C. Boyd
Last update: 2022/02/16

README: *Data as Assemblage* Dataset Documentation

Table of Contents

Introduction **1**

Data Files **2**
- 1) concept_assessment.tab 2
  - Description 2
  - Related Materials 3
  - References 3
- 2) concept_assessment_bibliography.pdf 3
  - Description 3
  - Related Materials 3
- 3) assemblage_ontology.owx and assemblage_ontology.tab 3
  - Description 3
  - References 4

Assemblage Model Concept Definitions **4**

Introduction

This file contains documentation for the *Data as Assemblage* dataset (https://doi.org/10.7910/DVN/20QVQB). The dataset contains files that support the following publications:

> Boyd, C. (2019). *Towards a domain ontology for data assemblages*. iConference, 2019, College Park, Maryland. https://doi.org/10.21900/iconf.2019.103347
>
> Boyd, C. (2022). (in press) Data as Assemblage. *Journal of Documentation*.

In addition to this README file, the dataset includes four files described in detail below. The last section of this document also contains a table of full definitions for the concepts associated with the assemblage model (see Boyd, 2022).

Data Files

1) `concept_assessment.tab`

Description

This data file is referenced in Boyd (2022) and contains the assemblage assessment of 60 definitions of data and data-related terms including research data, documents, texts, databases, and data repositories. Each definition describes a population of occurrences of data, documents, information, texts, datasets, and other data-related phenomena.

I evaluated each definition to identify concrete examples that met or failed to meet each of DeLanda's (2016) eight assemblage criteria. According to DeLanda, assemblages are (1) real; (2) composed of heterogeneous parts; (3) non-hierarchical and do not have intrinsic qualities; (4) exist in time, space, and can be nested within one another; (5) changeable; (6) have an adjustable underlying pattern of organization; (7) are members of a population, and (8) can be disconnected from one assemblage and connected to another.

For each definition and criteria, I assigned one of three category-values:

- Yes (2 points) - Occurrences of the definition met the criteria;
- Maybe (1 point) - Some occurrences of the definition failed to meet the criteria, the definition was ambiguous with respect to the specific criteria, or both;
- No (0 points) - Occurrences of the definition did not meet the criteria.

The definition assessment scores are shown in the table below. A perfect score is 16. Most definitions of the 60 definitions described occurrences that met assemblage criteria.

Score	Total # definitions	% Total
16	47	78%
15	4	7%
14	1	2%
13	0	0%
12	2	3%
11	3	5%
10	1	2%
9	2	3%

Related Materials

- `concept_assessment_bibliography.pdf`.

References

Boyd, C. (2022). (in press) Data as Assemblage. *Journal of Documentation*.

DeLanda, M. (2016). *Assemblage theory*. Edinburgh University Press.

2) `concept_assessment_bibliography.pdf`

Description

This document contains a bibliography of all sources for data and data-related definitions assessed in `concept_assessment.tsv`.

Related Materials

- `concept_assessment.tsv`.

3) `assemblage_ontology.owx` and `assemblage_ontology.tab`

Description

The file `assemblage_ontology.owl` encodes a *prototype* OWL formatted ontology of assemblage concepts and their relationships to one another. It was developed using WebProtégé (Horridge et al., 2019; https://webprotege.stanford.edu/), a Cloud-based ontology application. The file `assemblage_model.csv` is a comma-delimited format of the OWL file.

Please note, `assemblage_ontology.owl` is a *purely experimental model* that was developed as I conducted a close reading of DeLanda (2016) enroute to the research described in Boyd (2019). For instance, this ontology depicts 62 concepts in comparison to the greatly refined assemblage model described in Boyd (2022).

In addition, WebProtégé's underlying class-instance paradigm is at odds with DeLanda's neo-assemblage theory that rejects class hierarchies and envisions all occurrences as unique individuals as members of populations, rather than instances of abstract classes. Therefore, this prototype cannot truly be considered an assemblage model. However, I include it here in this dataset as a gesture towards how a machine-readable model *could be* constructed. Implementing a machine-readable assemblage model using Javascript is an alternative approach; Javascript's object model is based upon prototypes and does not depend on abstract classes.

References

Boyd, C. (2019). *Towards a domain ontology for data assemblages*. iConference, 2019, College Park, Maryland. https://doi.org/10.21900/iconf.2019.103347

DeLanda, M. (2016). *Assemblage theory*. Edinburgh University Press.

Horridge, M., Gonçalves, R. S., Nyulas, C. I., Tudorache, T., & Musen, M. A. (2019). WebProtégé: A Cloud-Based Ontology Editor. *Companion Proceedings of The 2019 World Wide Web Conference*, 686–689. https://doi.org/10.1145/3308560.3317707

Assemblage Model Concept Definitions

In Boyd (2022) I presented a theoretical model of an assemblage, described its anatomical elements, and mapped the elements to features of a dataset of U. S. Census population estimates. The table below provides definitions for each of the model's concepts.

Concept	Definition
Assembler	An assemblage is the result of a selection made by a living or nonliving *assembler*. The assembler has a *positionality* and chooses and arranges the assemblage components using a *selection procedure*.
Capacity	A *capacity* is the ability or potential of an entity to change or respond to an event. *Assemblages*, *components*, *properties*, and *mechanisms* have capacities, which are innumerable. Assemblage properties are the concrete results of it exercising its capacities.
Causality	An assemblage or component can assume the role of a *causal agent*. DeLanda's neo-assemblage theory explicitly mentions five types of *causal relations*: top-down causality, bottom-up causality, linear causality, nonlinear causality, and probabilistic causality. An assemblage's *parameters* (the value of a property at a specific time) change as its components interact with one another generating events--*causes*--whose effects are reflected in the values of its properties.
Components	A *component* is any element belonging to an assemblage. The nesting capacity of assemblages means that a component may also be an assemblage. Therefore, 1) the terms component and assemblage may be used interchangeably, and 2) the concept of a component may be described as a type of *role* that an assemblage can play within the scope of another assemblage. Components are independent entities and may not be aware of their membership in an assemblage.
Event	An *event* is a trigger or a catalyst and is the predecessor of one or more processes, effects, or states of a component or assemblage.

History	*History* refers to the sequence of events, processes, operations that have occurred to an entity and/or in which it has participated. Every assemblage will have a history, which is related to its identity. For instance, a dataset's history would include previous versions and its change history reflecting operations performed upon it such as additions or redactions.
Identity	The *identity* of an assemblage at any scale is the product of processes, including territorialization and coding, and is always precarious because other processes can destabilize that identity. The set of an assemblage's properties, parameters, interactions, and history are aspects of its identity. Assemblages such as data, datasets, and or data repositories may have enduring, unique identifiers including digital object identifiers (DOIs) and URNs.
Mechanism	*Mechanisms* are phenomena associated with an assemblage that bring about changes in properties or interactions between components or between other assemblages, thereby giving rise to causal relations (e.g., probabilistic causality) and processes (e.g., territorialization or coding). Mechanisms can be thought of as packages of causal relations that produce repeatable events. There are special mechanisms called *controls* that adjust assemblage parameters. DeLanda describes assemblages as being parameterized, having 'control knobs' with variable settings that can be adjusted. Mechanisms may be assemblages; for instance, algorithms are mechanisms that can be modeled as assemblages.
Motive	Deleuze and Guattari and DeLanda indicate that the assembler's *motives* and goals influence the selection process and the purpose of and final arrangement of the assemblage. A *point of view* will be encumbered with explicit or implicit motives and goals.
Point of View	*Point of view* denotes the unique perspective, degree of agency, philosophical underpinnings, and types of possible observations or actions for an assemblage or component given its degree of nesting within an assemblage of greater scale. *Point of view* is a property that all entities with agency have to make and act upon observations or events. A point of view includes one or more motives. It is the perspective of the data subject (or subjects) who may be a person (living or dead) or may be an instrument, or a system.
Population	In neo-assemblage theory, aggregates of assemblages, properties, capacities, and other elements are called *populations* rather than sets. The term *set* is not used because it is freighted with other connotations such as homogeneity, intrinsic characteristics, and exclusivity. In contrast, *populations* consist of diverse individuals that are similar according to some selection criteria. In the case of assemblages, a population is a group of assemblages having similar properties and virtual diagrams. There is no requirement that an assemblage is aware of its membership in a population or interacts with other members of the population.

Positionality	*Positionality* refers to the social and political context and characteristics of the assembler, which influences the selection process and the ultimate composition and behavior of the assemblage. Positionality is, itself, may be modeled as an assemblage.
Possibility Space	An assemblage's *possibility space* is expressed as the range (e.g., from a maximum to a minimum) for each of its properties. It is the set of all possibilities for the properties.
Process	Neo-assemblage theory is primarily concerned with *processes* that are causal and productive, producing changes in the properties of an assemblage, or population of assemblages or components. These processes are recurrent and repetitions can create or synthesize populations of assemblages. Common assemblage processes include *territorialization* (and *deterritorialization*) and *coding* (and *decoding*).
Property	*Properties* are the characteristics, qualities, features, or attributes of an assemblage or component that help to identify it as an individual or the member of a population. DeLanda indicates that properties are a "given and closed list" associated with an assemblage (DeLanda, 2006, p. 10). In addition to the three common properties of assemblages, DeLanda also describes five kinds of properties associated with assemblages that are derived from the physical sciences and philosophy: 1) intrinsic, 2) extrinsic, 3) intensive, 4) extensive, and 5) invariant. Each property has parameters that indicate the value of the property at a specific point in time.
Role	A *role* is a capacity of an assemblage denoting a set of assemblage properties dictating its behavior when the assemblage is nested within another assemblage. DeLanda explicitly defines one role: materiality and implies another: component. Using the lexicon of roles, an assemblage can assume the component role when it becomes the member of another assemblage. The assemblage can also play a material or expressive role in the scope of an assemblage. An assemblage may be composed of material and symbolic parts, but within an assemblage at a greater scale, it may play a more material role if its material components are more important, or it may play a symbolic role if required.
Scale	Assemblages and populations of assemblages exhibit properties of *scale* along at least three axes: time, space, and level of nesting. Time and space are invariant across assemblages, therefore one second or one meter is the same duration or length in every assemblage. The level of nesting, however, depends upon the specific assemblage under inspection, and cannot be compared across assemblages. For example, a nesting level of 3 is only meaningful in the context of an individual assemblage, it does not denote an ontological category at a specific location in a hypothetical hierarchy.

Scope	*Scope* is a computer science concept denoting an assemblage's sphere of influence or its context, including its components and their properties, relations, and events. Scope makes it possible to discuss the types of interactions that an assemblage may have with other assemblages and the degree of access that other assemblages have to its components.
Selection Procedure	A *selection procedure* is the procedure (e.g., set of methods) used by an assembler to create an assemblage. The selection process is informed by the assembler's *positionality* and *motives*.