HYDROGEOLOGICAL ASSESSMENT OF THE NORTHERN REGIONS OF GHANA

PRELIMINARY ASSESSMENT OF AVAILABLE ELECTRONIC DATABASES

August 2006


SNC•LAVALIN International In joint venture with


Université du Québec
Institut national de la recherche scientifique
Eau, Terre et Environnement

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Among the main objectives of the Hydrogeological Assessment Project of the Northern Regions of Ghana (HAP) is the establishment of the basis for current knowledge of the hydrogeological setting. One of the first steps to achieve this consists of assessing the content of the electronic hydrogeological databases made available by stakeholders and to eventually merge them into a consolidated database that will serve future hydrogeological projects in Northern Ghana as well as HAP downstream activities.

The main objectives of this preliminary assessment are to identify 1) the unique records among all available databases, 2) the reliable records among these unique records (N.B.: reliability based on the location data) and 3) the resultant data gaps within the unique and reliable records. The first objective is aimed at eliminating redundant information in view of a database consolidation while the second and the third are aimed at establishing the need (if any) for additional data acquisition. It is important to mention that this assessment is preliminary and is not to be held as thorough. In some cases, only parts of electronic databases were available at the time of the assessment and in all cases, a lot of the available data was left unverified. Only specific verifications that yielded results considered critical for the data consolidation process to go on were carried out. A more complete assessment will be done near the end of the database consolidation process in order to re-assess the data situation and confirm that minimum data requirements are met.

2 Available electronic databases

Through the key stakeholders, the six following hydrogeological electronic databases were obtained:

- Agence Française de Développement (AFD) database (MS Excel file)
- Canadian International Development Agency (CIDA) database (MS Excel file)
- European Union (EU) database (MS Excel file)
- Global Change in the Hydrological Cycle Project (GLOWA) database (MS Access file)
- Water Research Institute (WRI) database (Ground Water for Windows (GWW) file)
- World Vision (WV) database (Ground Water for Windows (GWW) file)

The first three databases, obtained from the Community Water Supply and Sanitation Agency (CWSA), contain only new records[1] created during their respective projects. All of these projects were carried out mainly in the Northern Region. As of the time of this assessment, only part of the expected data was available for these three databases as they originated from active projects that were not yet completed. It is expected that the additional data will be forthcoming by the end of 2006.

The GLOWA database contains records collected from different sources, notably contractors, regional CWSA offices and the WRI. Although research conducted under the GLOWA project mostly concerns the Volta Basin, this electronic database contains records for all of Ghana.

[1] In this document, unless other wise specified, a record of a specified database refers to a well with all its associated data (i.e. descriptive attributes).

The WRI database is considered the official hydrogeological electronic database for the Northern Regions of Ghana. It comprises records from many projects (e.g. Community Water Project (COWAP)) carried out in the Upper East, Upper West and Northern Regions.

Finally, the WV database contains available records created for World Vision projects. Some of the WRI records for the Northern Region were also appended to this WV database.

3 Methodology

The format and size of most of the available databases made it possible to carry out the analysis and queries in ArcGIS 9.0. For the GWW format files (i.e. WRI and WV databases), some manual editing was required before transfer into ArcGIS as the output format of GWW is an ASCII text file. The use of a Geographical Information System (GIS) such as ArcGIS was necessary for this preliminary assessment since spatial analysis functions were required.

All available databases were first examined to determine the total number of records. The assessment then began with the identification of the unique records in each database. The evaluation of location data reliability (i.e. the coordinates) followed. Finally, statistics were calculated for each of the selected data fields. Although the electronic databases available were analysed differently with respect to their content, the general procedure is described below for each objective.

3.1 Unique records

The identification of unique records did not require the same method for all databases as information was sometimes available concerning the origin of data. First, the records from the AFD, EU and CIDA databases were all considered unique since they came from recent projects and were unlikely to have been entered (and thus duplicated) in any other database yet. Consequently, verification of record uniqueness was not undertaken for these databases. For the GLOWA and WV databases, it was possible to determine the presence of redundant records through queries. It was determine that such records were all originally taken from the WRI database. Consequently, all records in the WRI were considered unique while redundant records in the GLOWA and WV databases were ignored. Different methods were used in identifying redundant records between GLOWA & WRI and WV & WRI.

3.1.1 GLOWA database

The identification of unique records in the GLOWA database was done by relating records of the GLOWA database with records of the WRI database. The creation of reliable link between the databases required the use of two key data fields (Well ID and Project ID) and the correction of syntax errors in the Project ID field. The use the Project ID field was required since many records present in both GLOWA and WRI databases had different Well IDs although they were clearly the same wells. On the other hand, the use of the Well ID field was also required because some records present in both databases had the same Project ID, which is to be expected since many wells could have drilled during the same project. Unfortunately, the use of both data fields was not always sufficient for identification and visual inspection based on other data fields (e.g. community name and well completion date) had to be done to identify some common

records. The majority of the redundant records were however identified following these steps:

1) Creation of a new temporary data field with the corrected project number
2) Determination of common records between the two databases based on Project IDs and Well IDs
3) Creation of a new data field to store a unique record identifier (unique record = 1 and common record = 0 (N.B.: data from WRI database was kept for common records))

3.1.2 World Vision database

The identification of unique records in the WV database was simpler since the Project IDs of the World Vision projects were known (personal communication with Enoch Asare from WRC, 2006). Therefore, records with the following project suffixes were identified as unique records: UNICEF, OIC, and WV.

3.2 Reliable records

Prior to the verification of the location data, regions and districts names in the databases had to be corrected for syntax errors. The table 1 presents the region and district names used. The official and updated districts names for the Northern Regions (available at the following URL http://ghanadistricts.com/home) were not used in this assessment because the associated file giving the location of each official district was not available at the moment of this assessment.

Table 1: Region and district names used

Region	District
Northern	Bole
Northern	East Gonja
Northern	East Mamprusi
Northern	Gushiegu Karaga
Northern	Nanumba
Northern	Saboba Chereponi
Northern	Savelugu Nanton
Northern	Tamale
Northern	Tolon Kumbungu
Northern	West Gonja
Northern	West Mamprusi
Northern	Yendi
Northern	Zabzugu Tatale
Upper East	Bawku East
Upper East	Bawku West
Upper East	Bolgatanga
Upper East	Bongo
Upper East	Builsa
Upper East	Kassena Nankana
Upper West	Jirapa Lambussie
Upper West	Lawra
Upper West	Nadowli
Upper West	Sissala
Upper West	Wa

For this preliminary assessment, the evaluation of reliability was carried out only for the data field containing the well coordinates (i.e. longitude and latitude). Because the original location data (e.g. paper logs or GPS datasheet) was not yet available, coordinates reliability was mainly evaluated using spatial analysis functions with respect to administrative boundaries from an independent data source[2]. The presence of syntax errors in coordinates and the inaccuracy of both coordinates and administrative boundaries made the use of decision trees appropriate for this task. Records identified as unreliable were kept for future and more thorough verifications (with original location data if possible). It is important to mention upfront that this was a time consuming task and that, therefore, it was not carried out with the same level of detail for all databases. Priority was given to the WRI database since it was considered to hold the larger number of unique records. The GLOWA and WV databases were also verified for location data reliability but to a lesser extent. As for the AFD, EU and CIDA databases, only minor verifications were made since many records were missing coordinates (N.B.: updated versions of these databases with coordinates for all records are expected by the end of 2006). While a complete analysis of all databases could have yielded a greater number of reliable records, the effort needed to accomplish this was considered disproportionate at that time in regards to the results that would be obtained.

3.3 Data gaps for selected data fields

For the HAP purposes, the following data fields were considered to be the minimum data requirements[3] to carry out the necessary analysis: 1) Well state, 2) Well depth, 3) Weathered layer thickness, 4) Lithology, 5) Groundwater level. 6) Yield, 7) Water quality.

The identification of data gaps for these data fields is twofold: 1) identification of gaps in terms of data quantity and 2) identification of data gaps in terms of spatial distribution of data. The first part was simply done with the help of statistics. Records containing information for each of these data field were compiled in order to evaluate the quantity of data available for each data field. The records resulting from this compilation were then plotted for each data field to assess their spatial distribution. This was done with regards to a 15kmx15km cell grid that was considered the minimum requirement in terms of data distribution for HAP purposes (i.e. at least one borehole must be present in each cell – minimal density of 1 borehole per 225 km²). The dimensions selected for the grid cell size is comparable to the size used in similar regional studies.

4 Results

Table 2 presents the total number of records for each database. The sum of these total records (15,092 records) does not give a representative idea of the amount of data that can actually be used for hydrogeological analysis. It is also important to mention that the total appearing in this table for the GLOWA database represents only the number of records relevant to the Northern Regions. A subset of records had to be selected[4] since the GLOWA database includes records for all of Ghana.

---

[2] At the moment of writing, the most reliable data source for regions and district boundaries was considered to be the Solar and Wind Energy Resources Assessment (SWERA) Project.

[3] These minimum data requirements only apply to hydrogeological data found in the available databases; other required data for HAP, such as meteorological data, are not included in these requirements.

[4] The selection was based on the Region data field; the actual total number of records that are available for all of Ghana in the GLOWA database is 15212.

Table 2: Total number of records for available databases

Database	Total number of records
Agence Française de Développement (AFD)	231
Canadian International Developement Agency (CIDA)	859
European Union (EU)	483
Global Change in the Hydrological Cycle Project (GLOWA)	6571
Water Research Institute (WRI)	5984
World Vision (WV)	964
Total	15092

4.1 Unique records

4.1.1 AFD, EU, CIDA databases

As mentioned previously, records in theses databases were all considered unique as they come from ongoing projects and were therefore considered to not be duplicated among the database examined. Consequently, the numbers of unique records, which also correspond to the total number of records for each database, are 231 for AFD, 483 for EU and 859 for CIDA.

4.1.2 GLOWA database

The identification of unique records for the GLOWA database was done with respect to the WRI database, for which all records are considered unique as mentioned earlier. The analysis yielded 1,406 unique records out of a total 6,571 records for the GLOWA database. The 5,165 other records in the GLOWA database are thus common to the WRI database. A quick examination of these 5,165 records revealed that, although they contain data common to both GLOWA and WRI databases, there is also specific data unique to each databases for some records (e.g. for one particular common record, the thickness of the weathered layer might be available in the GLOWA database while nonexistent in the WRI database). Although records from the WRI database were considered over the ones of GLOWA, a closer examination of redundant records in the GLOWA database should be carried out to extract the additional information in view of the consolidation process. The details of the analysis are presented below in table 3.

Table 3: Unique and common records for the GLOWA database

GLOWA (compared to WRI)	Records	Status
Common Well ID and Project ID	3881	Common
Common Project ID only (1)	1220	Common
Common Well ID only (2)	64	Common
No data fields in common (3)	37	Unique
No data fields in common	1369	Unique
Total	6571	

Notes:

(1) : the Project ID (and other data fields) were common but Well ID was different (N.B.: it was assumed that Well IDs were changed for specific project purposes)

(2) : the Project ID was missing for these records

(3) : all data fields were different but the Project ID of these records existed in both database (N.B.: this situation may arise if different wells of a same project were entered in the two databases analysed)

4.1.3 WRI database

Records from the WRI database were all considered unique as it is the reference database. It is however important to mention that no verification was done to identify record duplicates during this assessment. The number of unique records, which in this case also corresponds to the total number of records, is 5,984.

4.1.4 WV database

Unique records from the World Vision database were identified on the basis of the Project ID. The query made for the Project IDs identified as unique WV projects returned the following number of unique records: UNICEF Project (98), OIC Project (39) and WV Project (240). The total number of unique records is thus 238 out of a total 964 records in this database.

4.2 Reliable records with respect to location data

4.2.1 AFD, EU, CIDA databases

Location data (i.e. latitude and longitude) of the EU, AFD, CIDA databases were generally considered reliable if they fell within the Northern Regions. Minor verifications were done for these databases, notably to find and correct syntax errors and to identify records with coordinates falling slightly outside Northern Regions but still relatively near to their corresponding district[5]. A complete and more thorough assessment of location data should be carried out when all coordinates will be available for these databases. Meanwhile, the following records were found to have reliable coordinates: 82 out 231 for AFD, 397 out of 483 for EU and 435 out of 859 for CIDA.

4.2.2 GLOWA database

The decision tree used to assess location data reliability is illustrated in figure 1. Results, which are also shown on that figure, reveal that 6,136 records apparently had reliable coordinates while 435 records were flagged as unreliable. In order to make all coordinates reliable, access to original data sheets or additional field work (i.e. GPS survey) will be necessary (N.B.: 433 out of these 435 records don't have any coordinates).

4.2.3 WRI database

The location data reliability analysis carried out on this database is presented in figure 2 along with the results. From this preliminary assessment, there are 4,498 records with apparently reliable coordinates and 1,486 records that either had unreliable coordinates or were missing coordinates. Although further analysis could help reduce the latter number, it is considered that the amount of work needed would be disproportionate in regards of the results expected. This is notably explained by the problems arising from the use of community names to conduct further analysis on coordinates reliability. Such problems include:

- syntax errors in community names (manual corrections);
- absence of communities in one of the database (manual update of database);
- presence of new communities in one of the database;
- coordinate discrepancies for the same community.

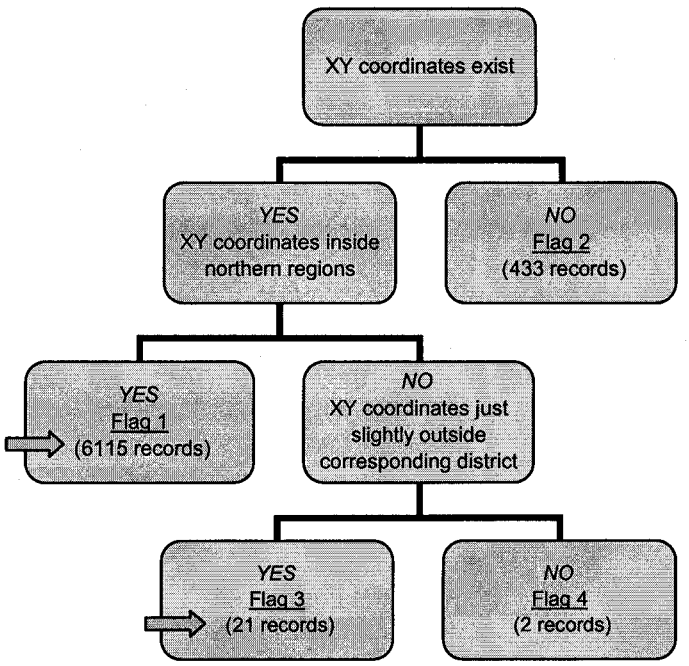
[5] The problem of coordinates falling outside the Northern Regions but near their corresponding district can be attributed to inaccuracy of administrative boundaries or of coordinates themselves.

Consequently, and depending on HAP needs, further coordinate verification could be done on a limited number of these 1,486 records. In any case, access to original data sheets or additional field work (i.e. GPS survey) will probably be necessary to make all coordinates reliable.

4.2.4 WV database

Results show 898 records with apparently reliable coordinates and 66 records that either had unreliable coordinates or were missing coordinates (N.B.: 21 out of these 66 records don't have any coordinates). Obvious longitude errors (i.e. East vs West direction) were corrected and 'flagged' reliable with a short description of the correction. Figure 3 illustrates the results.

Figure 1: Location data reliability assessment (GLOWA database)



Notes:

⇨ : Data selected as reliable

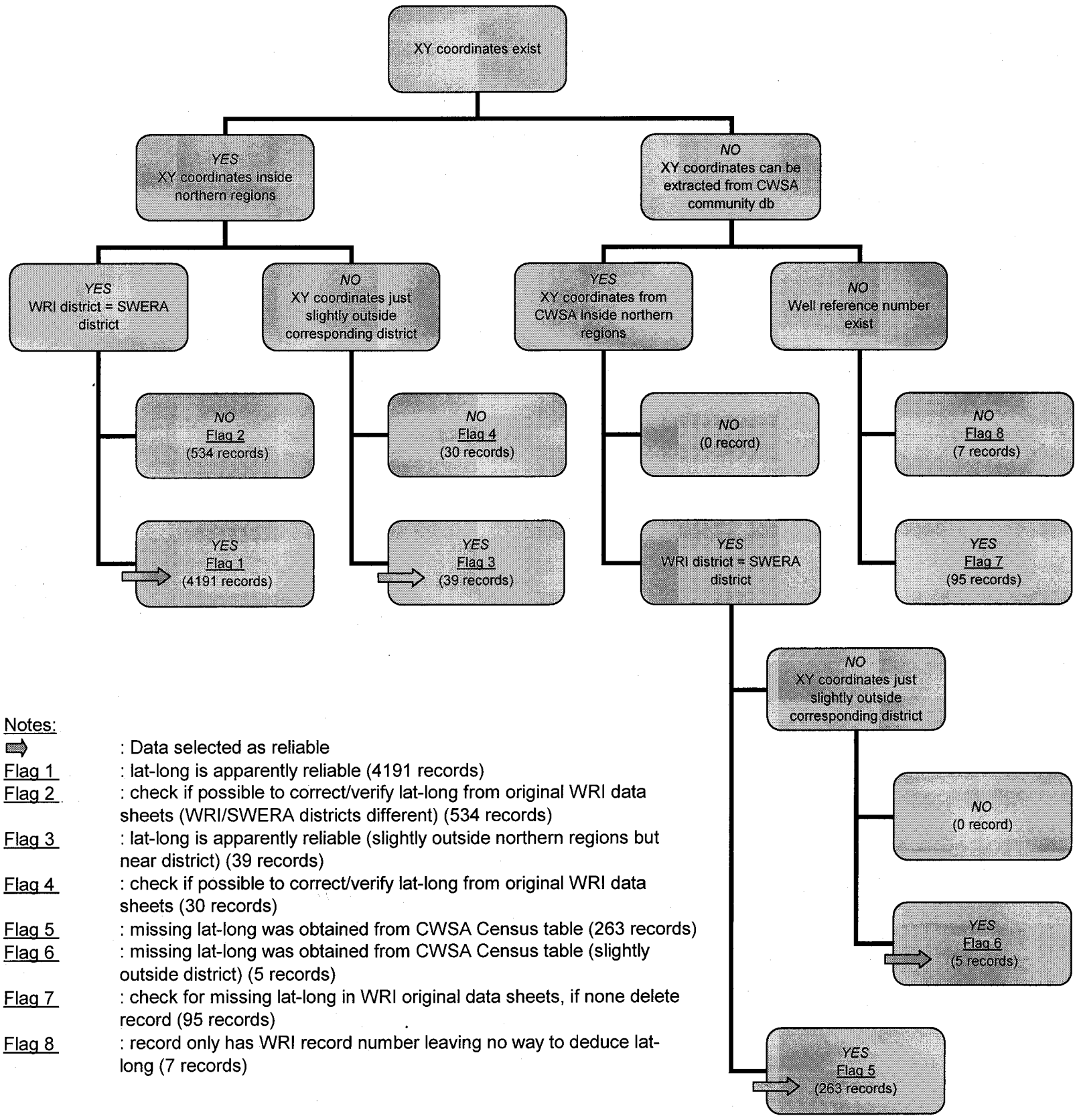
Flag 1 : lat-long is apparently reliable (6115 records)

Flag 2 : missing lat-long - check if possible to obtain from other data sources (433 records)

Flag 3 : lat-long is apparently reliable (slightly outside Northern Regions but near district) (21 records)

Flag 4 : check if possible to correct/verify lat-long from other data sources (2 records)

Figure 2: Location data reliability assessment (WRI database)



Notes:

⇨ : Data selected as reliable

Flag 1 : lat-long is apparently reliable (4191 records)

Flag 2 : check if possible to correct/verify lat-long from original WRI data sheets (WRI/SWERA districts different) (534 records)

Flag 3 : lat-long is apparently reliable (slightly outside northern regions but near district) (39 records)

Flag 4 : check if possible to correct/verify lat-long from original WRI data sheets (30 records)

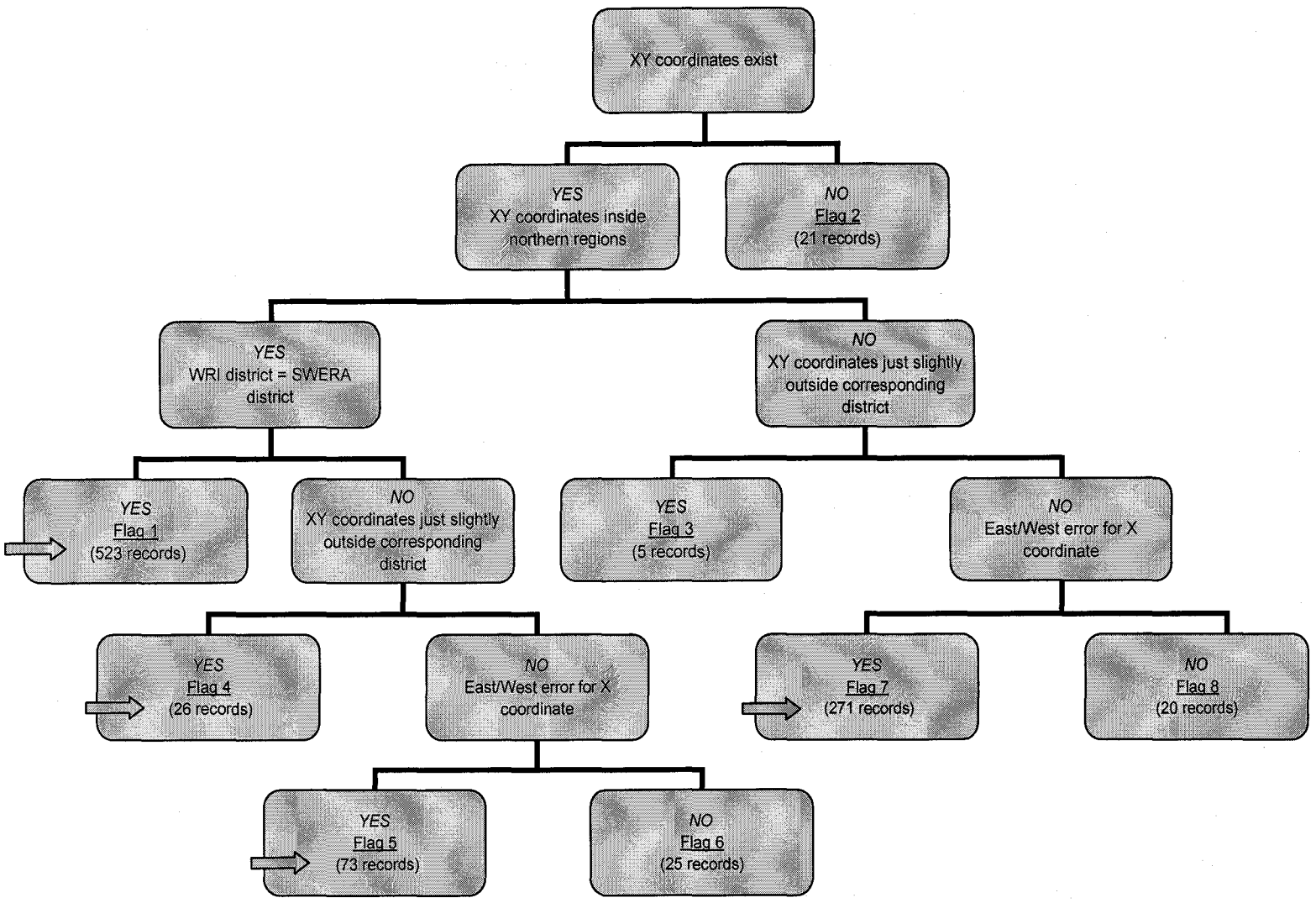
Flag 5 : missing lat-long was obtained from CWSA Census table (263 records)

Flag 6 : missing lat-long was obtained from CWSA Census table (slightly outside district) (5 records)

Flag 7 : check for missing lat-long in WRI original data sheets, if none delete record (95 records)

Flag 8 : record only has WRI record number leaving no way to deduce lat-long (7 records)

Figure 3: Location data reliability assessment (WV database)



Notes:

⇨ : Data selected as reliable

Flag 1 : lat-long is apparently reliable (523 records)

Flag 2 : missing lat-long - check if possible to obtain from other data sources (21 records)

Flag 3 : lat-long is apparently reliable (slightly outside northern regions but near district) (5 records)

Flag 4 : lat-long is apparently reliable (slightly outside district but near district) (26 records)

Flag 5 : lat-long is apparently reliable (after longE correction) (73 records)

Flag 6 : check if possible to correct/verify lat-long from other data sources (25 records)

Flag 7 : lat-long is apparently reliable (after longE correction) (271 records)

Flag 8 : check if possible to correct/verify lat-long

4.3 Data gaps for selected data fields

Following the analysis for unique and reliable records (reliability of location data only), statistics were calculated for each database. Table 4 shows that the total number of unique records is 9,340 and that 7,056 of these records are considered reliable (as far as location data is concerned). The last column reveals that there are major gaps in the lithology and weathered thickness[6] data fields for all databases except the WV database. For most databases, other major gaps are also found in the following data fields: groundwater level, yield and water quality.

Table 5 presents a summary of the previous table. From these results, it is plain that the number of records that can actually be used for hydrogeological analysis (i.e. last row of the table) is largely insufficient for HAP purposes.

While the tables revealed the obvious need for additional data, it could not tell where data is most needed geographically. Therefore, a verification of the spatial distribution of the 71 records identified above revealed that only 26 cells have one or more boreholes in them (figure 4). Considering that 507 cells (15x15km) are necessary to cover the entire study area, it is obvious that additional data is needed. The major spatial data gaps are easily spotted on figure 4 (i.e. basically every cell without a yellow dot, so all of the Northern and Upper West Regions and parts of the Upper East Region).

5 Conclusion

The results of this assessment indicate that additional data is required, both in terms of quantity and in terms of spatial distribution, over and above what is currently available in electronic databases in the sector. In the context of the HAP, it was proposed that a subset of 450-500 wells (~1/225 km^2) with more reliable information be created. In order to build this subset, the access to original documents is crucial. The most efficient and reliable way to select these reliable wells is to go through the hard copies of available documents. The selection, which would have to be carried out or supervised by a local hydrogeologist, could be based of the following criteria:

- location (one well per 15km by 15 km cells);
- information available (more than one well per cell could be selected if data requirements can not be met with only one well);
- contractor/consulting engineers (local knowledge of data reliability with respect to contractors/consulting engineers will help in the selection);
- representativeness (the selected well(s) would have to represent the average conditions encountered in the cell - this can be based on borehole logs inspection);
- well status (selected wells would have to be active/usable to allow for possible water level measurement for example).

It is without any doubts, a long process to go through but the resulting subset would represent a significant contribution to the hydrogeological database of the Northern Regions.

[6] Although weathered thickness is considered a required data field, it is not as critical as others since access to reliable and detailed lithological information can generally be used to define the limits of the weathered layer.

Table 4: Statistics for selected data fields in all databases

Database / Field	Records				
	Total	XY reliable	Unique	Unique/XY reliable	Unique/XY reliable (%)
WRI	5984	4498	5984	4498	100%
Well state	5919	4453	5919	4453	99%
Well depth	2890	1747	2890	1747	39%
Weathered layer thickness	0	0	0	0	0%
Lithology	374	333	374	333	7%
Groundwater level (survey or	478	388	478	388	9%
Yield (airlift or pumping test)	2421	2117	2421	2117	47%
Water quality (2)	175	145	175	145	3%
GLOWA	6571	6136	1406	1296	100%
Well state	6560	3959	1366	1261	97%
Well depth	6183	5822	1192	1122	87%
Weathered layer thickness	1112	1084	291	277	21%
Lithology (1)	3943	3911	511	494	38%
Groundwater level (survey or	4117	4076	655	628	48%
Yield (airlift or pumping test)	3985	3959	417	406	31%
Water quality (2)	0	0	0	0	0%
CIDA	859	435	859	435	100%
Well state	416	294	416	294	68%
Well depth	383	273	383	273	63%
Weathered layer thickness	0	0	0	0	0%
Lithology	0	0	0	0	0%
Groundwater level (survey or	153	149	153	149	34%
Yield (airlift or pumping test)	256	166	256	166	38%
Water quality (2)	110	106	110	106	24%
EU	483	397	483	397	100%
Well state	1	1	1	1	0%
Well depth	365	362	365	362	91%
Weathered layer thickness	0	0	0	0	0%
Lithology	0	0	0	0	0%
Groundwater level (survey or	105	104	105	104	26%
Yield (airlift or pumping test)	117	116	117	116	29%
Water quality (2)	2	2	2	2	1%
AFD	231	82	231	82	100%
Well state	11	11	11	11	13%
Well depth	82	82	82	82	100%
Weathered layer thickness	0	0	0	0	0%
Lithology	0	0	0	0	0%
Groundwater level (survey or	81	81	81	81	99%
Yield (airlift or pumping test)	82	82	82	82	100%
Water quality (2)	71	71	71	71	87%
WV	964	898	377	348	100%
Well state	964	898	377	348	100%
Well depth	963	897	377	348	100%
Weathered layer thickness	0	0	0	0	0%
Lithology	964	898	348	348	100%
Groundwater level (survey or	320	302	152	141	41%
Yield (airlift or pumping test)	199	189	56	51	15%
Water quality (2)	113	108	46	46	13%
Total	15092	12446	9340	7056	-

Notes:

(1) : lithology in the GLOWA database is limited to one column (no stratigraphic unit description)
(2) : records for which water quality data is available for at least: pH, EC, F, Fe and Mn

Table 5: Summary statistics for unique and reliable records in all databases

	Records	Area covered (1)	Required for HAP
Total records (in all databases)	15092	-	-
Unique records	9340	389	-
Unique & reliable records	7056	381	-
Unique & reliable records with well state	6368	352	-
Unique & reliable records with well depth	3934	352	-
Unique & reliable records with weathered layer	277	86	-
Unique & reliable records with lithology	681	99	-
Unique & reliable records with water level	1491	289	-
Unique & reliable records with yield	2938	297	-
Unique & reliable records with water quality (2)	370	163	-
Unique & reliable records with all required fields (3)	71	26	507 (4)

Notes:

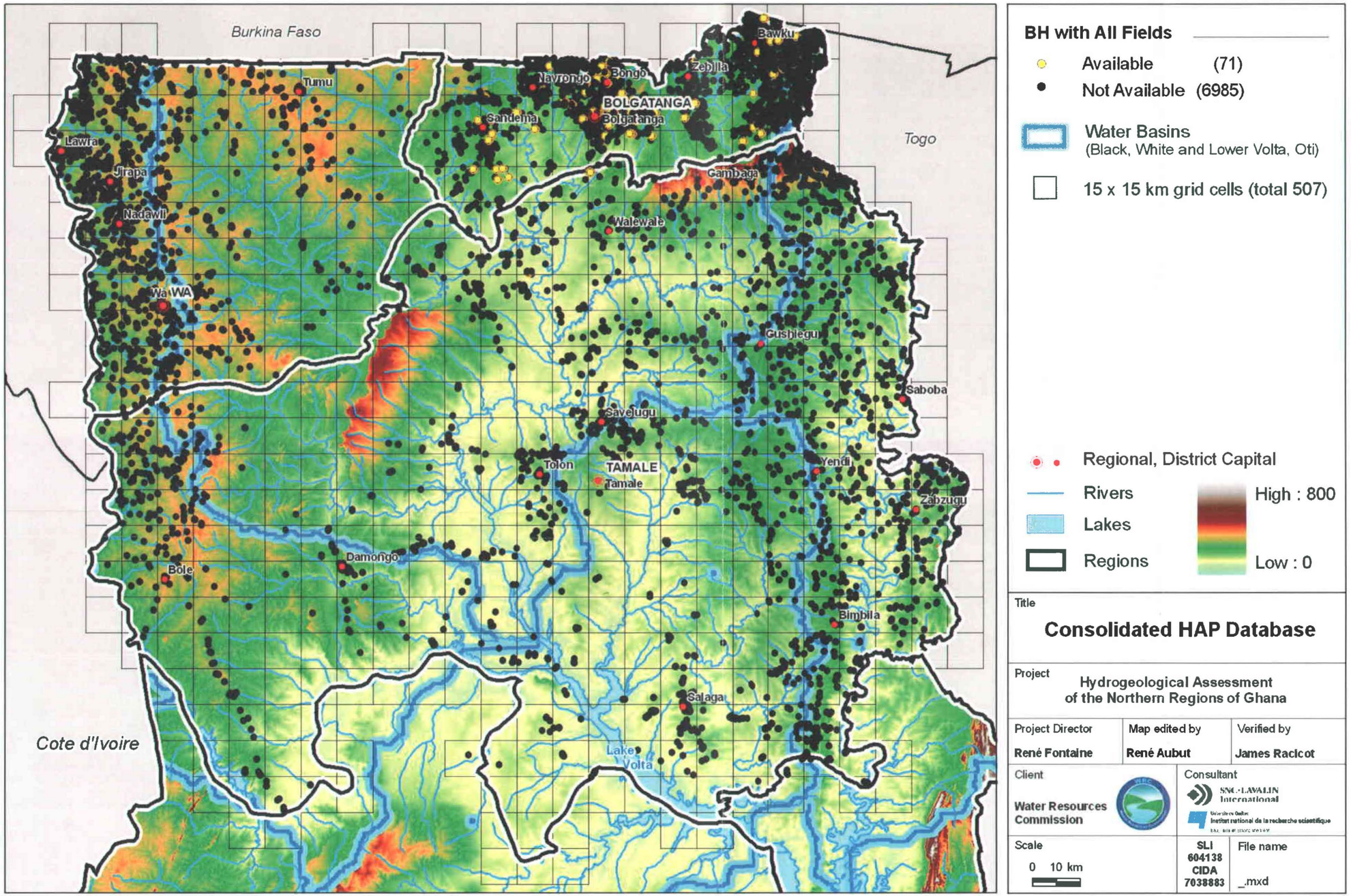
(1) : the area covered by each category is expressed by the number of 15x15km cells with one or more record in it

(2) : records for which water quality data is available for at least the following: pH, EC, F, Fe and Mn (N.B.: zeros exlcuded)

(3) : the field containing the weathered layer thickness was excluded from this calculation since it can be derived from lithological information

(4) : this target represents the total number of 15x15km cells in the northern regions (97721 km^2); for each cells, we need at least one borehole with reliable information for all required fields (N.B.: cell size is arbitrary but comparable to size used in similar regional studies; uniform data distribution

Figure 4: Spatial distribution of unique & reliable records with all required data fields



Prepared by: _____ Date: August 2006

Marc-André Carrier
GIS/Mapping advisor

Reviewed by: _____ Date: August 2006

James Racicot
Ghana-based Canadian advisor

_____ Date: August 2006

René Lefebvre
Technical Capacity Building Team Leader