Supporting documents for the Construction Classification System Database for Understanding Resource Use in Building Construction

[1. Introduction 1](#_Toc67554729)

[2. Guidelines for contributing to the database 4](#_Toc67554730)

[3. Variables 2](#_Toc67554731)

[4. Tables 4](#_Toc67554732)

[5. References 8](#_Toc67554733)

# 1. Introduction

Welcome to the Construction Classification System Database for Understanding Resource Use in Buildings.

This database provides a novel dataset and a building material data structure to facilitate study of resource use in building design and construction. The ontology developed for this database uses UniFormat (CSI and CSC, 2010) in conjunction with MasterFormat (CSI and CSC, 2016) for organizing and storing the building material data.

The dataset was developed by collecting design or construction drawings for the studied buildings and performing material take-offs based on the drawings. The ontology is based on Uniformat and MasterFormat to facilitate interoperability with existing construction management practices, and to suggest a standardized structure for future MI studies. The structure of the database and these guidelines builds on the structure presented by (Heeren & Fishman, 2019).

The initial database version is created by the research team supervised by Prof. Shoshanna Saxe at the University of Toronto and submitted to the journal Scientific Data (Guven et al. 2021) in October 2021 to describe the dataset and the associated methods and details.

Thank you for considering contributing to the database. Data contributors must follow the steps detailed below and must ensure that their inputs do not infringe any intellectual property or copyright agreements.

# 2. Variables

1. building\_identifier: A unique building identifier, e.g. 001, 002, 003.
2. country: The country where the building is located, based on the ISO 3166-1 alpha-2 code (International Organization for Standardization (ISO), 2021) (i.e. two-letter country codes defined in ISO 3166-1), e.g. CA for Canada.
3. city: The city where the building is located. First three letters of city names are used, e.g. TOR for Toronto.
4. quality\_or\_stage\_of\_data: Communicates the quality or stage of building drawings, demonstrating the level of completion of construction documents, such as Issued for Construction (00IFC) or Issued for Building Permit (0IFBP). Canadian Construction Association’s classification is used for the level of construction documents completion (Canadian Construction Association, 2012) (see Table 1 for full list).
5. construction\_date: It is the year (or projected year) of completion. For renovated buildings, construction date is the year the building was originally constructed.
6. building\_type: Shows the type of building that is quantified, e.g. single detached (SND), institutional (INS), educational (EDU). “R” in the building type code indicates that it is a renovated building (see Table 2 for full list).
7. contributor\_name: Identity of the contributor who added each subsequently-submitted column.
8. floor\_level: Describes the floor or part of building where the material is placed. 00R is for roof, 999 represents the whole building. Underground floors are named based on purpose of use (i.e. basement or parking). Basements are denoted with letter B (e.g. B01) while parking is denoted with letter P (e.g. P02). Foundation is 00F, ground floor is 000, mezzanine floors are denoted with letter M (i.e., M00), and above-ground floors are 002, 003, etc.
9. uf\_level\_1: Major categories of construction information separated by their special function, according to UniFormat (CSI and CSC, 2010) (e.g., Substructure, Shell and Interiors).
10. uf\_level\_2: UniFormat Level 1 categories divided into classes by separating the categories into the discrete concepts that compose them (e.g., Foundations, Slabs-on-Grade).
11. uf\_level\_3: Subdivisions of UniFormat Level 2 classes, (e.g., Standard Foundations, Special Foundations).
12. uf\_level\_4: Subdivisions of UniFormat Level 3 classes, (e.g., Wall Foundations, Column Foundations, Standard Foundation Supplementary Components).
13. uf\_level\_5: Additional elemental information that provides further details on certain structural elements (e.g., joists, beams, trusses) in some of the Substructure (i.e., Wall Foundations) and Shell categories (i.e., Floor Structural Frame, Roof Structural Frame). This variable is optional. For full list, please refer to Table 3.
14. mf\_level\_1: Divisions of a building by the related work results (i.e., Construction result achieved in the production stage and identified by one or more of the following: the particular skill or trade involved; the construction resources used) according to MasterFormat (CSI and CSC, 2016).
15. mf\_level\_2: Subdivisions of MasterFormat Level 1 (mf\_level\_1) titles, this variable is optional.
16. mf\_level\_3: Subdivisions of MasterFormat Level 2 (mf\_level\_2) titles, this variable is optional.
17. mf\_level\_4: Subdivisions of MasterFormat Level 3 (mf\_level\_3) titles, this variable is optional.
18. mf\_level\_5: Subdivisions of MasterFormat Level 4 (mf\_level\_4) titles, this variable is optional.
19. Unit: Amount of construction material calculated via material takeoff. Expressed in terms of mass (i.e. kg) or volume (i.e. m3).
20. uncertainty\_score: Communicates the uncertainty of the data sources used in the quantification process of the material on a scale of 1 to 6. The pedigree matrix originally developed by (Weidema & Wesnaes, 1996) is adapted (Table 4) to describe the uncertainty of the data sources used in the material quantification process.
21. quantity\_1: Two quantities (quantity\_1 and quantity\_2) are reported for each material take off, a minimum and a maximum. If the calculation of the takeoff is performed within a range of minimum and maximum, the minimum value must be written under the quantity\_1 column and the maximum value must be written under the quantity\_2 column. If the result of the takeoff is a discrete number, then the same amount must be entered under the quantity\_1 and quantity\_2 columns.
22. quantity\_2: Two quantities (quantity\_1 and quantity\_2) are reported for each material take off, a minimum and a maximum. If the calculation of the takeoff is performed within a range of minimum and maximum, the minimum value must be written under the quantity\_1 column and the maximum value must be written under the quantity\_2 column. If the result of the takeoff is a discrete number, then the same amount must be entered under the quantity\_1 and quantity\_2 columns.
23. In addition to the variables above, the gross floor area (GFA) of each building should be reported. The GFA is described as the total floor area of all floor levels including underground space and the area taken by external walls, internal walls, columns, and partitions. The unit of measurement for the GFA is square meters.

The data descriptor article submitted to the journal Scientific Data (Guven et al. XXXX) provides details on the method and data structure of the database.

# 3. Guidelines for contributing to the database

The database contains the variables that are described in the “Variables” section below. For contributing to this data repository, please follow these guidelines:

* Data must be provided for all variables that are not defined as optional.
* The amount of construction material that is calculated via material takeoff must be expressed in terms of mass (i.e. kg) or volume (i.e. m3). The amount must be entered to the database under two columns: Quantity 1 and Quantity 2. If the calculation of the takeoff is performed within a range of minimum and maximum, the minimum value must be written under the Quantity 1 column and the maximum value must be written under the Quantity 2 column. If the result of the takeoff is a discrete number, then the same amount must be entered under the Quantity 1 and Quantity 2 columns.
* Contributors must identify themselves in the contributor column and should adhere to the given data format.
* Contributors to add new data to the dataset should create copies of the repository (i.e. forks) and update the dataset this way.

# 4. Tables

*Table 1. List of construction stages for documents*

|  |  |
| --- | --- |
| Quality / Stage of Data | Code |
| Concept sketch design | 00CSD |
| <33% design development | L33DD |
| 33% design development | G33DD |
| 50% design development | 050DD |
| 66% design development | 066DD |
| 95% design development | 095DD |
| Issued for Tender | 00IFT |
| Issued for Building Permit | 0IFBP |
| Issued for Tender and Building Permit | IFTBP |
| Issued for Construction | 00IFC |
| Issued as Record Drawings | 0IARC |

*Table 2. List of building types*

|  |  |
| --- | --- |
| Building Type | Code |
| Apartment building | APB |
| Institutional | INS |
| Office | OFF |
| Educational | EDU |
| Single detached | SND |
| Semi detached | SMD |
| Detached accessory dwelling units | ADU |
| Secondary units | SEC |
| Townhouse | TWN |
| Mixed Use | MIX |
| Laneway Suites | LNW |
| Renovated single detached | SNR |
| Renovated semi detached | SMR |

*Table 3. Additions to UniFormat*

|  |  |  |
| --- | --- | --- |
| Additions to UniFormat | Levels | Code |
| Substructure Interior | Level 2 | A50 |
| Floor Construction | Level 3 | A5010 |
| Interior Partitions | Level 3 | A5020 |
| Ceiling Finishes | Level 3 | A5030 |
| Floor Structural Frame - A50 | Level 4 | A5010.10 |
| Floor Decks, Slabs, and Toppings - A50 | Level 4 | A5010.20 |
| Ramps - A50 | Level 4 | A5010.30 |
| Interior Fixed Partitions - A50 | Level 4 | A5020.10 |
| Interior Partition Supplementary Components - A50 | Level 4 | A5020.20 |
| Plaster and Gypsum Board Finish - A50 | Level 4 | A5030.10 |

*Table 3. List of UniFormat Level 5 elements*

|  |  |
| --- | --- |
| Name of element | Code |
| Continuous Footings | 0CF |
| Foundation Walls | 0FW |
| Spread Footings | 0SF |
| Column Piers | 0CP |
| Columns Supporting Floors | CSF |
| Floor Girders and Beams | FGB |
| Floor Trusses | 0FT |
| Floor Joists | 0FJ |
| Columns Supporting Roofs | CSR |
| Roof Girders and Beams | RGB |
| Roof Trusses | 0RT |
| Roof Joists | 0RJ |
| Parking Bumpers | 0PB |
| Precast Concrete Stair Treads | PCS |
| Roof Curbs | 0RC |
| Exterior Wall Construction | EWC |
| Composite Decking | CPD |
| Cast-in-Place concrete | CIC |
| Floor Structural Frame | FSF |
| Associated Metal Fabrications | AMF |
| Floor Construction Supplementary Components | FCS |
| Roof Construction Supplementary Components | RCS |
| Residential Elevators | 0RE |
| Vegetated Low-Slope Roofing | VLR |
| Swimming Pools | SWP |
| Excavation Soil Anchors | ESA |
| Roof Window and Skylight Performance | RWS |
| Rainwater Storage Tanks | RST |
| Gray Water Tanks | GWT |

*Table 4. Uncertainty according to the reliability of the data sources used for quantification of building materials (definitions adapted from Weidema and Wesnaes, 1996).*

|  |  |  |
| --- | --- | --- |
| Pedigree matrix indicator scores | Reliability indicator pedigree matrix definition | Definition adapted to material quantification process |
|
| 1 | Verified data based on measurements | Material quantification based on measurements performed on site, reflecting as-built conditions |
| 2 | Verified data based on measurements | Material quantification based on building drawings and details (e.g. specifications, notes, legends) |
| 3 | Verified data partly based on assumptions or non-verified data based on measurements | Material quantification based on trusted references for information not included in the building drawings (e.g. local and/or national building codes, brochures, literature) |
| 4 | Non-verified data partly based on qualified estimates | Material quantification based on consultation with industry experts and estimators for information not included in the building drawings |
| 5 | Qualified estimates (e.g. by industrial party) | Material quantification based on proxy data for information not included in the building drawings |
| 6 | Non-qualified estimate | N/A |

# 4. Acknowledgements

Below is the names of people contributed to the creation of the original dataset and the ontology published in (Guven et al. 2021):

* Aldrick Arceo
* Allison Bennett
* Alex W. Olson
* Batool S. Zehra
* Bolaji Olanrewaju
* Charu Tyagi
* Gursans Guven
* Kaan Isin
* Melanie Tham
* Molly McGrail
* Rose Zhang
* Salim Hourieh
* Shoshanna Saxe

The paper should be cited as: Guven, G., Arceo, A., Bennett, A., Tham, M., Olanrewaju, B., McGrail, M., Isin, K., Olson, A.W., and Saxe, S. (2021). “A Construction Classification System Database for Understanding Resource Use in Building Construction”, submitted October XX, 2021 to Scientific Data, Nature.

# 5. References

Canadian Construction Association. (2012). *Guide to Cost Predictability in Construction : An analysis of issues affecting the accuracy of construction cost estimates*. (1), 1–26.

CSI and CSC. (2010). *UniFormat - A Uniform Classification of Construction Systems and Assemblies*. Constructions Specification Institute (CIS) and Construction Specifications Canada (CSC).

CSI and CSC. (2016). *MasterFormat Numbers & Titles* (pp. 1–186). pp. 1–186. Constructions Specification Institute (CIS) and Construction Specifications Canada (CSC).

Heeren, N., & Fishman, T. (2019). A database seed for a community-driven material intensity research platform. *Scientific Data*, 1–10. https://doi.org/10.1038/s41597-019-0021-x

International Organization for Standardization (ISO). (2021). ISO Online Browsing Platform. Retrieved from Country codes website: https://www.iso.org/obp/ui/#search

Weidema, B. P., & Wesnaes, M. S. (1996). Data quality management for life cycle inventories-an example of using data quality indicators. *Journal of Cleaner Production*, *4*(3–4), 167–174.