Conversion of Lithological Data in the Manitoba Water Well Database (GWDrill) to a Mappable Format

By L. H. Thorleifson[1] and D. M. Pyne[2]

[1]Geological Survey of Canada
Current address: Minnesota Geological Survey
2642 University Ave W
St Paul, MN 55114-1057
Telephone: 612-627-4780 ext 224, Fax: 612-627-4778
e-mail: thorleif@umn.edu

[2]Geological Survey of Canada
601 Booth Street
Ottawa ON K1A 0E8

ABSTRACT

A project designed to gain access to all usable data to aid construction of 3D geological models for southern Manitoba has included significant effort to appropriately utilize lithological data of variable quality in the Provincial water well database known as "GWDrill". In order to use the data, it was necessary first to assign x, y, and z coordinates to each site on the basis of the existing designation, usually a quarter section (one-half mile by one-half mile) or river lot location. Secondly, it was necessary to convert the lithological data to a classification and terminology that could be queried and mapped. The lithological data were converted by correcting spelling, obtaining an inventory of words, deleting unusable words, identifying synonyms and changing them to a single term, and parsing and interpreting the remaining information into several lithological, colour, structure, consistency, hydrogeological, and stratigraphic variables. The resulting database has exceeded expectations with respect to apparent location accuracy and geological coherence.

INTRODUCTION

As a result of the foresight of Manitoba Water Resources Branch staff, the Province of Manitoba has an excellent digital database of water well records, known as GWDrill. The database includes several tables, including a table of verbal stratum descriptions. These lithological data typically appear unmodified from the driller's reports, providing very useful site-specific information in the driller's original wording.

Progress in application of computer-based mapping methods has, however, increased recognition that data would ideally also be formatted for use in these applications. In doing so, caution is required because data originally reported with a low level of certainty may be formatted for computer-based applications in a manner that implies a higher level of certainty than was originally intended. One approach to this challenge is to set up procedures that allow many drillholes to be viewed at once, allowing anomalous data to become more apparent. Doing so requires that sites be assigned x, y, and z coordinates so that sites can be graphically portrayed relative to each other, and that lithological data of varying format and terminology be translated to a set of attributes under which consistent, although evolving, terminology is used.

This paper describes an experiment in the implementation of this approach to the Manitoba water well database. The objective was to permit lithological data to be queried and mapped in 3D, thus aiding the construction of geological models for southern Manitoba.

LOCATION AND ELEVATION

Virtually all of the records in the database are located to a legal survey polygon such as a quarter section or a river lot, so a database of x and y locations for centroids of the legal survey polygons was required. Two formats are present: Dominion Land Survey (DLS) quarter sections, and Parishes, consisting of river lots and other survey systems that predate the DLS grid. Legal survey data obtained from Manitoba Conservation were used to obtain x and y values for the centroid of each polygon. Sites located to quarter sections were assigned the quarter section coordinates. Wells located to a section (1 mile by 1 mile) were assigned a location at the centroid of the section, obtained by averaging the quarter section xy values. Wells assigned locations no more specifically than section, for example to a township (6 miles by 6 miles), were not as-

signed coordinates for estimated location. Wells assigned to a river lot were given coordinates for the centroid of the lot. Where river lots are divided into two or more portions, such as inner and outer lots divided by a road, the average of the coordinates was used for a well whose legal survey location did not indicate which portion of the lot the well was located in.

Elevations were assigned by intersecting the legal survey polygon centroid x and y values with a digital elevation model derived from legal survey data that typically is accurate to +/- 3m.

PREPARATION OF THE DATA

Processing was first applied to a November 1998 version of the database that consisted of 83,597 sites (fig. 1) and 402,461 strata. Subsequently, the database was updated with a July 2000 version consisting of 87,992 sites and 422,917 strata. Ten sites that had occurred in the 1998 version had been deleted prior to release of the 2000 version, and 35 strata in the 1998 version do not appear in the 2000 version. Using both versions, a database of 88,002 sites and 422,943 strata resulted.

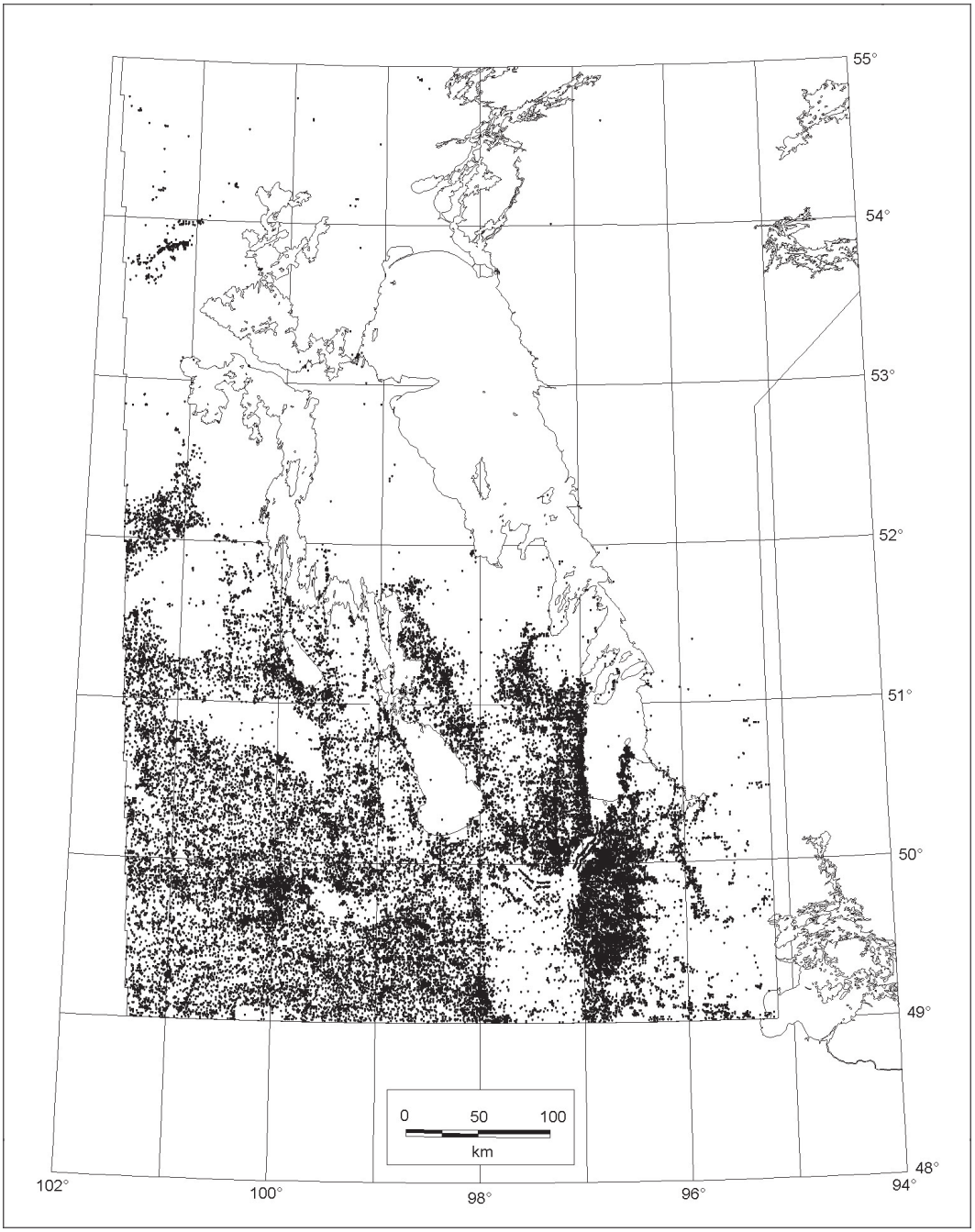

Figure 1. Location of GWDrill sites south of 55°N; 166 additional sites occur farther north.

The database download in 1998 was acquired from Manitoba Water Resources Branch twice, due to media format problems. The first version used memo fields of unrestricted length for the lithologic description, whereas the subsequent version used 255 character text fields, among which 15 records were truncated. The truncated records were restored after a method was found to read the first download. In addition, the first version contained both UTM zones 14 and 15, whereas the second, which was implemented, exclusively used zone 14.

It was agreed by all involved that no changes would be made to the driller's original lithological description as it appeared in the database, and Geological Survey of Canada (GSC) staff have no access to the permanent data archive. The intention was to generate a table of parsed information that could be linked to the original stratum file, with no change in the number of strata.

In addition to this table, however, an add-on table of zero-thickness strata was created to capture information regarding the nature of material at a lower contact. This information typically takes the form of a statement such as: 20 to 30' sand, rock at 30'. The new zero-thickness stratum was given an incremented sequence number, and the original description was modified in the add-on table, for example, sequence x, 20-30' sand, sequence x+1, 30-30', rock. These data were considered important for mapping the bedrock surface, as well as the top of Precambrian rock where it underlies Phanerozoic cover.

Records for the 88,002 sites include descriptions for 422,943 strata. An average of about 3 words is used per stratum, hence a total of about 1.2 million words occur in the original data. These words are arranged in 77,457 different combinations. Conversion of the lithological data commenced by correcting spelling errors using a word-processor spell-check of a copy of the descriptions. Having done so, the number of unique stratum descriptions was reduced to a total of 72,467. These records are constructed from 1,956 different words, excluding numbers and punctuation.

A table of these words was obtained by inserting a carriage return in a copy of the descriptions after all words, and by using database procedures to obtain a word inventory with frequency of occurrence. Each of the 1,956 words was classified with respect to whether the word was to be deleted, substituted with a synonym, or retained, and all words were classified with respect to a list of topics that emerged as the analysis progressed. These included lithological, colour, structure, consistency, hydrogeological, and stratigraphic variables.

Over half of the words were deleted, including terms such as 'and', 'with', and 'of' (table 1) that were of use in a phrase, but are not essential in parsed information. It should be noted that several thousand strata were re-interpreted in the late stages of processing on the basis of searches and inspection of the original description, and at this stage, the word 'and' was a factor. Hence some words that were deleted for the first iteration may later have been considered during a final iteration of interpretation.

Over one-third of the words were synonyms, and were changed to common terms. Among terminology related to lithology (table 2), the words most commonly replaced were synonyms of 'gravel', including 21 different words. Other words that occurred in many synonym forms are 'previous well', 'fill', and 'fossiliferous'. Some synonym replacements are at equal levels of specificity, such as 'hardpan' and 'till', while other substitutions are from differing levels, such as 'fossiliferous' and 'stromatolitic'.

Many changes were made to terminology related to colour (table 3), in order to convert all colours to Munsell word equivalents. The colour to which the greatest number of changes was made was 'light yellowish brown', including the commonly-occurring words 'buff' and 'beige'.

In the case of words related to structure and consistency (table 4), the words most commonly replaced were synonyms of 'stratified', including 55 different words. Other words that occurred in many synonym forms, or in some cases subordinate classes, are 'hard', 'soft' and 'firm'. Among words related to hydrogeology (table 5), the term with the largest number of synonyms was 'fractured/permeable', including 37 different words. Other words that occurred in many synonym forms, or in some cases subordinate classes, are the term 'saturated' and terms related to effervescence or odour.

A total of 92 words was retained from the original description without modification (table 6). In addition, numerous words related to stratigraphic nomenclature were retained (table 7).

Following consultations with co-operating agencies regarding the scheme, the deletion and synonym substitution procedures were applied to a copy of the spell-checked lithological data. Multiple copies of the result were made, and each copy was assigned to a topic such as lithology, colour, structure, or hydrogeology. In each copy, only words considered relevant to that topic were retained. This was done by substituting an ASCII symbol for each of the eligible words, deleting all letters of the alphabet, and then changing the symbols back to the desired words. The retained words were then either parsed into multiple fields, or interpreted. During this procedure, it often was necessary to refer to the original record and apply judgement based on geological knowledge, including the context of the stratum in a sequence. Location was not, however, considered.

CLASSIFICATION AND TERMINOLOGY

In order to generate a database that could readily be queried and mapped, it was essential to parse the information into a set of variables, under which a minimum number of common terms with no synonyms is used. Synonyms were not allowed, so that a query can be carried out without having to list the multiple words that may

Table 1. Most commonly occurring words that were deleted, with frequency of occurrence.

AND	64758	AS	189	CHANGING	90
WITH	15199	HEAVY	189	MATRIX	90
OF	11704	LIKE	179	FINER	88
FEET	10715	RIVER	173	DRILL	86
AT	8405	MAINLY	171	LESS	85
SOME	5240	CAVING	163	LOST	85
THIN	1688	DUG	162	IGPM	84
SMALL	1684	PURE	162	LOSS	82
ODD	1187	HIGHLY	152	DRILLED	82
IN	1080	SMOOTH	146	TAKING	81
OR	946	AFTER	144	SAME	80
FEW	785	LAKE	140	HIGH	79
SLIGHTLY	762	THAN	137	MODERATELY	79
PIT	733	ZONE	136	AMOUNT	74
LITTLE	608	GRAINS	134	ALMOST	74
A	521	GOOD	132	IT	71
NO	500	THROUGHOUT	132	REPORTED	71
FAIRLY	498	PACKED	129	TURNING	70
TRACES	480	OCCASIONAL	127	SIZE	70
GRAINED	468	MATERIAL	124	LOOKS	69
ROUGH	428	MINOR	120	PROBABLY	69
OVERLAY	403	BASE	119	GETTING	68
LEVEL	399	HOLE	115	DARKER	67
GLACIAL	398	OPEN	115	BIG	66
MORE	373	LACUSTRINE	113	SEMI	65
COLOURED	362	COLOUR	107	POORLY	63
MOSTLY	362	BACK	105	UNKNOWN	63
DEPTH	362	NUMEROUS	105	SURFACE	63
COARSER	345	TOP	102	PLATES	63
BOTTOM	323	MUCH	102	LOSING	63
LEDGES	306	ON	100	PREDOMINATE	63
LARGE	274	BELOW	98	THE	62
ABOVE	266	GPM	96	POSSIBLY	61
ALLUVIAL	242	NICE	95	ETC.	61
LOTS	233	BECOMING	95	SHARP	59
NOT	230	BUT	94	DIAMETER	59
FROM	228	IS	94	OFF	58
INCH	207	LOOSELY	92	BETWEEN	58
UNIFORM	205	BIT	92	DECAY	58
DRILLING	195	INCHES	91	NEAR	57

Table 2. Lithology-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

GRAVEL	11593	stones (5818), pebbles (3249), cobbles (597), fragments (538), pieces (275), particles (269), chips (267), bits (244), cobble (135), pebble (70), granular (34), cobblestones (24), granules (23), cobblestone (16), gravels (11), clasts (7), granulars (7), chip (5), chippings (2), aggregate (1), gravel-cobbles (1)
TILL/DIAMICT	10328	hardpan (10213), till-like (62), tilly (41), colluvium (3), semihardpan (3), tillish (3), tills (2), claybound (1)
SOIL	6630	topsoil (5289), loam (934), sod (274), dirt (85), earth (15), roots (12), vegetation (11), manure (5), turf (3), mould (1), soils (1)
GRAVELLY	6208	stoney (5607), pebbley (577), cobbley (18), stonier (6)
UNDIFFERENTIATED ROCK	4332	bedrock (2531), stone (1748), caprock (51), sedimentary (2)
INTRUSIVE OR HIGH GRADE METAMORPHIC	3789	granite (3343), granitic (333), igneous (90), schist (9), gneiss (8), pegmatite (4), diorite(2)
PREVIOUS-WELL	2369	well (1259), old (667), existing (227), casing (31), previously (29), crib (28), cribbed (26), excavated (24), abandoned (23), cribbing (16), cased (15), previous (8), pipe (8), deepened (4), trench (3), wells (1)
BOULDER(S)	1749	rocks (1743), block (3), boulders-cobbles (2), blocks (1)
SILICA/SILICEOUS	1578	silica (889), quartz (621), siliceous (54), agate (7), agates (3), tripolized (1), quartzose (1), silica-like (1), chalcedony (1)
BENTONITIC SHALE/MUDSTONE	1052	bentonite(1052)
FILL	889	road (364), backfill (291), roadbed (136), concrete (28), ash (23), pavement (9), asphalt (8), roadfill (6), driveway (5), rockfill (3), culvert (3), garbage (2), backfilled (2), sawdust (2), landfill (1), fill-type (1), runway (1), bricks (1), roadway (1), junk (1), dike (1)
EVAPORITES	788	gypsum (752), gyprock (21), anhydrite (15)
CARBONATE	783	lime (763), tyndall (8), carb (6), carbonates (5), calcarenite (1)
DOLOSTONE	756	dolomite(737),dolostone(19)
SAND	549	quicksand (482), beach (22), sands (21), grit (21), alluvium (2), sand-like (1)
LOW GRADE METASEDIMENTARY OR METAVOLCANIC	431	soapstone (223), slate (118), volcanic (33), greenstone (18), quartzite (11), marble (9), basalt (9), metamorphic (7), andesite (1), argillite (1), volcanics (1)
CLEAN	400	washed (249), cleaner (149), cleaned (1), cleanest (1)
FINE-TO MEDIUM	360	fine-medium (253), medium-fine (106), fine-medium-fine(1)
UNCLASSIFIED	297	log (297) *(inherited from 'no log' after 'no' was deleted)*
FOSSILIFEROUS	264	shells (163), shell (51), clam (12), fossils (12), fossil (7), snail (7), bone (3), crinoid (2), brachiopods (2), bones (1), mollusk (1), snails (1), stromatolitic (1), vertebrae (1)
PEAT/ORGANIC	258	muskeg (85), swamp (75), muck (33), moss (30), quagmire (17), swampy (16), peatmoss (2)
SULPHIDE-BEARING	257	pyrite (228), pyritic (12), pyrites (5), sulfide (5), marcasite (2), arsenopyrite (2), sulfides (1), ore (1), metallic (1)
MEDIUM-TO COARSE	238	medium-coarse (232), coarse-medium(6)
SOFT CLAY	236	gumbo(183), putty(53)
FINE-TO COARSE	220	fine-coarse (215), coarse-fine (3), fine-medium-coarse (2)
SILTY CLAY	217	mud(216),silt-clay(1)
UNDIFFERENTIATED SOIL/SEDIMENT	217	overburden (168), drift (19), sediment (18), surficial (7), sediments (4), regoliths (1)
SANDY	193	gritty (184), sandier (5), loamy (4)
WOOD-BEARING	182	wood (179), woody (2), stump (1)
SHALE/MUDSTONE	181	siltstone (105), shale-like (31), shales (28), claystone (7), mudstone (4), quartz-shale (2), clay-shale (2), shalestone (1), diatomaceous (1)
BOULDERY	179	rocky (179)
CLAYEY	154	muddy (113), clayed (25), clayish (14), clay-type (1), clayier (1)
COAL	144	lignite (142), coals (2)
PETROLIFEROUS	144	oily (112), oil (27), petroleum (4), hydrocarbons (1)

Table 2. Continued.

FERRUGINOUS	137	iron (77), oxide (30), hematite (17), oxides (10), ferrous (1), sideritic (1), ironstone (1)
CONCRETIONARY/NODULAR	114	nodules (49), concretions (33), concretion (29), nodular (3)
WELL SORTED/GRADED	71	sorted (48), graded(23)
VOID	70	sinkhole (48), cavern (7), cavity (5), drop (3), cave (3), sink (1), cavernous (1), caves (1), cavities (1)
SILT	67	fines (30), silts (29), flour (8)
CALCAREOUS	51	calcite (29), limey (13), calcium (7), calcitic (1), marly (1)
CHERTY	50	chert (32), flinty (14), flint (4)
MEDIUM	48	pea (47), pea-size (1)
RUBBLE	47	rubbley (26), breccia (20), rubbles (1)
CALCAREOUS SILT	39	marl (39)
AUTHIGENIC-XL-BEARING	38	crystals (24), selenite (12), roses (1), rosettes (1)
MICACEOUS	38	mica (29), biotite (4), biotitic (3), muscovite (1), schistose (1)
SILTY	36	silted (20), siltier (14), silting (2)
CLAY	35	clays (34), soil-clay (1)
CARBONACEOUS/BITUMINOUS	20	tar (10), charcoal (3), coaly (2), bituminous (2), sooty (1), carboniferous (1), tar-like (1)
FELDSPATHIC	12	feldspar (12)
ROUNDED	11	round (9), spherical (2)
ORGANIC	10	peaty (6), plant (2), carbon (1), marshy (1)
DIRTY	9	impure (9)
KAOLINITIC	8	kaolin (5), kaolinized (3)
LIMESTONE	2	ls (1), quasi-limestone(1)
CHALKY CARBONATE	5	chalk(5)
OOLITIC	3	oolites (2), oolite (1)
SANDY SILT	3	sandy-silty (3)
CARBONATE & PRECAMBRIAN-RICH	2	granite-limestone (2)
MEDIUM-TO VERY	2	medium-very(2)
SHALEY	2	shalier (2)
SILTY SAND	2	silt-sand(2)
SUBROUNDED	2	subspherical (2)
ARGILLACEOUS	1	argillitic (1)
CARBONACEOUS/BITUMINOUS SHALE/MUDSTONE	1	coal-shale(1)
FINE-TO VERY	1	fine-very(1)
PRECAMBRIAN-RICH	1	granity (1)
SANDY CLAY	1	silt-sand-clay(1)

Table 3. Colour-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

LIGHT YELLOWISH BROWN	1708	buff (1167), beige (531), blonde (2), buff-brown (2), brownish-buff (2), buff-tan (1), tan-yellow (1), buff-grey (1), beige-grey (1),
VERY-LIGHT BROWN	361	cream (359), whitish-brown (1), cream-white (1)
BROWNISH GREY	347	brown-grey (264), brownish-grey (77), tan-grey (2), grey-brownish (2), grey-cream (1), grey-buff (1)
GREYISH BROWN	336	grey-brown (287), greyish-brown (46), grey-tan (2), brown-greyish (1)
MULTICOLOURED	327	varicoloured (90), specks (73), speckled (68), flecks (36), spots (25), patches (15), speckles (6), mottling (5), marbled (3), spotted (3), variegated (2), mottle (1), light-dark (1)
LIGHT BROWN	326	tan (279), cream-brown (29), brown-white (5), tan-brown (4), white-brown (3), light-brown (3), brown-light (1), light-buff (1), tan-coloured (1)
LIGHT	304	whitish (179), creamy (78), lighter (36), milky (9), light-medium (1), bland (1)
REDDISH	211	rusty (210), rusty-yellow (1)
GREYISH BLACK	179	grey-black (179)
REDDISH BROWN	163	rust (67), reddish-brown (54), red-brown (23), orange-buff (7), brown-red (7), buff-orange (3), rust-brown (1), rust-red (1), reddish-buff (1)
BLUISH GREY	122	blue-grey (89), bluish-grey (32), grey-blue-grey (1)
GREENISH GREY	109	greenish-grey (57), green-grey (50), greenish-grey-white (1), grey-greenish (1)
YELLOWISH BROWN	90	yellow-brown (90)
DARK BROWN	78	chocolate (69), black-brown (7), brown-blackish (2)
GREYISH WHITE	76	grey-white (74), greyish-white (2)
SPECKLED	68	salt (36), pepper (30), salt-pepper (2)
LIGHT GREY	66	white-grey (41), light-grey (13), cream-grey (8), grey-light (2), silver (2), grey-whitish (1)
BLUISH GREEN	57	blue-green (53), bluish-green (4)
PALE PURPLE	51	mauve (51)
DARK	44	blackish (43), blackened (1)
LIGHT GREENISH BROWN	36	khaki (36)
GREYISH GREEN	36	grey-green (29), greyish-green (7)
BRILLIANT	27	bright (26), vivid (1)
BROWNISH BLACK	25	brown-black (24), black-brownish (1)
REDDISH ORANGE	23	orange-brown (11), orange-red (5), reddish-orange (3), red-orange (2), red-yellow (2)
GREENISH BROWN	15	green-brown (13), greenish-brown (1), brown-greenish (1)
DARK GREY	14	black-grey (12), blackish-grey (2)
GREENISH WHITE	14	greenish-white (11), green-white (3)
BROWNISH YELLOW	13	brown-yellow (11), brownish-yellow (2)
GREYISH BLUE	11	grey-blue (11)
BROWNISH	10	brick (10)
VERY-LIGHT GREY	10	whitish-grey (10)
REDDISH GREY	10	reddish-grey (8), red-grey (2)
WHITE	9	off-white (8), white-light (1)
BLUISH BLACK	8	blue-black (7), bluish-black (1)
LIGHT RED	7	rose (7)
LIGHT GREENISH BLUE	7	turquoise (7)
YELLOWISH GREY	7	yellow-grey (6), yellowish-grey (1)
GREENISH BLUE	6	green-blue (5), blue-greenish (1)
RED	5	ochre (5)
BROWNISH BLUE	4	brown-blue (4)

Table 3. Continued.

BROWNISH GREEN	4	brown-green (4)
GREY	4	greyer (3), pearly (1)
GREENISH BLACK	4	green-black (3), greenish-black (1)
PURPLISH GREY	4	purplish-grey (2), grey-purple (1), purple-grey (1)
GREENISH YELLOW	3	greenish-yellow (3)
BRILLIANT YELLOW	3	gold (3)
LIGHT PINKISH GREY	3	pink-grey (3)
MULTICOLOURED BLUE	3	yellow-blue (3)
PALE YELLOW	3	yellow-white (3)
PALE	3	faint (3)
BROWNISH RED	3	brownish-red (3)
GREYISH RED	3	greyish-red (2), grey-red (1)
BROWN	3	drab (2), brown-brownish (1)
YELLOWISH WHITE	3	white-yellow (2), yellowish-white (1)
YELLOWISH RED	3	yellow-rusty (2), yellow-red (1)
MULTICOLOURED GREY	3	grey-white-black (2), blue-grey-green (1)
DARK BLUE	3	blue-dark (1), black-blue (1), blackish-blue (1)
MULTICOLOURED RED	3	blue-green-reddish (1), black-red-white (1), red-blue (1)
PINKISH WHITE	3	white-reddish (1), white-pink (1), white-pink (1)
MULTICOLOURED ORANGE	2	orange-grey (2)
STRONG	2	blood (2)
LIGHT GREYISH	2	silvery (2)
LIGHT PINKISH BROWN	2	pink-beige (2)
GREEN	2	greener (2)
MULTICOLOURED BROWN	2	white-brown-black (1), grey-white-brown (1)
PURPLISH RED	2	purplish-red (1), purple-red (1)
DARK GREENISH BLUE	1	teal (1)
GREYISH YELLOW	1	grey-yellow (1)
YELLOWISH GREEN	1	yellow-green (1)
YELLOWISH	1	golden (1)
VERY-LIGHT YELLOW	1	whitish-yellow (1)
OLIVE BROWN	1	brown-olive (1)
DARK YELLOW	1	mustard (1)
LIGHT REDDISH BROWN	1	red-cream (1)
PINKISH	1	flesh (1)
BRILLIANT RED	1	scarlet (1)
PINKISH BROWN	1	brown-pink (1)
PINKISH GREY	1	grey-pinkish (1)
BLUISH BROWN	1	Blue-brown (1)
BROWNISH PURPLE	1	brownish-purple (1)
LIGHT YELLOWISH RED	1	beige-red (1)
SALT-AND-PEPPER	1	peppered (1)
LIGHT PINK	1	pink-white (1)
LIGHT ORANGE	1	orange-white (1)
LIGHT GREYISH YELLOW	1	grey-beige (1)
LIGHT REDDISH	1	light-reddish (1)
LIGHT BROWNISH	1	light-brownish (1)
ORANGE	1	orange-yellow (1)
PURPLISH BROWN	1	purplish-brown (1)

Table 4. Structure and consistency-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

STRATIFIED	28496	layers (17473), layer (3972), stringers (1147), streaks (1095), lenses (757), layered (733), mixed (540), trace (532), seams (324), mixture (260), pockets (238), mix (192), stringer (153), interbedded (151), seam (137), zones (113), streak (79), bed (67), lens (64), laminated (61), strips (45), interlayered (36), beds (36), laminations (28), alternating (27), veins (25), varved (23), inclusions (21), strip (19), bands (17), alternate (16), pocket (14), bedded (11), sections (10), interbeds (9), partings (7), layering (7), streaked (6), intermittent (6), intervals (5), vein (5), strata (5), horizons (5), intermixed (4), stripes (4), pods (3), band (3), banding (2), horizon (2), interval (2), lensing (1), intermingled (1), veinlets (1), inclusion (1), intermittently (1)
HARD	2111	solid (1273), dense (398), consolidated (191), harder (130), cement (47), compact (38), compacted (21), denser (6), cementing (2), cementation (1), hard-medium (1)
VERY SOFT	1059	sticky (974), soupy (36), tacky (22), mucky (13), gooey (9), gummy (4), soup-like(1)
SOFT	464	softer (230), greasy (203), clay-like (21), putty-like (7), mushy (2), semiplastic (1)
PLASTIC	347	slick (152), plasticity (146), slippery (49)
FIRM	338	firmer (224), waxy (72), rubber (22), cohesive (4), semistiff (4), semisoft (4), semisolid (3), pliable (2), rubbery (1), semi-firm (1), firmness (1)
WEATHERED	321	rotten (320), weathering (1)
VERY HARD	117	brittle (117)
UNOXIDIZED	85	nonoxidized (85)
FRIABLE	83	crumbley (74), crumbled (3), shattery (2), crumbles (2), crumbling (1), semiconsolidated (1)
OXIDIZED	19	oxidization (10), oxidation (9)
STIFF	17	stiffer (9), semihard (5), medium-hard (3)
UNWEATHERED	15	fresh (15)
MASSIVE	15	homogeneous (15)
LOOSE	4	uncemented (4)

Table 5. Hydrogeology-related synonyms that were changed to common terms, with frequency of occurrence in parentheses.

SATURATED	9559	water (7784), wet (772), water-bearing (415), moist (349), damp (93), flowing (53), frozen (31), watertable (31), ice (9), aquifer (9), moisture (2), frost (2), flows (2), artesian (2), watery (2), flowed (1), waterlogged (1), permafrost (1)
FRACTURED/PERMEABLE	7740	broken (5971), fractures (482), fracture (392), shattered (233), circulation (156), crushed (130), seepage (87), stream (77), porous (56), cracks (37), crevices (15), cracked (14), flakey (11), crack (11), fissures (11), fragmental (9), vuggy (7), permeable (6), brecciated (4), fissile (4), spongy (3), fracturing (2), fissured (2), blocky (2), fragmented (2), crevice (2), porosity (2), pores (2), microfractures (2), pervious (1), honeycomb (1), honeycombed (1), spally (1), platy (1), splintery (1), fissure (1), fault (1)
UNSATURATED	1361	dry (1361)
SALINE	945	salty (945)
EFFERVESCENT/GASSY	108	gas (40), smell (40), sulfur (10), gaseous (6), smelling (4), putrid (2), egg (2), bubbly (1), carbonated (1), methane (1), sulfurous (1)
TIGHT/IMPERMEABLE	82	tight (74), impervious (4), bonded (3), fractureless (1)

Table 6. Words that were retained, with frequency of occurrence.

CLAY	124039	BROWNISH	1124	CHERTY	9
SAND	87191	REDDISH	938	PREVIOUS	8
TILL	85317	CALCAREOUS	645	UNWEATHERED	5
GREY	80416	WEATHERED	592	OOLITIC	3
BROWN	59053	OLIVE	538	VOID	3
GRAVEL	45573	YELLOWISH	511	NODULAR	3
BLUE	29241	STIFF	491	GLAUCONITIC	3
FINE	25071	ORGANIC	481	FELDSPATHIC	3
SILTY	18439	CEMENTED	432	COALY	2
YELLOW	18049	ORANGE	420	ARKOSIC	2
SANDY	15483	MOTTLED	396	CONGLOMERATIC	1
SOFT	15445	GREENISH	377	PETROLIFEROUS	1
HARD	13175	PINK	358		
COARSE	11920	BOULDERY	287		
MEDIUM	10689	PLASTIC	287		
WHITE	10222	BLUISH	275		
LIGHT	9055	UNOXIDIZED	229		
SILT	8515	NONCALCAREOUS	221		
RED	7992	CARBONACEOUS	195		
DARK	7846	ROUNDED	185		
TO	7708	PURPLE	157		
FRACTURED	6300	PALE	134		
BLACK	5849	ARGILLACEOUS	126		
CLAYEY	5344	ANGULAR	108		
FIRM	4115	PINKISH	86		
SANDSTONE	4000	MICACEOUS	64		
GRAVELLY	3672	STRATIFIED	37		
RUBBLE	2694	DOLOMITIC	36		
OXIDIZED	2456	SUBROUNDED	30		
CLEAN	2271	GYPSIFEROUS	26		
GREYISH	2074	BENTONITIC	24		
GREEN	2073	CHALKY	24		
DIRTY	1975	PURPLISH	24		
FILL	1939	SUBANGULAR	22		
SOIL	1694	FOSSILIFEROUS	21		
SHALEY	1640	SALINE	20		
COAL	1355	MASSIVE	18		
WELL	1259	KAOLINITIC	17		
CARBONATE	1251	FRIABLE	14		
LOOSE	1223	ARENACEOUS	10		

Table 7. Stratigraphic terminology occurring in the database, with frequency of occurrence.

ODANAH MEMBER	917	RESTON FORMATION	3
MILLWOOD MEMBER	481	MORDEN SHALE	2
SWAN RIVER FORMATION	108	COULTER MEMBER	2
BOISSEVAIN FORMATION	51	GUNTON MEMBER	2
PIERRE SHALE	35	WILLIAMS MEMBER	2
TURTLE MOUNTAIN FORMATION	19	WINNIPEGOSIS FORMATION	2
ASHVILLE FORMATION	9	WINNIPEGOSIS FORMATION	2
RED RIVER FORMATION	6	UPPER FORT GARRY MEMBER	1
STONY MOUNTAIN FORMATION	5	SELKIRK MEMBER	1
WINNIPEG FORMATION	5	STONEWALL FORMATION	1
FAVEL FORMATION	5	BLAIRMORE FORMATION	1
DAWSON BAY FORMATION	4	UPPER RED RIVER FORMATION	1
VERMILLION RIVER FORMATION	4	ASHERN FORMATION	1
RIDING MOUNTAIN FORMATION	3	AMARANTH FORMATION	1
		INTERLAKE GROUP	1

have been applied to a single entity. Many classification schemes have been published, but the approach taken here was to build a classification that could capture as much information as possible from the database as it stands. To some extent, additional terminology was added to accommodate potential future usage, but this was limited in order to avoid an unnecessarily complicated system with terms that might never be used. It should be noted however, that the terminology could readily be extended to accommodate geology in other areas, or increasing geological sophistication in the drilling community.

All information herein makes reference to a stratum intersected by a drill hole. The entities in the database therefore are identified on the basis of a site identifier, a stratum sequence number starting at 1 at the top, a depth to top of the stratum, and a depth to the stratum's lower contact (table 8). A distinction is then made whether the interval has been described or not, whether it is a previous well, or whether it is a geological material, either sediment or rock. If it is a sediment or rock that was described, a basic description was parsed into five variables (table 9). If more detailed information is provided, typically by a geologist, this more detailed information was placed into supplementary tables that provide an extended description of the geology (tables 7 and 10 to 15).

Rules of precedence were required, and typically it was only the first piece of information on a topic in a stratum description that was parsed under the relevant attribute. The other information after the first word can now only be retrieved by searching the original descriptions or, preferably, the spell-checked original description. An exception to this rule is some cases of composite adjectives, where it is the final adjective that may take precedence. There also are cases where a composite noun is modified by a composite adjective, and in this case there is a tendency for precedence to be given to the first noun, but sometimes the first, but elsewhere the second, adjective. Also, judgment was used to identify the most noteworthy of multiple observations with respect to geology.

Some compound nouns were parsed as a noun and a modifier, where geological knowledge implies that the two entities were blended in a homogeneous manner. Other compound nouns imply to a geologist a heterogeneity that would have resulted in the two nouns being parsed to the 'lithology' and the 'interbedded with' fields. For example, 'silt and clay' implies homogeneity to a geologist, where 'limestone and shale' implies heterogeneity.

MATERIAL

The most general level of stratum classification is the material. From driller's records, we classify the materials as soil/sediment, rock, ice, water, unclassified, or as a pre-existing well (table 9). The term soil/sediment accommodates the engineering view that anything above bedrock is soil, as well as the geologists' term 'sediment', including deposits described as unconsolidated, unlithified, drift, sediment, surficial deposits, or overburden. 'Rock' encompasses all hard rock as well as in situ weathered rock, excluding boulders. This includes deposits referred to as 'bedrock', 'caprock', and 'stone'. 'Ice' and 'water' were added to the table of options in anticipation of data from lakes, frozen lakes, or glaciers. Unclassified intervals are those for which no information was available, such as blanks where no sample was recovered from the borehole, or where the driller recorded 'pit', 'open pit', 'no log', or an adjective with no noun. Previous wells that were being deepened or re-drilled were described using terminology such as 'old well', 'well pit', 'existing well', 'dug well', 'previously drilled', 'cement crib', 'cribbing', or 'casing'.

Table 8. Designation of the stratum.

Site ID	Unit #	From depth	To depth	Material
				unclassified previous-well ice water soil/sediment rock

Table 9. Basic rock or sediment description.

Principal colour	Textural modifier	Texture	Lithology	Interbedded with
black	clayey	clay	fill	fill
blue	silty	silt	topsoil	topsoil
brown	sandy	sand	peat/organic	peat/organic
green	gravelly	very-fine	silt/clay/fines/mud	silt/clay/fines/mud
grey	bouldery	fine	sand	sand
multicoloured	very-fine	medium	gravel	gravel
olive	fine	coarse	boulder(s)	boulder(s)
orange	medium	very-coarse	till/diamict	till/diamict
pink	coarse	small	soil/sediment	soil/sediment
purple	very-coarse	medium	coal	coal
red		large	evaporites	evaporites
white		very-large	carbonate	carbonate
yellow			shale/mudstone	shale/mudstone
			sandstone	sandstone
			conglomerate	conglomerate
			igneous/metamorphic	igneous/ metamorphic
			rock	rock
			void	void

LITHOLOGY

Following designation of the general material type, as much additional information as possible was captured for every stratum designated soil/sediment or rock. The first two distinct lithological entities described for each stratum were assigned to the 'lithology' and 'interbedded with' attributes (table 9).

Sediments referred to as fill included such descriptions as 'backfill', 'roadbed', 'concrete', 'dyke fill', 'pavement', and 'driveway'. These are human-built deposits of limited thickness at land surface.

Deposits referred to by the driller as topsoil include references to 'soil' at surface, and typically are the thin, organic-rich A-horizon deposits, including 'topsoil', 'loam', 'sod', 'dirt', and 'black earth'. These occurrences are distinct from the engineering usage of the term 'soil', where soil refers to everything above bedrock. The latter we call undifferentiated soil/sediment. Organic sediments we refer to as peat/organic include those described as 'peat', 'muskeg', 'swamp decay', 'organic', 'quag', and 'swamp'.

Driller's descriptions of clay or silt were lumped together as fines, also known as mud in geological terminology. This was done in recognition of the likelihood that silt deposits, especially fine silt, would normally be described as clay, even by trained geologists. Designation as clay or silt, however, is retained in the texture variable (also in table 9). Sediments referred to as sand include those described as 'sand and gravel', on the basis of our opinion that sand and gravel deposits consist mostly of sand, and therefore were classified as gravelly sand. Sediments that we refer to as gravel occur in many original descriptions as various synonyms, as already discussed. The term 'boulder(s)' was retained, although it could be argued that these intervals could be labelled 'rock', while still using the word 'bouldery' in an adjective field. The term 'rubble' is commonly used, but for the sake of simplicity, was changed to 'angular boulders'.

The lithological theme that required the most judgment was selection of driller's descriptions that are regarded by the authors as likely to be glacial diamicts, in most cases deposited as till by glacial ice. Diamicts occurring as diamictites or diamictons are relatively unsorted sediments containing a range of grain sizes from clay to gravel. In the database, the term till/diamict was used to broaden recognition of the term, although other terms exist in the formal scientific and technical literature, such as matrix-supported gravel, morainal material, and loam. Strata coded as till/diamict included those described as 'clay and stones', 'clay and boulders', 'clay and rocks', 'gravel and clay', 'hardpan', and 'boulder clay'. In other words, any stratum that included reference to both clay and gravel was designated as till. Although many till deposits in the region are very silty, descriptions of gravelly silt were not changed to till on the basis of the expectation that a driller who uses the word silt likely possesses, at least in the majority of cases, a relatively high level of geological sophistication and therefore is more likely to be distinguishing an actual silt deposit, perhaps with thin gravelly interbeds, from a till. However, even experienced geologists frequently have difficulty distinguishing till from poorly sorted silty deposits in drillhole samples, and doing so from washed cuttings and drilling rate presumably is also difficult.

Actual interpretation of the driller's description, and hence conversion to something that was not said by the driller was strictly avoided. It could even be argued that no interpretation was made, and reformatting was limited to synonym substitution, if it is accepted that 'clay with stones' is a synonym of 'till'. For example, clay interbedded with limestone was left as such, rather than intervening to designate the limestone a boulder, and translating the description to something like carbonate-rich bouldery till. This was done in recognition that a deposit described as clay interbedded with limestone could be soft shale interbedded with limestone, and designation as till would have been significantly in error. Furthermore, sand was coded as sand, even when interbedded with sandstone and/or shale, and clay was left as clay, even if interbedded with sandstone and/or shale.

Rocks that we coded as coal were described by the driller as either 'coal' or 'lignite'. In a manner similar to the line of reasoning discussed above with respect to silt and clay, it was our opinion that anhydrite deposits, if present, would likely have been referred to as gypsum, so the term 'evaporites' was used for deposits described as 'gypsum', 'anhydrite', or 'gyprock'. Similarly, for carbonates, it was assumed that many strata described as limestones are in fact dolomites, so the term 'carbonate' was used, although the driller's specification of limestone or dolostone was retained as carbonate modifier.

Shales, mudstones, siltstones, and claystones were lumped under the term 'shale/mudstone'. These included those described as 'shale', 'bentonite', 'soapstone', and 'slate', although the latter two terms were included among unmetamorphosed rocks only after inspection of the context. Rocks referred to as sandstone are in nearly every case described as such in the original data, but also include 'silica sand' where included in a rock sequence. The term 'conglomerate' is rarely reported in the region, but was retained.

Also occurring in rock sequences are voids, which are reported as 'cavity', 'sinkhole', 'cavern', 'void', 'cave', 'open cavity', 'free fall', 'empty space', or 'large crevice'. This is not unexpected in the karst terrain found in the region.

All igneous and metamorphic rocks were lumped in one category at the level of basic description, including rocks described as 'granite', 'igneous bedrock', or 'Precambrian rock'. Distinction of intrusive and high-grade-

metamorphic rocks, such as granite, gneiss and schist, from low-grade-metasedimentary and metavolcanic, such as greenstones, was retained in another field.

Where two distinct lithologies were indicated by words such as ‘and’ or ‘layers’, the second named lithology was placed in the ‘interbedded with’ variable, although no provision was made for description of the interbedded material. A description such as ‘till and sand’, for example, was parsed as till interbedded with sand, and the till was given no textural modifier related to the interbedded sediment.

TEXTURE

Adjectives that provide information about grain size were divided into three variables. The principal modifier, here referred to as ‘Texture’ (table 9) includes three groups of modifiers. The words ‘clay’, ‘silt’, and ‘sand’ are intended to be modifiers of fines and till/diamict, as in ‘silt diamict’. The terms ‘very fine’, ‘fine’, ‘medium’, ‘coarse’, and ‘very coarse’ are primarily intended as modifiers for sand, but also can be used for silt and gravel. The terms ‘small’, ‘medium’, ‘large’, and ‘very large’ also were included in this variable for possible future use with respect to boulders.

An agreed-upon set of definitions for these terms would ideally be adopted for the future, but it is recognized that the existing database likely was built using several slightly differing schemes. For example, the exact definition of clay and sand, with respect to quantitative grain size cut-offs, varies slightly among the many classification schemes currently in use.

Also appearing in the basic lithological description (table 9) is a second texture-related adjective, here called the textural modifier. The words ‘clayey’, ‘silty’, ‘sandy’, ‘gravelly’ and ‘bouldery’ are intended to be used as modifiers with or without use of another word under ‘Texture’. For example, ‘silty sand’ is accommodated, although future usage should encourage the dominant sand texture class to be identified, as in ‘silty very fine sand’. Usage such as ‘silty sand diamict’ also is accommodated. The words ranging from ‘very fine’ to ‘very coarse’ also are included here to permit usage such as ‘coarse silt’.

Where boulders were mentioned in a stratum description, this observation took precedence as a textural modifier. In the case of till initially described as some combination of clay and gravel, if the gravel fraction was described as something like an odd pebble, a few pebbles, or a trace of gravel, the bed was classified a clay till/diamict. Deposits described as ‘clay, gravelly’ were simply classed as till with no textural modifier, because any till/diamict deposit has a gravel fraction. In contrast, ‘till, gravelly’ was parsed with a textural modifier, due to the implication that the deposit contains more gravel than the amount adequate for designation as till. Similarly, ‘till, clayey’, ‘till, silty’, or ‘till, sandy’, were parsed with a textural modifier. Multiple textural modifiers were disregarded due to ambiguity, although the information remains available in the original text.

In addition to the above, there is additional textural information in the database related to the range of sand texture, as well as sorting and clast shape. In order to minimize the complexity of the basic description, and because this information typically is only available for strata described by a geologist, the information was placed in a subordinate table referred to as an extended sediment description (table 10).

The terms ‘very fine-to’, ‘fine-to’, ‘medium-to’, ‘coarse-to’ are meant to be used in combination with

Table 10. Extended sediment description.

Weathering	Structure	Consistency	Lithological modifier	Carbonate modifier	Textural range	Sorting	Roundness 1	Roundness 2
oxidized	massive	loose	carb/evap&clstcsedrx-rich	noncalcareous	very-fine-to	well-sorted/ graded	very-angular-to	very-angular
unoxidized	stratified	friable	carb/evap&ig/met-rich	slightly-calcareous	fine-to-	poorly-sorted/ graded	angular-to	angular
		very soft	carb/evap-rich	calcareous	medium-to-		subangular-to	subangular
		soft	clstcsedrx&ig/met rich	strongly-calcareous	coarse-to-		subrounded-to	subrounded
		plastic	clstcsedrx-rich				rounded-to	rounded
		firm	ig/met-rich					well-rounded
		medium stiff	mixed-lithology					
		stiff	very-carb/evap-rich					
		very stiff	very-clstcsedrx-rich					
		hard	very-ig/met-rich					
		very hard	clean					
			dirty					
			authigenic-xl-bearing					
			organic					
			wood-bearing					

'fine', 'medium', 'coarse', and 'very coarse', in order to specify a textural range. This primarily is intended for sand, but also can be used for silt and gravel. Sorting and roundness/angularity rarely are described in the database. Where the information is provided, sorting was categorized as 'well sorted/graded' or 'poorly sorted/graded' (table 10). A range was used for roundness, with the first variable giving the more angular limit of the range, including 'very angular-to', 'angular-to', 'subangular-to', 'subrounded-to', and 'rounded-to', and the second roundness variability giving the more rounded limit, either 'very angular', 'angular', 'subangular', 'subrounded', 'rounded, or 'well rounded'. A full set of terms in the second roundness variable permits one-term descriptions, such as 'very angular gravel'.

SEDIMENT STRUCTURE/COMPOSITION

In many extended descriptions of sediments there is additional information on structure and composition (Table 10). The structure of sediments was classified as either massive or stratified within a defined unit. The term massive is rarely specified, but a large number of words convey the presence of stratification. The consistency of sediment, where described, was classified as either 'loose', 'friable', 'very soft', 'soft', 'plastic', 'firm', 'medium stiff', 'stiff', 'very stiff', 'hard', or 'very hard'.Information on sediment composition includes indications of weathering, whether 'oxidized' or 'unoxidized', and carbonate content, classified as either 'noncalcareous', 'slightly calcareous', 'calcareous', or 'strongly calcareous'.

The lithology of the gravel fraction in sand and gravel or till deposits commonly is reported in well descriptions. A scheme was developed to classify these gravels based on varying proportions of composition: 1) carbonate and evaporite rocks; 2) clastic sedimentary rocks including shale, sandstone, and conglomerate; and 3) igneous and metamorphic rocks. A composition classification is contained in table 10 under the "Lithological modifier". Where one composition class strongly dominates (>~90%), the composition is prefaced by 'very', as in 'very carbonate-rich'. Where there is an indication that one class is dominant (>~50%), that class is named, as in 'carbonate-rich. Where two lithologies are named, implying that the two classes make up more than ~75% but neither of them exceed 50%, a designation such as 'carbonate-and-shale-rich' (which is listed as 'carbonate/evaporite' and 'clastic-sedimentary-rock-rich' in the generic language of Table 10) was used. In cases where each of the three classes exceeds a quarter, but none exceed a half, a term such as mixed-lithology would be used. Other indications of sediment composition included clean, dirty, authigenic-crystal-bearing, organic, and wood-bearing.

EXTENDED ROCK DESCRIPTION

Numerous sedimentary rock descriptions include information more detailed than that suitable for the basic description, including information on weathering, detailed lithological attributes, and carbonate content. This information is rarely found. Where information on weathering of sedimentary rocks is found in drill records, it was parsed as either weathered or unweathered (table 11).

Detailed information on lithology that could readily be parsed as a lithological modifier (table 11) included the terms 'arenaceous', 'argillaceous', 'arkosic', 'bentonitic', 'carbonaceous/bituminous', 'chalky', 'cherty', 'concretionary/nodular', 'conglomeratic', 'dolomitic', 'feldspathic', 'ferruginous', 'fossiliferous', 'glauconitic', 'gypsiferous', 'kaolinitic', 'micaceous', 'oolitic', 'petroliferous', 'shaley', 'silica/siliceous', and 'sulphide-bearing'.

In a manner similar to that used for sediments, provision was made to permit description of sedimentary rocks as either 'noncalcareous', 'slightly calcareous', 'calcareous', or 'strongly calcareous' (see "Carbonate modi-

Table 11. Extended sedimentary rock description.

Weathering	Lithological modifier	Carbonate modifier
weathered	arenaceous	noncalcareous
unweathered	argillaceous	slightly-calcareous
	arkosic	calcareous
	bentonitic	strongly-calcareous
	carbonaceous/ bituminous	limestone
	chalky	dolostone
	cherty	
	concretionary/ nodular	
	conglomeratic	
	dolomitic	
	feldspathic	
	ferruginous	
	fossiliferous	
	glauconitic	
	gypsiferous	
	kaolinitic	
	micaceous	
	oolitic	
	petroliferous	
	shaley	
	silica/siliceous	
	sulphide-bearing	

fier” in table 11). In addition, the words ‘limestone’ and ‘dolostone’ are included under this attribute as modifiers of carbonate.

Among igneous and metamorphic rocks, indications of weathering were frequently reported, and these were captured as either ‘weathered’ or ‘unweathered’. Where possible, these rocks were further subdivided as either intrusive/high-grade metamorphic or low-grade metasedimentary/metavolcanic (table 12).

Table 12. Extended igneous/metamorphic rock description.

Weathering	Lithological modifier
weathered	intrusive/high-grade-metamorphic
unweathered	low-grade-metasedimentary/metavolcanic

COLOUR

All colour terminology interpreted from the well descriptions was coded according to the Munsell Book of Colour, a system that can readily be converted to other systems. The most important colour information is the single word such as ‘brown’ or ‘grey’ that was placed in an attribute called ‘principal colour’ (table 9). To capture other colour information, a colour such as ‘light yellowish brown’ was parsed into three variables (principal colour, principal value/chroma, and principal colour modifier; tables 9 and 13). In addition, in the rare cases when it was described, the pattern of multiple colours, whether ‘mottled’ or ‘speckled’, was coded. If the first two colours were distinct, the word ‘and’ was used as a colour link, whereas gradational colour relationships were indicated using the word ‘to’. This allowed a description such as ‘mottled light yellowish brown to dark greyish green’ to be fully captured. The term ‘salt and pepper’ was recorded as ‘speckled’.

Multiple colours were handled arbitrarily by assigning the first named colour to principal colour and the second named colour to the secondary colour. It is acknowledged that there commonly will be ambiguity in such cases, although these concerns relate only to a small number of strata, where multiple lithologies and multiple colours are mentioned. Table 14 presents examples of hypothetical cases that range from clear to ambiguous when parsed. While the first example in table 14 is clear, the second will be ambiguous when parsed, if not misleading, due to the tendency of the user to assume that a single colour relates to the first listed lithology. The third example, ‘red shale and blue clay’ will be ambiguous, but also misleading with respect to the tendency for the user to read the parsed data as red and blue shale interbedded with clay. Other examples in the table indicate, for example, capture of only the first two reported colours. The level of ambiguity in these situations is considered acceptable due to their low frequency of occurrence, and also due to recognition of the database complexity that would have been required to avoid ambiguity.

HYDROGEOLOGY

Information related to hydrogeology was captured in three variables (table 15). Strata were classified where possible as either ‘unsaturated’ or ‘saturated’ with water. Indications of gases based on effervescence or odour were captured as ‘effervescent/gassy’ under a solutes attribute, as were indications of salinity. Indications of permeability were classified as either ‘fracture/permeable’ or ‘tight/impermeable’.

STRATIGRAPHY

Stratigraphic nomenclature was captured (table 7), except where it made reference to the provenance of a gravel, such as ‘gravel composed of Pierre shale’.

QUALITY ASSURANCE

A key issue is the level of reliability that can be granted to the database. In answering this concern, a crucial distinction has to be made between applications that make reference to one site, versus applications that make reference to many sites in a manner that takes the location into consideration. When the information from one site is being inspected, there is little to be gained by referring to the parsed information rather then the original description, and a high level of caution is required in case the location is in error or the geological description is misguided. In contrast, however, are cases where the data from a large number of sites can be viewed, with location taken into consideration. By enabling the presentation of several hundred sites in map or section view, with the aid of x, y, and z coordinates along with categorized geological information that can be depicted using colours or symbols, a powerful means to assess reliability is obtained, because anomalous sites will differ from adjacent sites in a manner that is unreasonable from a geological perspective. Applications of the database are now indicating a surprisingly low frequency of apparent location errors or mistaken geological categorizations.

RECOMMENDATIONS

1. It is recommended that the parsed information be audited by members of the group responsible for maintenance of GWDrill, so that they may make their own conclusions regarding its validity.
2. Ongoing changes should be made to the classifica-

Table 13. Extended colour description.

Colour pattern	Principal value/chroma	Principal colour modifier	Principal colour (*from main table*)	Colour link	Secondary value/chroma	Secondary colour modifier	Secondary colour
mottled	brilliant	blackish	black	and	brilliant	blackish	black
speckled	dark	bluish	blue	to	dark	bluish	blue
	dusky	brownish	brown		dusky	brownish	brown
	light	greenish	green		light	greenish	green
	medium	greyish	grey		medium	greyish	grey
	moderate	olive	multicoloured		moderate	olive	multicoloured
	pale	orange	olive		pale	orange	olive
	strong	pinkish	orange		strong	pinkish	orange
	very-dark	purplish	pink		very-dark	purplish	pink
	very-dusky	reddish	purple		very-dusky	reddish	purple
	very-light	yellowish	red		very-light	yellowish	red
	very-pale		white		very-pale		white
	weak		yellow		weak		yellow

Table 14. Examples: parsing of colour information.

Hypothetical original description	Principal colour	Colour link	Secondary colour	Lithology	Interbedded with
red shale	red			shale	
shale and blue clay	blue			shale	clay
red shale and blue clay	red	and	blue	shale	clay
red to blue clay	red	to	blue	clay	
red and blue clay	red	and	blue	clay	
shale and red to blue clay	red	to	blue	shale	clay
shale and red and blue clay	red	and	blue	shale	clay
red to blue clay and green to black shale	red	to	blue	clay	shale
red to blue clay and green and black shale	red	to	blue	clay	shale
red and blue clay and green to black shale	red	and	blue	clay	shale
red and blue clay and green and black shale	red	and	blue	clay	shale

Table 15. Extended hydrogeological description.

Water	Solutes	Permeability
unsaturated	effervescent/gassy	fractured/permeable
saturated	saline	tight/impermeable

tion and its implementation on the basis of audits, test applications, and user consultations. These changes could initially be made by the authors during a transition period, but this responsibility should be transferred to the managers of the database at the earliest opportunity.

3. Following satisfactory audit and user consultations, the revised parsed information should be attached to the existing database using site ID and stratum sequence number as unique identifiers for the strata. At least the basic description, consisting of five variables, should be attached, if not the more detailed tables as well.

4. Data acquired in the future should be parsed to the adopted structure. Various means to accomplish this task on an ongoing basis should be tested. The simplest means of doing so, in our opinion, would be for new stratum descriptions, following correction of spelling, to be compared to the existing database, and for the existing parsed information to be duplicated where a match to a previously acquired record is found. In cases of newly submitted stratum descriptions that have never appeared before, the record could be referred to a geologist for interpretation, with the aid of a checklist of attributes and terminology. Although software that would process all conceivable wording could potentially be developed, this is not seen as a favourable option from the point of view of cost/benefit analysis.

5. The potential should be considered for water well drillers to utilize: A) a GPS receiver capable of obtaining adequate location data, and B) a checklist of attributes and words with which to describe strata. Doing so would require training and interaction with the drilling community on an ongoing basis, and would require provision for unrestricted comments to accompany the coded information.

ACKNOWLEDGMENTS

This work was done in increments over several years, and interaction over that time with R. Bezys, R. Betcher, L. Donnelly, R. Eilers, R. Fulton, L. Gray, C. Kaszycki, G. Keller, G. Matile, L. Murray, and F. Render is gratefully acknowledged.