Data Archiving and Networked Services

DANS

Data and Metadata Instantiation

Use Cases and a Conceptual Model

Richard P. Smiraglia, Visiting Professor

Richard.smiraglia@dans.knaw.nl

DCMI, Crystal City, VA, October 26-29, 2017

Abstract: Instantiation describes the phenomenon of variation in representation of information objects over time. Smiraglia (2008) describes it as the diatonic problem of both clustering and disambiguation of groups of what appear to be, but are not quite, iterations of the same object. Although the problem is well-known in bibliographic information retrieval (Smiraglia 2001), it also is well-documented among other kinds of information objects. Greenberg (2009) demonstrated instantiation among metadata records of evolutionary biologists, Coleman (2002) drew an analogy to instantiation among scientific models, and Smiraglia (2005) found instantiation among archival records of artifacts in a museum of archeology. As Greenberg points out (399), the problem is particularly acute in digital repositories where "automatic propagation, metadata inheritance, and value system adoption" contribute to a "lifecycle" that creates potentially ambiguous clusters.

Digital repositories are particularly susceptible to the problem of uncontrolled data and metadata instantiation because of the complex lifecycles of data deposit, use, and reuse. In repositories that require deposit of research data on a large scale, instantiation can become particularly acute. DANS (Data Archiving and Networked Services), a division of the Royal Netherlands Academy of the Arts and Sciences, is the "institute for permanent access to digital research resources" in The Netherlands (DANS 2017). The role of DANS is to encourage scholars to make their data accessible, interoperable and resusable, in a sustainable environment. In addition to serving as a host repository for tens of thousands of datasets, DANS also manages the NARCIS gateway to more than 160,000 datasets generated by Dutch scholars.

Recent research (Smiraglia and Park 2016) demonstrated one approach to a conceptual model of instantiation among open government data records, deriving core attributes "information object," "expression," "manifestation product type," "actor, "expression creation," and "information carrier" from the FRBRoo ontology of bibliographic instantiation. The proposed presentation combines these and other FRBRoo attributes with the generations of lifecycle modeling identified by Greenberg, as applied to a series of use cases from DANS.

Instantiation typology

Derivations		Bibliographic Works	Artifacts-Metadata	Artifacts-Representations	Personal Papers
	instantiation	simultaneous editions	finding aids	field photos	Photocopies
		successive editions	field notes	working images	Carbon copies
	accretion	amplifications	letters	exhibition color images	Photos
	depletion	extractions	conservation treatment notes	digitized exhibition images	postcard with photo
	re-presentation	musical presentations	register descriptions; object cards	conservation photos	digitized scan of postcard with photo
		notational transcriptions	image order invoices	archived photographic negatives	reprint of photo
		performances	museum database records	archived photographic prints	digitized scan of photo
			catalog card records	archived photographic transparencies	
Mutations	**alteration**	translations		object reproductions	
		adaptations		drawings	
				3D models	

(adapted from Smiraglia 2008, 15-16)

Life-cycle from DRYAD

Propagation

Reuse … may result in the creation of an offspring or progeny, connoting a sense of propagation, stemming from an initial progenitor (Greenberg 2009, 385)

Inheriting characteristics

Given propagation … seemingly organic progenies, do inherit characteristics from their original source and adopt values from outside systems (Greenberg 2009, 392)

Adopting values

Mapping to standardized KOS (Greenberg 2009, 394)

Value propagation vs. inheritance

Fido—Dog—Mammal—Thermoregulation=Warm-blooded; subclass and instance do not match *IsGatheredInto* (Renear et al. 2008, 84)

Value constraint

Values vary with the semantics of related attributes (Renear et al. 2008, 85)

Deriving core attributes (FRBRoo)

Instantiation of Open Government Data

information object

information carrier

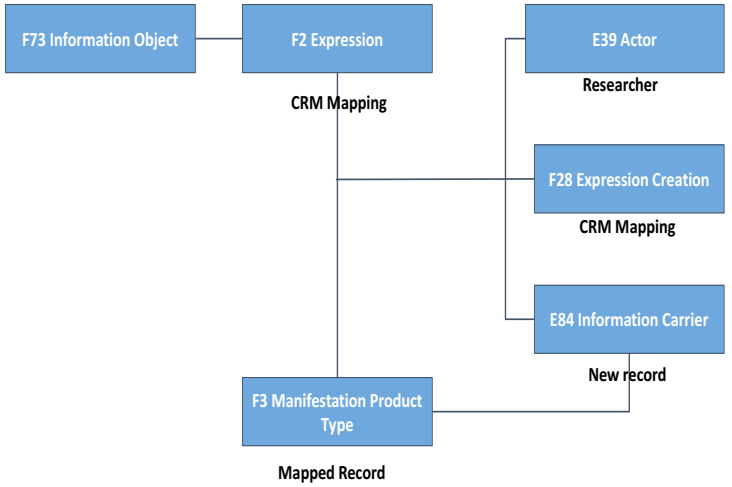
expression

expression creation

manifestation product type

actor

(Smiraglia and Park 2016; 2017 forthcoming)



Example 1:Derivation (accretion), Propogation, Inherited characteristics, New information object

OViN2015, access category: *Restricted (Request permission)*

Centraal Bureau voor de Statistiek (CBS); Rijkswaterstaat (RWS) (2016): *Onderzoek Verplaatsingen in Nederland 2015 - OViN 2015*. DANS. https://doi.org/10.17026/dans-z38-prz4

Has version:

Centraal Bureau voor de Statistiek (CBS); Rijkswaterstaat (RWS) (2017): *Onderzoek Verplaatsingen in Nederland 2015 - OViN 2015 versie 2.0*. DANS. https://doi.org/10.17026/dans-z2v-c39p

Remarks

Per 8 augustus is een verbeterde versie beschikbaar, OVIN 2015 versie 2.0, Centraal Bureau voor de Statistiek (CBS); Rijkswaterstaat (RWS) (2017): Onderzoek Verplaatsingen in Nederland 2015 - OVIN 2015 versie 2.0. DANS. DOI: 10.17026/dans-z2v-c39p https://doi.org/10.17026/dans-z2v-c39p. Zie ook het veld 'Relation' voor de link naar deze nieuwe versie.

Per 4 mei 2017 is een nieuwe versie van de Plausibiliteitsrapportage 2015, v. 1.1 30-11-2016 geupload in de dataset. In het Erratum van deze rapportage staan de wijzigingen vermeld t.o.v. de oude versie 1.0. Op verzoek van de deponeerder is de oude versie niet meer downloadbaar.

Rijkswaterstaat heeft de databestanden geleverd in SAV en XLS formaat. DANS heeft voor het SAV bestand de POR en DTA formaten toegevoegd en voor de XLS bestanden de CSV formaten toegevoegd. Deze bestanden zijn te vinden in de map 'DANS toegevoegde formaten'.

Description

***LET OP: dit is een oudere versie van OVIN 2015. Per 8 augustus 2017 is OVIN 2015 versie 2.0 beschikbaar. De DOI, tevens link naar OVIN 2015 versie 2.0 vindt u in het veld 'Relation' hieronder. ***

Het Onderzoek Verplaatsingen in Nederland is een continu dagelijks onderzoek naar het verplaatsingsgedrag van de Nederlanders. Aan de respondenten wordt gevraagd om voor één bepaalde dag van het jaar bij te houden waar ze die dag naartoe gaan, met welk doel, met welk vervoermiddel en hoe lang het duurt om er te komen. Daarnaast wordt gevraagd naar enkele algemene persoons- en huishoudkenmerken en naar zaken zoals rijbewijs- en vervoermiddelenbezit. Op grond van dit onderzoek kan informatie worden verkregen over alle dagelijkse verplaatsingen door Nederlanders op Nederlands grondgebied.

OVN is een voortzetting van het Mobiliteitsonderzoek Nederland (MON) dat van 2004 t/m 2009 is uitgevoerd door Rijkswaterstaat. Voor 2004 werd het onderzoek onder naam van Onderzoek Verplaatsingsgedrag (OVG) door CBS uitgevoerd.

In 2013 is een vraagstelling over bezit en gebruik van elektrische fiets aan OVN toegevoegd. De vragen over de elektrische fiets voor het waarnemen van vervoermiddelenbezit in het huishouden bleek daarbij tot onduidelijkheid te leiden. Daardoor is de informatie over bezit van fietsen in huishoudens in OVIN 2013 en 2014 niet goed waargenomen. De informatie over de fietsmobiliteit / fietsgebruik, zowel voor gewone fiets als elektrische fiets, in OVIN 2013 en 2014 zijn wel betrouwbaar waargenomen. Met ingang van 2015 is de vraagstelling van fietsbezit in het huishouden verbeterd, waardoor fietsbezit weer correct wordt waargenomen.

is version of • Onderzoek Verplaatsingen in Nederland 2015 - OViN 2015

has version • Onderzoek Verplaatsingen in Nederland 2015 - OViN 2015 versie 2.0

Example 2:Derivation (accretion), Propogation, Inherited characteristics, Constrained inheritance, New information object

FEM growth and yield data - Poplar roadside plantations, access category: *Restricted (Request permission)*

Ouden, Dr. ir. J. den (Wageningen UR); Mohren, Prof. dr. ir G.M.J. (Wageningen UR); Jansen, Ir. J.J. (Wageningen UR) (1997): *FEM growth and yield data - Poplar roadside plantations*. DANS. https://doi.org/10.17026/dans-zb2-p2h2

Has version:

Ouden, Dr.ir. J. den (Wageningen University & Research, Forest Ecology and Forest Management); Mohren, Prof.dr.ir. G.M.J. (Wageningen University & Research, Forest Ecology and Forest Management); Jansen, Ir. J.J. (Wageningen University & Research, Forest Ecology and Forest Management) (2016): *FEM growth and yield data - Poplar roadside plantations (revised version)*. DANS. https://doi.org/10.17026/dans-xky-zh99

Description	The current database is part of the FEM growth and yield database, a collection of growth and yield data from even-aged monocultures (douglas fir, common oak, poplar, Japanese Larch, Norway spruce, Scots pine, Corsican pine, Austrian pine, red oak and several other species with only a few plots, even-aged mixed species forest plots, uneven-aged natural forest, uneven-aged selection forest and roadside plantations of poplar. The FEM growth and yield data base is currently supervised by Jan den Ouden and Frits Mohren. The growth and yield research on poplar roadside and other line plantation was initiated by prof. dr. J.H. Becking in 1951. During almost five decades the project was supervised by A. van Laar, P.G. de Vries, J.H. Hildebrand and H.H. Bartelink. The project stopped after measuring the last remaining stand in 1997. Within the overall database it is called study 13.

is part of • Thematic collection: FEM growth and yield data

is version of • FEM growth and yield data - Poplar roadside plantations (old version)

Example 3:Mutation, Propogation, Inherited characteristics, New information object

Kenya public weather processed by the Global Yield Gap Atlas project, access category: *Open Access (CC0 Waiver)*

Groot, Ir. H.L.E. de (Wageningen UR, Alterra, Earth Observation and Environmental Informatics); Adimo, O. (Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya); Claessens, L. (International Crops Research Institute for the Semi-Arid Tropics, Nairobi, Kenya); Wart, J. van (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Bussel, Dr. ir. L.G.J. van (Wageningen UR, Plant Production Systems); Grassini, P. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Wolf, J. (Wageningen, Plant Production Systems); Guilpart, N. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Boogaard, Ir. H.L. (Wageningen UR, Alterra, Earth Observation and Environmental Informatics); Oort, Dr. ir. P.A.J. van (Wageningen UR, Centre for Crop System Analysis); Yang, H.S. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Ittersum, Prof. dr. ir. M.K. van (Wageningen UR, Plant Production Systems); Cassman, K.G. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA) (2016): *Kenya public weather processed by the Global Yield Gap Atlas project*. DANS. https://doi.org/10.17026/dans-xc8-3a2q

Has version:

Groot, Ir. H.L.E. de (Wageningen Environmental Research (Alterra)); Adimo, O. (Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya); Claessens, L. (International Crops Research Institute for the Semi-Arid Tropics, Nairobi, Kenya); Wart, J. van (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Bussel, Dr. ir. L.G.J. van (Plant Production Systems, Wageningen University & Research); Grassini, P. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Wolf, J. (Plant Production Systems, Wageningen University & Research); Guilpart, N. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Boogaard, Ir. H.L. (Wageningen Environmental Research (Alterra)); Oort, Dr. ir. P.A.J. van (Plant Production Systems, Wageningen University & Research); Yang, H.S. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Ittersum, Prof. dr. ir. M.K. van (Plant Production Systems, Wageningen University & Research); Cassman, K.G. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA) (2016): *Kenya public weather processed by the Global Yield Gap Atlas project (revised version)*. DANS. https://doi.org/10.17026/dans-zyu-xkhc


KENYA PUBLIC WEATHER PROCESSED BY THE GLOBAL YIELD GAP A

Overview Description Data files (16)

Cite as:

Groot, Ir. H.L.E. de (Wageningen UR, Alterra, Earth Observation and Environmental Informatics); Adimo, O. (Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya); Claessens, L. (International Crops Research Institute for the Semi-Arid Tropics, Nairobi, Kenya); Wart, J. van (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Bussel, Dr. ir. L.G.J. van (Wageningen UR, Plant Production Systems); Grassini, P. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Wolf, J. (Wageningen, Plant Production Systems); Guilpart, N. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Boogaard, Ir. H.L. (Wageningen UR, Alterra, Earth Observation and Environmental Informatics); Oort, Dr. ir. P.A.J. van (Wageningen UR, Centre for Crop System Analysis); Yang, H.S. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Ittersum, Prof. dr. ir. M.K. van (Wageningen UR, Plant Production Systems); Cassman, K.G. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA) (2016): *Kenya public weather processed by the Global Yield Gap Atlas project.* DANS. https://doi.org/10.17026/dans-xc8-3a2q

2016-01-05 | Groot, Ir. H.L.E. de (Wageningen UR, Alterra, Earth Observation and Environmental Informatics); Adimo, O. (Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya); Claessens, L. (International Crops Research Institute for the Semi-Arid Tropics, Nairobi, Kenya); Wart, J. van (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Bussel, Dr. ir. L.G.J. van (Wageningen UR, Plant Production Systems); Grassini, P. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Wolf, J. (Wageningen, Plant Production Systems); Guilpart, N. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Boogaard, Ir. H.L. (Wageningen UR, Alterra, Earth Observation and Environmental Informatics); Oort, Dr. ir. P.A.J. van (Wageningen UR, Centre for Crop System Analysis); Yang, H.S. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA); Ittersum, Prof. dr. ir. M.K. van (Wageningen UR, Plant Production Systems); Cassman, K.G. (Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, USA) | 10.17026/dans-xc8-3a2q

A revised version of this dataset has been published: https://doi.org/10.17026/dans-zyu-xkhc. The files of this dataset are therefore no longer accessible.

Relations

is version of
- Kenya public weather processed by the Global Yield Gap project (old version)

- HasVersion: Kenya public weather processed by the Global Yield Gap Atlas project (revised version)

Model

Both the data and the metadata show:

Derivation/Mutation

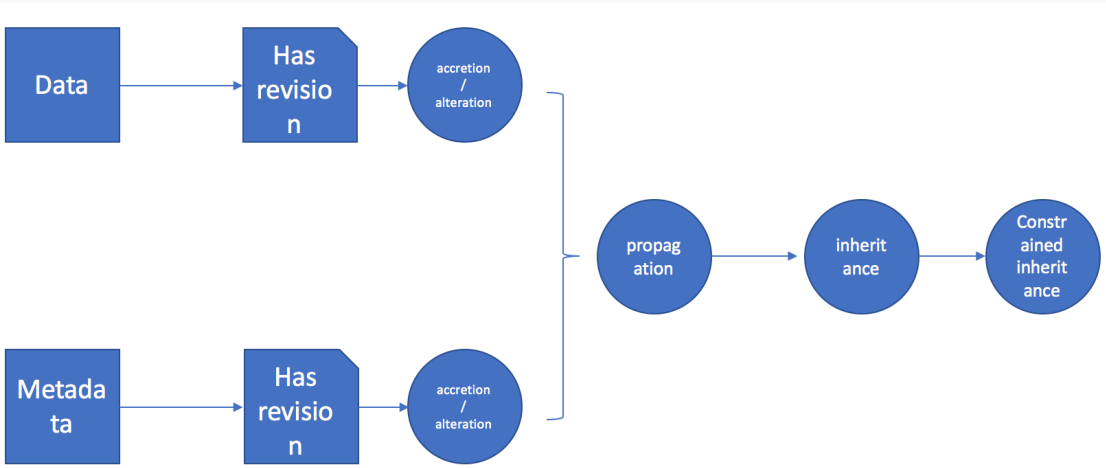
accretion/alteration

propogation

inherited characteristics

constrained inheritance

new information object



Thanks!

richard.smiraglia@dans.knaw.nl

Works Cited

Data Archiving and Networked Services (DANS). https://dans.knaw.nl/en/about accessed 12 June 2017.

FRBRoo (Functional Requirements for Bibliographic Records—object oriented), an extension of the CIDOC-CRM. http://old.cidoc-crm.org/frbr_inro.html# accessed 12 June 2017.

Coleman, Anita S. 2002. Scientific models as works. *Cataloging & classification quarterly* 33 no. 3/4:129-59

Greenberg, Jane. 2009. Theoretical considerations of lifecycle modeling: An analysis of the Dryad repository demonstrating automatic metadata propagation, inheritance, and value system adoption. *Cataloging & classification quarterly* 47, no.3:380-402.

Renear, Allen H., Karen M. Wickett, Richard J. Urban, Dave Dubin, and Sarah L. Shreeves. 2008. Collection/Item metadata relationships. *Proceedings of the International Conference on Dublin Core and Metadata Applications 2008 Berlin*, ed. J. Greenberg and W. Klaus. Dublin Core Metadata Initiative, 80-9.

Smiraglia, Richard P. 2001. *The nature of a work: Implications for the organization of knowledge.* Lanham, Md.: Scarecrow Press.

Smiraglia, Richard P. 2005. Content metadata: An analysis of Etruscan Artifacts in a museum of archeology. *Cataloging & classification quarterly* 40 no. 3/4:135-51.

Smiraglia, RIchard P. 2008. A meta-analysis of instantiation as a phenomenon of information objects. *Culture del testo e del document* 9, no. 25:5-25.

Smiraglia, Richard P. and Hyoungjoo Park. 2016. Using Korean Open Government Data for data curation and data integration. Presentation, International Conference on Dublin Core and Metadata Applications, DC-2016, Copenhagen, Denmark. http://dcevents.dublincore.org/IntConf/dc-2016/paper/view/447/509 Accessed 12 June 2017.

Smiraglia, Richard P. and Hyoungjoo Park. 2017. Ontological data-sharing of Open Government Data for data curation. Unpublished manuscript under review.