WORLD OIL RESERVES DATA: INFORMATION QUALITY ASSESSMENT ANALYSIS

(Completed Academic Paper)

Yusuf Yiliyasi
University of Arkansas at Little Rock
yxyiliyasi@ualr.edu

Daniel Berleant
University of Arkansas at Little Rock
jdberleant@ualr.edu

Abstract: While high quality oil data can help oil companies and governments reduce risk of investments and increase profits, bad oil data can lead to economic losses. In this research, we propose a framework for assessing the information quality of world oil reserves data. The framework is applied by calculating the information quality ratings of relevant information sources. The information quality dimensions assessed in this framework include Data Decay, Data Integrity and Reputation, Data Coverage and Completeness, Degree of Compliance with Data Standards, Expertise of Data Source, and Degree to which Data was Vetted. Using the proposed framework, we assess the information qualities of some well-known sources of oil reserve data.

Keywords: Data Quality, Information Quality, Information Quality Framework, Information Product, World Oil Reserves, Information Quality Dimension

GENERAL BACKGROUND

Information quality (IQ) theories and frameworks have been increasingly studied and practiced in various organizations to assess, improve and monitor the quality of their information products [1]. However, information quality issues remain pervasive in organizations, industries and government institutions. An example of this pervasiveness is the energy industry, where IQ problems in the crude oil exploration and production industry have important implications due to the heavy reliance of modern economy on petroleum.

Bad data can increase risks of petroleum-related investments and lead to economic damage. Various government as well as non-government organizations, institutions and companies collect, compile and distribute oil related data in order to satisfy the needs of information consumers such as investors, companies, and governments. However, unbiased oil data are often either regarded as secret by some oil companies and governments or are not freely available. In some cases, oil reserve figures are exaggerated for economic and political purposes [2]. These add to the challenges of obtaining high quality oil data. In order to help data consumers make better decisions, our research addresses the question of assessing the qualities of oil information products. In doing this, we develop a framework for assessment based on widely accepted foundations in the information quality field. Using this framework, we analyze the qualities of several sets of data on international oil reserves provided by well-known information sources.

Some United States federal government agencies have enacted information quality guidelines [3]. Among these, the Department of Energy and the Energy Information Administration specialize in producing and disseminating energy-related data such as crude oil reserves and production levels [4]. However, in-depth analysis, assessment and quantification of the information quality of oil reserve data and their sources

against an explicit information quality framework is currently lacking. Thus there is a need for such an information quality framework and its application to the assessment of the IQs of world oil reserves data and their sources.

AN INFORMATION QUALITY FRAMEWORK FOR OIL DATA

Information can be defined as a product of an information producing system that processes raw data into an information product that adds value for the information consumer [5]. Accordingly, we consider world oil reserve data as an information product whose quality can be measured. We assess the quality on certain dimensions for each data source. The qualities of the dimensions can be combined to obtain the overall IQ of each source. Here are the IQ dimensions of our framework.

- Data Decay
- Data Integrity and Reputation
- Data Coverage and Completeness
- Degree of Compliance with Data Standards
- Expertise of Data Source
- Degree to which Data was Vetted

We have selected these IQ dimensions both because they are critical quality indicators of oil data, and because they can be quantified in a transparent manner. However, it should be noted that there are other IQ dimensions such as *Accuracy*, *Consistency* and *Transactability* that might also be useful to include in a framework of this type if the practical difficulties in assessing and quantifying them can be solved. The selected IQ dimensions are discussed in more detail next along with their application to sources of oil reserve information.

Data Decay. A measure of the rate of deterioration of the quality of the data. Data decay rates can help determine the frequency of updates necessary for the data. Volatile data that require high reliability need more frequent updates compared to similar data with a lower decay rate [6]. *Data Decay* is closely related to *timeliness*. Lee and Pipino *et al.* proposed a metric to calculate the *Timeliness* of data as follows [7]:

$$\text{Timeliness rating} = \{\max [(1 - \frac{\text{currency}}{\text{volatility}}), 0]\}^{s} \qquad (1)$$

We adapt that result slightly and propose the following to quantify *Data Decay* [7]:

$$\text{Data Decay rating} = \min (1, \frac{\text{currency}}{\text{volatility}}) \qquad (2)$$

where currency is the age of data and volatility is the age at which the data is considered decayed and thus worthless. The volatility is best determined by an IQ assessor based on context and characteristics of data, data source and expert opinion. Because a high *Data Decay* rating implies low information quality, we invert formula (2):

Information quality rating for the *Data Decay* dimension = 1 - Data Decay rating (3)

We consider oil reserves data outdated and therefore unsuitable for decision making if it is 30 or more years old. This means that oil reserve data becomes decayed at 30 years of age, that is, volatility = 30, in the formula.

Data Decay ratings of specific sources.
Wired.com. The most recent world oil reserve data is dated Jan. 1, 2007 [8]. Thus, its data is not very current. According to eqs. (2) and (3) above, *Data Decay* rating = (2011 - 2007) / 30 = 0.133, so the information quality rating in terms of the *Data Decay* dimension is 1 – 0.2 = 0.867.

Energy Information Administration (EIA). The most recent world oil reserve data is dated 2011 [9]. Its data is current. According to eqs. (2) and (3) above, *Data Decay* rating = (2011 - 2011) / 30 = 0, so the information quality rating in terms of the *Data Decay* dimension is 1 – 0 = 1.

Europe's Energy Portal. The most recent world oil reserve data is dated 2009 [10]. Its data is relatively current. According to eqs. (2) and (3) above, *Data Decay* rating = (2011 - 2009) / 30 = 0.0667, so the information quality rating in terms of the *Data Decay* dimension is 1 – 0.0667 = 0.933.

British Petroleum (BP). The most recent world oil reserve data is dated 2009 [11]. Its data is relatively current. According to eqs. (2) and (3) above, *Data Decay* rating = (2011 - 2009) / 30 = 0.0667, so the information quality rating in terms of the *Data Decay* dimension is 1 – 0.0667 = 0.933.

Organization of the Petroleum Exporting Countries (OPEC). The most recent world oil reserve data is dated 2009 [12]. Its data is relatively current. According to eqs. (2) and (3) above, *Data Decay* rating = (2011 - 2009) / 30 = 0.0667, so the information quality rating in terms of the *Data Decay* dimension is 1 – 0.0667 = 0.933.

Central Intelligence Agency (CIA). The most recent world oil reserve data is dated 2010 [13]. Its data is relatively current. According to eqs. (2) and (3) above, *Data Decay* rating = (2011 – 2010) / 30 = 0.0333, so the information quality rating in terms of the *Data Decay* dimension is 1 – 0.0667 = 0.967.

Titi Tudorancea Bulletin. The most recent world oil reserve data is dated 2010 [14]. Its data is relatively current. According to eqs. (2) and (3) above, *Data Decay* rating = (2011 – 2010) / 30 = 0.0333, so the information quality rating in terms of the *Data Decay* dimension is 1 – 0.0667 = 0.967.

Data Integrity and Reputation. This measure reflects an overall subjective quality assessment of data based on a period of use by its consumers. If data is accurate but lacks a good reputation, data consumers might hesitate to make decisions using the data. In research on the relation of PageRank to co-citation of authors, Ding *et. al* [15] noted a high correlation of citation rank and website's PageRank. We consider the PageRank score as an indicator of the overall reputation of a website. Therefore, we propose to quantify *Data Integrity and Reputation* with the Google PageRank score [16] for the homepage of the data source website:

Data Integrity and Reputation = Google PageRank score for data source website (4)

Data Integrity and Reputation of specific sources.
Wired.com. The Google PageRank for wired.com was 8/10. Therefore, based on the above formula, the IQ rating for the dimension of *Data Integrity and Reputation* for wired.com is 0.80.

EIA. The Google PageRank for www.eia.gov was 8/10. Therefore, based on the above formula, the IQ rating for the dimension of *Data Integrity and Reputation* for EIA is 0.80.

Europe's Energy Portal. The Google PageRank for www.energy.eu was 7/10. Therefore, based on the above formula, the IQ rating for the dimension of *Data Integrity and Reputation* for Europe's Energy Portal is 0.70.

BP. The Google PageRank for www.bp.com was 8/10. Therefore, based on the above formula, the IQ rating for the dimension of *Data Integrity and Reputation* for BP is 0.80.

OPEC. The Google PageRank for "www.opec.org/opec_web/en/index.htm" was 8/10. Therefore, based on the above formula, the IQ rating for the dimension of *Data Integrity and Reputation* for OPEC is 0.80.

CIA. The Google PageRank for "www.cia.gov" was 7/10. Therefore, based on the above formula, the IQ rating for the dimension of *Data Integrity and Reputation* for CIA is 0.70.

Titi Tudorancea Bulletin. The Google PageRank for "http://www.tititudorancea.com" was 3/10. Therefore, based on the above formula, the IQ rating for the dimension of *Data Integrity and Reputation* for The Titi Tudorancea Bulletin is 0.30.

Data Coverage and Completeness. *Completeness* measures the extent to which data records are available in the data collection. It can describe the extent to which all records of all variables are present for use [1]. *Coverage* measures the availability of data relative to the total data universe [6].

In this research on the IQ of world oil reserves data, one important aspect is the availability of oil reserves data for all countries. This motivates assessing the *Coverage and Completeness* dimension of world oil reserves data. Lee and Pipino *et al.* proposed a metric to calculate the *completeness* of data in a given database as follows [17]:

$$\text{Completeness rating} = 1 - \left(\frac{\text{Number of incomplete items}}{\text{Total number of items}}\right) \qquad (5)$$

We consider oil reserve figure for one country as one data item, and the data universe as the total number of countries in the world. Based on above formula, we propose the following formula to quantify the *Data Coverage and Completeness* of world oil reserves data from an information source:

$$\text{Data Coverage and Completeness rating} = 1 - \frac{\text{Number of incomplete countries}}{\text{Total number of countries}} \qquad (6)$$

Data Coverage and Completeness of specific sources.
Wired.com. We found that it listed oil reserve figures for the top 20 countries and a single figure for the rest of the world. This data is incomplete since the oil reserve figures for many countries are not listed separately; instead, a single summed figure is listed for them [8].
According to above formula, the rating of wired.com is:
Data Coverage and Completeness rating = 1 - (209 - 20) / 209 = 0.096.

EIA. We found that it listed oil reserve figures for the 209 countries of the world [9]. According to above formula, the rating of EIA is:
Data Coverage and Completeness rating = 1 - (209 - 209) / 209 = 1.

Europe's Energy Portal. We found that it listed oil reserve figures for 44 countries of the world [10].
According to above formula, the rating of Europe's Energy Portal is:
Data Coverage and Completeness rating = 1 - (209 - 44) / 209 = 0.211.

BP. We found that it listed oil reserve figures for 48 countries of the world [11]. According to above formula, the rating of BP is:
Data Coverage and Completeness rating = 1 - (209- 48) / 209 = 0.230.

OPEC. We found that it listed oil reserve figures for 44 countries of the world [12]. According to above formula, the rating of OPEC is:
Data Coverage and Completeness rating = 1 - (209 - 44) / 209 = 0.211.

CIA. We found that it listed oil reserve figures for 209 countries of the world [13]. According to above formula, the rating of CIA is:
Data Coverage and Completeness rating = 1 - (209 - 209) / 209 = 1.

Titi Tudorancea Bulletin. We found that it listed oil reserve figures for 97 countries of the world [14]. According to above formula, the rating of the Titi Tudorancea Bulletin is:
Data Coverage and Completeness rating = 1 - (209 - 97) / 209 = 0.464.

***Degree of Compliance with Data Standards*.** The existence of organization-wide IQ awareness and observance of an established IQ standard is a significant indication of a high quality information provider. To assess this dimension, we determined if the data source has demonstrated IQ awareness by explicitly following industry, national, or international standards in information manufacturing and maintenance.

We use the following formula:

$$\text{Degree of Compliance with Data Standards} = \begin{cases} 0\text{: No IQ standards followed.} \\ 0.25\text{: Demonstrated IQ awareness.} \\ 0.5\text{: Industry level IQ standards followed.} \\ 1\text{: National or international IQ standards followed.} \end{cases} \quad (7)$$

Degree of Compliance with Data Standards of specific sources.
Wired.com. Based on the information available on its website, no description of specific information quality standards followed was in evidence. Accordingly, *Degree of Compliance with Data Standards* = 0 for this source.

EIA. We noted EIA has implemented an information quality standard [4]. EIA is a U.S. federal government agency, therefore its IQ standards are considered national. Accordingly, *Degree of Compliance with Data Standards* = 1 for this source.

Europe's Energy Portal. Based on the information available on its website, no description of specific information quality standards it has followed was in evidence. Accordingly, *Degree of Compliance with Data Standards* = 0 for this source.

BP. When researching BP's IQ/DQ standards, we noted it has adopted an approach called Data Quality Management (DQM) [18], which we classify as a companywide IQ standard. Accordingly, *Degree of Compliance with Data Standards* = 0.5 for this source.

OPEC. We searched for "information quality" and "data quality" standards on OPEC's homepage www.opec.org as well as on Google. We have not found descriptions about specific IQ framework or standards OPEC has adopted. OPEC seems to have not established an IQ/DQ standard yet. However, there were efforts in the organization to tackle IQ/DQ problems in their oil data [19]. This demonstrates

OPEC's IQ awareness. Therefore, we rated *Degree of Compliance with Data Standards* = 0.25 for this source.

CIA. We searched for "information quality," "data quality" and "quality standards" on its homepage http://www.cia.gov as well as on Google. No description of specific information quality standards the CIA has followed was in evidence. However, the CIA, which is a U.S. federal government agency, is assumed to follow the U.S. national IQ guidelines issued by the Office of Management and Budget of the White House [3]. Since this was not specifically stated on its website, however, compliance is less certain. Therefore we gave the CIA half credit for national/international standards. Accordingly, *Degree of Compliance with Data Standards* = 0.5 for this source.

Titi Tudorancea Bulletin. We searched for "information quality" and "data quality" standards on its homepage http://www.tititudorancea.com as well as on Google, but no description of specific information quality standards the Titi Tudorancea Bulletin has followed was in evidence. Accordingly, *Degree of Compliance with Data Standards* = 0 for this source.

Expertise of Data Source. Oil reserve figures are highly processed by human actors, not automatically recorded instrument readings. Hence people are critical to the creation of high quality oil reserve information. Qualifications and experience of the relevant staff are thus an indication of the quality of the information an organization produces. To assess this dimension, we determined if the data source is backed by the expertise or qualifications needed for their data to be considered of high quality.
We use the following formula:

$$\text{Degree of Expertise of Data Source} = a + b + c + d \quad (8)$$

Where the data source being assessed for IQ has staff with the following stated qualifications or credentials:

a) Mathematical and or survey statisticians with advanced degrees: 0.25,
b) Graduates in IQ and or Certified IQ Professionals (CIQP): 0.25,
c) Petroleum engineers and or geologists with advanced degrees: 0.25,
d) Economists and or operations research analysts with advanced degrees: 0.25.

Expertise of Data Source for specific sources.
Wired.com. Based on information on its website, wired.com is a marketing company that advertises various data. It also provides various services. We looked for wired.com's employee qualifications and credentials. No credentials or qualifications matching eq. (8) were found. Wired.com does not claim professional expertise in the oil industry. Therefore, the quality rating for the dimension of *Expertise of Data Source* = 0 + 0 + 0 + 0 = 0.

EIA. Based on descriptions on job qualifications and requirements on official EIA website [20], EIA is hiring employees with all qualifications listed in eq. (8). EIA claims professional expertise in the oil industry. Therefore, the quality rating for the dimension of *Expertise of Data Source* = 0.25 + 0.25 + 0.25 + 0.25 = 1.

Europe's Energy Portal. We looked for its employee qualifications and credentials. Based on descriptions on its website [21], Europe's Energy Portal claims to have a team of talented and professional energy analysts and researchers. Accordingly, we rated *Expertise of Data Source* = 0.25 + 0 + 0 + 0 = 0.25 for this source.

BP. We looked at its "Career" information at its official website www.bp.com. BP is hiring employees in all qualification categories in eq. (8) [22]. BP claims professional expertise in the oil industry. Therefore, the quality rating for the dimension of *Expertise of Data Source* = 0.25 + 0.25 + 0.25 + 0.25 = 1 for this source.

OPEC. We looked for "Career" information at its official website www.opec.org. Based on information there, OPEC is hiring employees of various qualifications and credentials and strives to meet the standards of its member states [23]. OPEC's Research Division, which has a Data Services Department, has employees with PhD degrees [24]. Based on criteria of eq. (8), OPEC claims professional expertise in the oil industry. Therefore, the quality rating for the dimension of *Expertise of Data Source* = 0.25 + 0 + 0 + 0.25 = 0.50 for this source.

CIA. We looked for "Career" information at its official website www.cia.gov. The CIA hires employees with all the categories or qualifications but petroleum engineers or geologist listed in eq. (8) [25]. Therefore, the quality rating for the dimension of *Degree of Expertise of Data Source* = 0.25 + 0.25 + 0 + 0.25 = 0.75 for this source.

Titi Tudorancea Bulletin. We looked for information on its career and employee requirements. Titi Tudorancea is a company that markets various data and also provides various other services and products. We could not find career credentials or employment qualification requirements that match the criteria of eq. (8) [26]. Titi Tudorancea Bulletin does not claim professional expertise in the oil industry. Therefore, the quality rating for the dimension of *Degree of Expertise of Data Source* = 0 + 0 + 0 + 0 = 0 for this source.

Degree to which Data was Vetted. Peer acceptance of data can help enhance the confidence of data users in decision making. Furthermore, number of citations to an information source can be expected to correlate with its esteem by both authors and peer reviewers of refereed articles. Thus frequent citation of a source suggests high quality of the source. To assess this dimension, we determined if the data was peer reviewed, and if so, to what extent. We propose to quantify the IQ dimension of *Degree to which Data was Vetted* by normalizing the numbers of citation counts returned on Google Scholar [27] citation search for each of the data sources. We use the following formula:

Degree to which Data was Vetted = Normalized source citation counts on Google Scholar

$$= \frac{\text{number of citation count of source}}{\text{total number of citations for all sources}} \quad (9)$$

As shown in eq. (9), *Degree to which Data was Vetted* is calculated by dividing the number of citation of source by the total number of citations of all information sources being assessed.

Degree to which Data was Vetted for specific sources.
Wired.com. Multiple users commented on its data, but without in-depth critiquing and analysis about the quality of its data and its sources. We searched for citation counts for the terms "www.wired.com" and "World Oil Reserves" on Google Scholar. The hits returned were 9. The normalized figure for this result is 0.003, so *Degree to which Data was Vetted* = 0.003 for this source.

EIA. We searched for citation counts for the terms "Energy Information Administration" and "World Oil Reserves" on Google Scholar. The hits returned were 610. The normalized figure for this result is 0.198, so *Degree to which Data was Vetted* = 0.198 for this source.

Europe's Energy Portal. We searched for citation counts for the terms "www.energy.eu" and "Oil Reserves" on Google Scholar. The hits returned were 18. The normalized figure for this result is 0.006, so *Degree to which Data was Vetted* = 0.006 for this source.

BP. We searched for citation counts for the terms "BP" and "World Oil Reserves" on Google Scholar. The hits returned were 869. The normalized figure for this result is 0.282, so *Degree to which Data was Vetted* = 0.282 for this source.

OPEC. We searched for citation counts for the terms "OPEC" and "World Oil Reserves" on Google Scholar. The hits returned were 1390. The normalized figure for this result is 0.450, so *Degree to which Data was Vetted* = 0.450 for this source.

CIA. We searched for citation counts for the terms "Central Intelligence Agency" and "World Oil Reserves" on Google Scholar. The hits returned were 190. The normalized figure for this result is 0.062, so *Degree to which Data was Vetted* = 0.062 for this source.

Titi Tudorancea Bulletin. We searched for citation counts for the terms "Tudorancea" and "World Oil Reserves" on Google Scholar. The hit returned was 0. The normalized figure for this result is 0, so *Degree to which Data was Vetted* = 0 for this source.

Summary. Table 1 summarizes our IQ assessments and calculates an overall IQ rating for each of the information sources.

- ΣIQ represents the overall information quality of an information product. The ΣIQ of each source is calculated by summing the IQ ratings of all IQ dimensions being assessed for the source.
- The Normalized ΣIQ for each source is calculated by dividing its corresponding ΣIQ by the sum of ΣIQ of all information sources being assessed. Thus, the normalized ΣIQ values add up to 1.

Information Quality (IQ)	**Information Sources**						
Information Quality Dimensions	**EIA**	**CIA**	**BP**	**OPEC**	**Energy**	**Wired**	**Titi Tudorancea**
Data Decay	1.0000	0.9670	0.9330	0.9330	0.9330	0.8670	0.9670
Data Integrity and Reputation	0.8000	0.7000	0.8000	0.8000	0.7000	0.8000	0.3000
Data Coverage and Completeness	1.0000	1.0000	0.2300	0.2110	0.2110	0.0960	0.4640
Degree of Compliance with Data Standards	1.0000	0.5000	0.5000	0.2500	0.0000	0.0000	0.0000
Degree of Expertise of Data Source	1.0000	0.7500	1.0000	0.5000	0.2500	0.0000	0.0000
Degree to which Data was Vetted	0.1980	0.0620	0.2820	0.4500	0.0060	0.0030	0.0000
ΣIQ	**4.9980**	**3.9790**	**3.7450**	**3.1440**	**2.1000**	**1.7660**	**1.7310**
Normalized ΣIQ	**0.2329**	**0.1854**	**0.1745**	**0.1465**	**0.0978**	**0.0823**	**0.0807**
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**

Table 1. Summary of IQ assessments. Lower rank numbers are better.

Based on our IQ framework, Table 1 ranks Wired, EIA, Europe's Energy Portal, BP, OPEC, the CIA and the Titi Tudorancea Bulletin in terms of overall composite IQ ratings of their oil reserve information products. This can provide useful insight and guidance to oil data consumers in decision making using these information sources.

VALIDATING THE INFORMATION QUALITY FRAMEWORK

It would be convenient indeed if the Earth had its own dipstick. One could simply pull it out and read it, rank the information sources according to how closely they match the reading, and check that gold standard against the ranks given to the information sources in Table 1. However, reality demands validating our ranking by other means. We might then contemplate the reality-based approach of waiting enough decades, observing past production figures from that future vantage point, and determining what the reserves were based on the acid test of actual production. Then at that future point we would be able to see which information sources were best in hindsight, and validate (or not) the information quality ranking of Table 1. Unfortunately by that time the ranking would have long outlived its usefulness. Luckily, more plausible approaches to validation exist. A panel of experts could be convened and a Delphi Method-based approach employed to see how well the ranking of Table 1 matches expert concensus. If it matches well enough, the approach of this paper would be validated, making it plausible to use for related problems for which getting expert concensus is impractical, slow, or expensive. A related validation approach would be to present the results to experts and tabulate the degree to which they agree with them. Similarly, partial validation would follow if experts at least found the results useful and interesting. Then the results would be shown to help experts though not replace them.

For this paper, we addressed validation by examining if the results are robust to changes in the relative importance of the different information quality dimensions we used. If robust, the results merit greater credence than if they are sensitive, changing greatly when problem parameters are varied by, for example, changing the importances of the dimensions.

McGilvray [6] provides guidance in determining which dimensions are more important than which others, by proposing a best practice of assessing IQ dimensions based on the following two strategies.

1) Use only dimensions that can be reasonably assessed and for which there is a necessity for assessing the data.
2) List and rank the selected IQ dimensions using a scale based on possible benefits of the dimensions.

With these guidelines in mind, the IQ dimensions selected for this project were weighted using a 1 to 6 weighting scale based on their relative importance. The resultant weight vector was then used to compute a composite, overall quality for the oil reserve information provided by each source.

Ranking the Dimensions

The dimension *Data Integrity and Reputation* was deemed more important than the dimension *Degree of Compliance with Data Standards* because the former enables users to trust the data. Otherwise, data consumers would have insufficient reason to use the data in the first place. The latter is comparatively less important because its benefits depend on the former. By demonstrating the professionalism of the data source, this dimension adds value but only if the data itself is trustworthy. (Otherwise, "garbage in, garbage out.")

The dimension *Degree of Expertise of Data Source* is less important than *Data Integrity and Reputation* because it is auxiliary to it: it contributes to integrity and reputation but cannot substitute for it, as expertise does not necessarily imply trustworthiness or reliability. On the other hand, *Degree of Expertise* was judged more important than *Degree of Compliance with Data Standards* because it is more fundamental: expert data can be made more compliant, but compliant data is not necessarily backed up by expertise.

The dimension of *Data Decay* was assessed of lesser importance than the three dimensions previously discussed because, while users will likely want to know how current their data is, they would be more concerned with their data's reputation, standards and expertise.

The dimension of *Data Coverage and Completeness* should be even lower in importance than *Data Decay* because the user might want to know if their information product includes all data items they would need, but most likely only after being satisfied with the timeliness (i.e. lack of excessive decay) of their information product.

The dimension of *Degree to which Data was Vetted* should be 6th in importance because, before making a final decision about the overall quality of the information product, consumers might want to investigate how other users are evaluating the information product and its source, as a check on their own analysis based on the other 5 dimensions. They would likely find it advisable to use the citation rates by others as a check on their own evaluations, rather than as a substitute, or why do an information quality analysis at all?

The above analysis ranks the six IQ dimensions in our framework. We defined *raw weight* as the inverse of rank, and normalized the weights so that they add up to 1, as summarized in Table 2.

Rank	IQ Dimensions	Raw Weight Vector	Normalized Weight
1	Data Integrity and Reputation	6	**0.29**
2	Degree of Expertise of Data Source	5	**0.24**
3	Degree of Compliance with Data Standards	4	**0.19**
4	Data Decay	3	**0.14**
5	Data Coverage and Completeness	2	**0.10**
6	Degree to which Data was Vetted	1	**0.05**

Table 2. IQ dimension ranks and weights.

Table 3 and Figure 1 together summarize the IQ assessment and IQ dimension weight factor analysis for the oil reserve information sources.

Information Quality	EIA	BP	CIA	OPEC	Energy	Wired	Titi Tudorancea
ΣIQ (from Table 1)	**5**	**3.75**	**3.98**	**3.14**	**2.1**	**1.77**	**1.73**
Normalized ΣIQ (from Table 1)	**0.2329**	**0.1745**	**0.1854**	**0.1465**	**0.0978**	**0.0823**	**0.0807**
Weighted ΣIQ	**0.9047**	**0.7305**	**0.7101**	**0.57**	**0.4132**	**0.3617**	**0.268**
Normalized Weighted ΣIQ	**0.2285**	**0.1846**	**0.1794**	**0.144**	**0.1044**	**0.0914**	**0.0677**

Table 3. Summary of IQ assessments.

Weighted ΣIQ and Normalized Weighted ΣIQ are defined as follows.

- The Weighted ΣIQ for a source is calculated by multiplying its rating on each dimension (from Table 1) by the Normalized Weight for that dimension in Table 2, then summing the resulting terms.

- The Normalized Weighted ΣIQ for each source is calculated by dividing its corresponding Weighted ΣIQ by a constant so that the Normalized Weighted ΣIQ values for all the sources add up to 1.

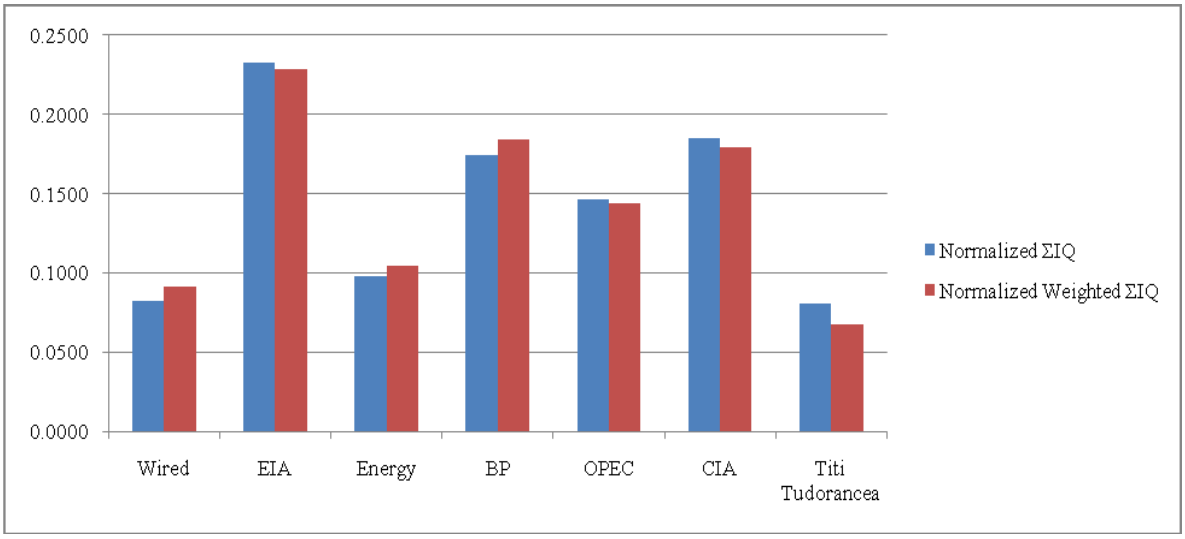

Figure 1. Normalized ΣIQ and Normalized Weighted ΣIQ for oil reserve information sources.

Figure 1 compares weighted and un-weighted ΣIQs. The results in both cases are clearly similar. We further tested this apparent similarity phenomenon using five different weight vectors, as shown in Figure 2 and Table 4. The resulting rankings of information sources in terms of *Normalized Weighted ΣIQ* remained similar to the rankings obtained using the Raw Weight Vector of Table 2, as shown in Figure 2. This generally validates our ranking of the IQ dimensions.

IQ Dimensions	Raw Weight Vector	Weight Vector 2	Weight Vector 3	Weight Vector 4	Weight Vector 5	Weight Vector 6
Data Integrity and Reputation	6	1	2	5	4	2
Degree of Expertise of Data Source	5	2	6	3	1	6
Degree of Compliance with Data Standards	4	3	4	6	5	4
Data Decay	3	4	1	2	3	1
Data Coverage and Completeness	2	5	3	1	2	3
Degree to which Data was Vetted	1	6	5	4	6	5

Table 4. The different weight vectors. "Raw Weight Vector" is from Table 2.

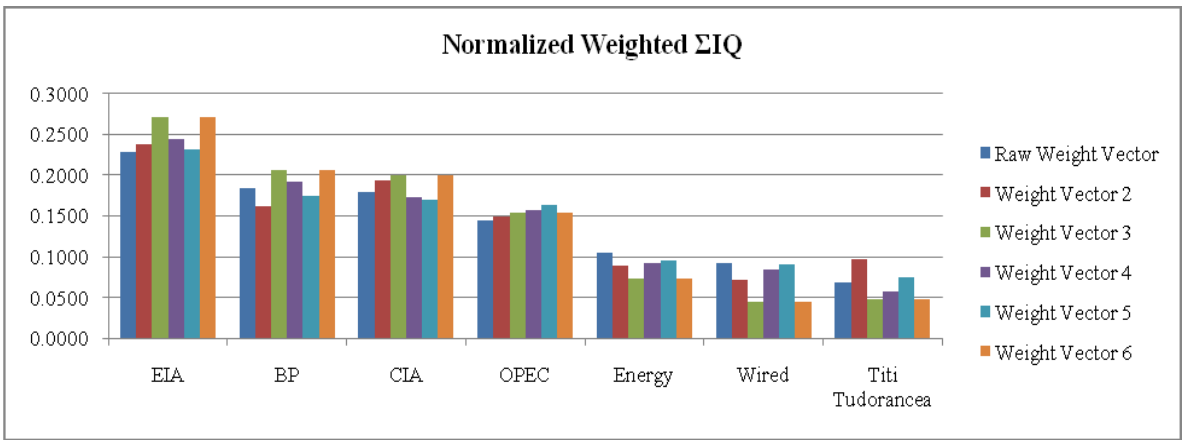

Figure 2. Normalized Weighted ΣIQ for six random weight vectors

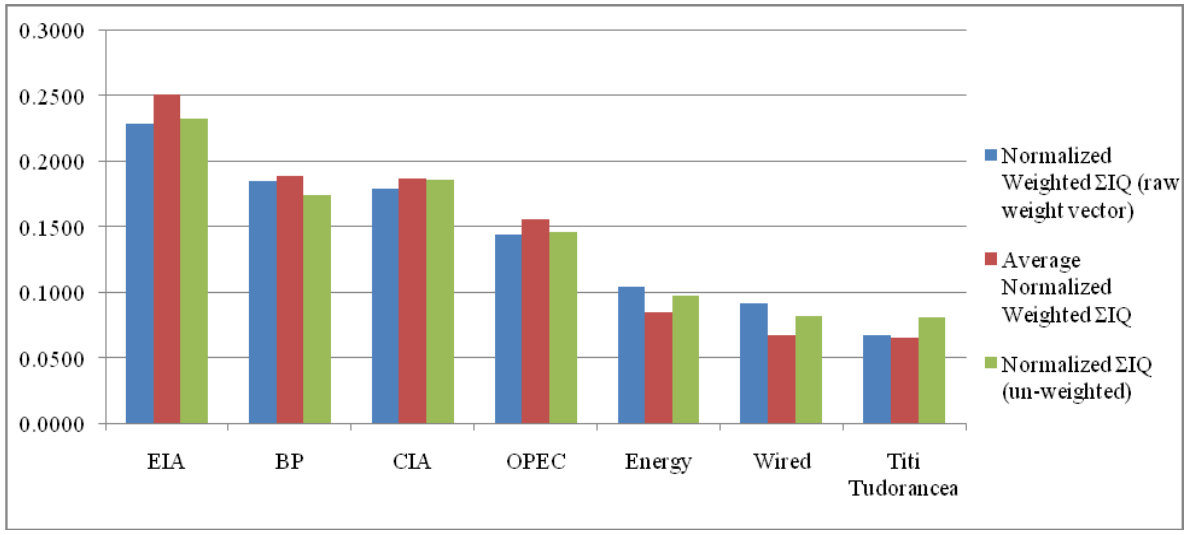

Figure 3. Comparison of unweighted and weighted conditions, along with the results of averaging five diverse incorrect weight vectors, to illustrate the robustness of the results to weightings.

As a further validation test, we averaged the heights of the histogram bars for each source in Figure 2 for Weight Vectors 2 through 6. Figure 3 shows these averages in comparison to the two other conditions discussed. The results are similar for all three conditions, further validating the IQ framework proposed here.

CONCLUSION

In this research, we proposed an information quality framework and formulas to objectively quantify the information quality of world oil reserves data from various sources. This framework is based on several information quality dimensions relevant to oil reserves data. The framework was applied to calculating the IQ ratings of oil reserves information products from several sources. The work draws importance in part from the importance of oil reserves data, which derives from the key role of oil in modern society.

This research has further value in providing a framework for oil data consumers to objectively assess the information qualities of information products from various information sources in order to assist their decision making.

ACKNOWLEDGEMENTS

We express our gratitude to Elizabeth Pierce for valuable suggestions and input to this research. This research was supported in part by funding from the United States Air Force Research Laboratory at Wright Patterson Air Force Base, Dayton, Ohio.

REFERENCES

[1] Fisher, C., Lauria, E., Smith, S. C., & Wang, R. *Introduction to Information Quaity.* s.l. : MIT Information Quality Program, 2008.

[2] Reuters. Venezuela: Oil reserves surpasses Saudi Arabia's. *Ahram Online.* [Online] January 16, 2011. [Cited: June 28, 2011.] http://english.ahram.org.eg/NewsContent/3/14/4060/Business/Markets--Companies/Venezuela-Oil-reserves-surpasses-Saudi-Arabias.aspx.

[3] Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated . *The White House Office of Management and Budget.* [Online] January 2002. [Cited: May 25, 2011.] http://www.whitehouse.gov/omb/fedreg_reproducible/.

[4] Information Quality Guidelines. *U.S. Energy Information Administration.* [Online] 2009. [Cited: May 25, 2011.] http://tonto.eia.doe.gov/abouteia/information_quality_guidelines.cfm.

[5] *A Product Pespective on Total Data Quality Managment.* Wang, Richard Y. 2, February 1998, Communications of the ACM, Vol. 41.

[6] McGilvray, Danette. *Executing Data Quality Projects.* Burlington : Morgan Kaufmann Publishers, 2008.

[7] Yang W. Lee, Leo L. Pipino, James D. Funk, Richard Y. Wang. *Journey to Data Quality.* 1st. Cambridge : The MIT Press, 2006. p. 59.

[8] Squatriglia, Chuck. Who's Gushing Now: World Oil Reserves. *wired.com.* [Online] 2011. [Cited: January 31, 2011.] http://www.wired.com/special_multimedia/2008/oilreserves.

[9] International Energy Statistics. *U.S. Energy Information Administration.* [Online] 2011. [Cited: June 3, 2011.] http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=5&pid=57&aid=6.

[10] Oil Reserves by Country at the End of 2009. *Europe's Energy Portal.* [Online] 2009. [Cited: June 3, 2011.] http://www.energy.eu/stats/energy-oil-reserves.html.

[11] BP Statistical Review of World Energy. *BP Global.* [Online] 2010. [Cited: June 3, 2011.] http://www.bp.com/liveassets/bp_internet/globalbp/globalbp_uk_english/reports_and_publications/statistical_energy_review_2008/STAGING/local_assets/2010_downloads/statistical_review_of_world_energy_full_report_2010.pdf.

[12] World Proven Crude Oil Reserves by Country, 1960-2009. *Organization of the Petroleum Exporting Countries.* [Online] 2009. [Cited: June 3, 2011.] http://www.opec.org/library/Annual%20Statistical%20Bulletin/interactive/2009/FileZ/XL/T31.HTM.

[13] The World Factbook Country Comparison Oil - Proved Reserves. *The Central Intelligence Agency.* [Online] 2010. [Cited: June 3, 2011.] https://www.cia.gov/library/publications/the-world-factbook/rankorder/2178rank.html.

[14] World Crude Oil Proved Reserves. *The Titi Tudorancea Bulletin.* [Online] 2010. [Cited: June 3, 2011.] http://www.tititudorancea.com/z/ies_world_crude_oil_proved_reserves.htm.

[15] *PageRank for Ranking Authors in Co-citation Networks.* Ding, Ying, et al. 11, 2009, Journal of the American Society for Information Science and Technology, Vol. 60, pp. 2229–2243.

[16] Page Rank Checker. *Page Rank Checker.* [Online] 2011. [Cited: May 19, 2011.] http://www.prchecker.info/.

[17] Yang W. Lee, Leo L. Pipino, James D. Funk, Richard Y. Wang. *Journey to Data Quality.* 1st. Cambridge, Massachusetts : The MIT Press, 2006. p. 56.

[18] Data, Data, Everywhere. *BP Global.* [Online] 2011. [Cited: June 7, 2011.] http://www.bp.com/sectiongenericarticle.do?categoryId=9013437&contentId=7026364.

[19] Joint Oil Data Exercise. *Organization of the Petroleum Exporting Countrie.* [Online] April 25, 2002. [Cited: June 7, 2011.] http://www.opec.org/opec_web/en/1085.htm.

[20] Career Opportunities. *U.S. Energy Information Administration.* [Online] 2011. [Cited: June 7, 2011.] http://www.eia.gov/about/careers/.

[21] Europe's Energy Portal: One-stop Shop for EU-Related Energy Intelligence. *Europe's Energy Portal.* [Online] June 2011. [Cited: June 25, 2011.] http://www.energy.eu/about/.

[22] Search Openings. *BP Gloabal.* [Online] 2011. [Cited: June 7, 2011.] https://careers.bpglobal.com/2057/ASP/TG/cim_advsearch.asp?SID=^kpsxkyjogNwp9Fjbm3ywHbOkMBUKeXuI8h1wl3EGU9swwQdzpDI3vN9Ry4v_slp_rhc_SrzJ&ref=672011194030.
[23] Employment Information. *Organization of the Petroleum Exporting Countries.* [Online] 2011. [Cited: June 8, 2011.] http://www.opec.org/opec_web/en/employment/55.htm.
[24] Research Division. *Organization of the Petroleum Exporting Countries.* [Online] 2011. [Cited: June 25, 2011.] http://www.opec.org/opec_web/en/about_us/154.htm.
[25] Career Opportunities. *The Central Intelligence Agency.* [Online] 2011. [Cited: June 8, 2011.] https://www.cia.gov/careers/opportunities/cia-jobs/index.html#Engineering.
[26] The Titi Tudorancea Bulletin. *The Titi Tudorancea Media Network.* [Online] 2011. [Cited: June 8, 2011.] http://www.tititudorancea.com/.
[27] Google Scholar. *Google Scholar.* [Online] 2011. [Cited: June 25, 2011.] http://scholar.google.com/schhp?hl=en&tab=ws.