard voor. Op schilderijen en foto’s kan een kapiteinsvlag aangeven bij welk zeemanscollege de kapitein hoorde en zijn specifieke nummer geeft dan de specifieke kapitein weer; zie hierover <http://nl.wikipedia.org/wiki/Zeemanscollege> , geraadpleegd 03-05-2013.
6. Koppeling met reder. Bij niet-VOC schepen wat problematisch: wie gaat dit in lijsten van bv. Lloyd’s of Sweijs allemaal na? Staat namelijk niet vermeld in de monsterrollen, evenmin de thuishaven.
7. Koppeling met afbeelding. Aangezien het meestal om hele oude afbeeldingen zal gaan, is de betrouwbaarheid van de koppeling aan een naam niet eenvoudig na te gaan, behalve in het geval van officiële schilderijen met bv. kapiteinsvlaggen of foto’s zoals voor een proefvaart. Een aparte categorie vormen technische bouwtekeningen van een schip. Handig hierbij is de website [www.maritiemdigitaal.nl](http://www.maritiemdigitaal.nl) , maar die bevat niet alle afbeeldingen in Nederlandse maritieme musea. Zie verder wat vermeld onder V over kapiteinsvlaggen.
8. Koppeling met document. Voor VOC-schepen is het DAS-nummer een belangrijk identificatielabel voor koppeling van een persoon aan een scheepssoldijboek. Scheepspapieren zijn leidend.
9. Koppeling met een plaats, gebeurtenis of event. In kranten zijn Scheepsberichten, Scheepstijdingen en Zeetijdingen belangrijke rubrieken, evenals “scheepsramp”, “scheepsdoop”, “scheepslading”, “scheepswerf”, “schipbreuk”, “waterlating” en “zeereis”. Ernstige verminkingen van scheepsnamen geen zeldzaamheid. Koppeling aan een reis met ev. ligplaats, vertrekhaven en bestemming is bij sommige typen reizen wat moeizaam, zoals bij vissersschepen, walvisvaarders, marineschepen op patrouille en koopvaardijschepen die een route (circuit) varen. Aparte koppelingen vormen de berichten van het vergaan van schepen bij grote stormen, zoals die bv. gerapporteerd zijn door historisch geograaf Jan Buisman of het KNMI. Duikers zoals Arent Vos (leider van het voormalige Nederlandse nationale duikteam) gebruiken dit soort koppelingen.
10. Koppeling met roepletters voor radiotelefonie (vanaf begin 20e eeuw). Bij uitschrijving van een schip uit het scheepsregister kan een roeplettercombinatie opnieuw in gebruik worden genomen door een ander schip.

**Om voldoende zekerheid over het bestaan van een uniek schip te krijgen, zijn tenminste Naam, Scheepstype, Tijd en Gezagvoerder belangrijk, eventueel aangevuld met Grootte en/of Plaats: I, II, IV en V, met III of IX erbij .**

1. *Data discoverable*

Om de gebruikte primaire bron(nen) terug te kunnen vinden, is een verwijzing vereist. Dit kan naar het origineel zelf, een scan of andere vorm van afbeelding ervan of beide. Op dit moment spreiden de archiefverwijzingen in de databases zich uit over meerdere velden. Hoe wil Victor dit?

1. *Data usable*

Dit heeft betrekking op de queries die we met zijn allen, nu en in de toekomst, kunnen bedenken. Om vragen uit het onderzoeksveld te kunnen afvangen, dient de portal hier een loket voor te hebben. Als er ten aanzien van cijfermatige gegevens kanttekeningen moeten worden gemaakt om te vermijden dat statistische analyses worden uitgevoerd waar de bron dat eigenlijk niet toelaat, moet dat duidelijk in een inleiding of handleiding worden aangegeven. Het kan handig zijn om iets te maken voor het omrekenen van bv. lasten naar tonnen, kubieke meters en brutoregistertonnage (BRT of GRT) of nettoregistertonnage (NRT); zie hierover <http://nl.wikipedia.org/wiki/Tonnenmaat> . Zie voor omrekeningen van een muntsoort naar een andere: <http://www.iisg.nl/hpw/calculate-nl.php>