clarifying the results in this paper is greatly appreciated. The author also expresses his gratitude to the reviewers for their helpful comments.

REFERENCES AND NOTES

(1) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983.
(2) Gati, G. *J. Graph Theory* **1979**, *3*, 95.
(3) Read, R. C.; Corneil, D. G. *J. Graph Theory* **1977**, *1*, 339.
(4) Weisfeiler, B. On Construction and Identification of Graphs. Lectures in Mathematics; Springer-Verlag: Berlin, 1976.
(5) Shalabi, A. Hessenberg Matrix in Molecular Graph Theory. *Chem. Scr.* **1985**, *25*, 252-256.
(6) Randić, M. Random Walks and their Diagnostic Value for Characterization of Atomic Environments. *J. Comput. Chem.* **1980**, *1*, 386-399 and references cited therein.
(7) Randić, M.; Wilkins, C. L. Graph-Theoretical Approach to Recognition of Structural Similarity in Molecules. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 30-35 and references cited therein.
(8) Bertz, S. H.; Herndon, W. C. The Similarity of Graphs and Molecules. In *Artificial Intelligence in Chemistry*; Pierce, T. H., Hohne, B. A., Eds.; ACS Symposium Series 306, American Chemical Society: Washington, DC, 1986.
(9) Basak, S. C.; et al. *Discr. App. Math.* **1988**, *19*, 17.
(10) See, for example, Randić, M.; et al. *J. Comput. Chem.* **1979**, *3*, 5.
(11) Project SERAPHIM software, IB 1404, National Science Foundation Education, Department of Chemistry, University of Wisconsin, Madison, WI 53706. Details concerning the application of HMO theory can be found in: Greenwood, H. H. *Computing Methods in Quantum Organic Chemistry*; Wiley-Interscience: New York, 1972.
(12) Helpin, W.; Burkart, A. *Tables for Laboratory and Industry*; Wiley Eastern Ltd.: New Delhi, 1979.

The NEIC Organic Analysis Reporting System

JOHNNY LEE,* K. ERIC NOTTINGHAM, and LAURENCE W. STRATTAN

Data Analysis Section, Chemistry Branch, Laboratory Services Division, National Enforcement Investigations Center, Environmental Protection Agency, Denver, Colorado 80225

Received July 2, 1990

The Organic Analysis Reporting System processes data from the Finnigan GC-MS Formaster Data System into a "matrix" report format. The standard report created by the GC-MS system is the "QUAN" report which, though very informative, does not present the data in a format that meets the Agency's needs in explaining analytical results to the lay public. Accordingly, the Organic Analysis Reporting System was created to convert the essential data from the GC-MS reports into a form more understandable by the nonscientist.

INTRODUCTION

The National Enforcement Investigations Center (NEIC) is an investigative unit of the Environmental Protection Agency's (EPA) Office of Enforcement. The NEIC works on cases that involve potential civil and criminal violations of environmental laws. These cases result in litigation that is very often adversarial. We have found that the presentation of final analytical results in a matrix-like format best portrays our findings. The Organic Analysis Reporting System (OARS) is a PC-based system that extracts organic compound data from the Finnigan Formaster Data System and generates such matrix-type reports.

METHODS

The goal of the Organic Analysis Reporting System (OARS) is to present organic compound results in a format that is definitive, complete, pertinent, and compact. The report from the Formaster Data System's QUAN Report, though informative and complete, contains instrumental and analytical conditions that are pertinent to the generation of quality data but superfluous and confusing when the results are presented in court. In an adversarial situation it is often prudent to present only the essential results of the analyses which show a violation of environmental law without including the intricacies of the required analytical chemistry.

A typical Finnigan Formaster Data System produces a report such as the one shown in Figure 1.

This report is important to the case development because it has information which demonstrates that each sample was analyzed by using appropriate analytical procedures. The disadvantage of this report is that it is cumbersome. Using this format, it is very difficult to cross reference the analytical results from one sample with results from other samples. The quantitative results are also reported to three significant figures, implying a precision that may not exist.

The NEIC Organic Analysis Reporting System reformats the information from the Formaster QUAN report into a matrix-type report, like that shown in Figure 2.

This NEIC matrix-type report is easier to read and more informative than the Formaster QUAN report. Seven sample results are presented on one page whereas the Formaster report gives the results of one sample per page. A particular advantage of OARS is the ease of comparison between samples. On a potential hazardous waste site, the extent and amount of contamination from different sampling stations can be easily ascertained. The spread of the contaminants could be of interest to the public as well as the courts.

An example of an actual case where the NEIC data led to a guilty plea by a polluter shows the advantage of matrix-type data presentation.

Investigators in Buffalo, NY, discovered that large amounts of solvents were accumulating in a storm drain under a baseball field. The investigators collected samples through a manhole cover in the playing field and from drums from a plant upstream of the storm drain. The NEIC analyses of the samples from both sources determined that they contained percent level solvents. All the samples exhibited the RCRA characteristic of ignitability when they flashed at less than 20 °C (68 °F) using Method 1020 of SW-846.

The EPA's definition of the characteristics of hazardous waste for a liquid is published in the Code of Federal Regulations (CFR), Title 40, Part 261, Subpart C, Section 261.21, paragraphs (a)(1).

261.21. Characteristics of Ignitability.

(a) A solid waste exhibits the characteristic of ignitability if a representative sample of the waste has any of the following properties:

(1) It is a liquid, other than an aqueous solution containing less than 24% alcohol by volume and has a

page 1

SEMIVOLATILE COMPOUND SINGLE SAMPLE REPORT
NATIONAL ENFORCEMENT INVESTIGATIONS CENTER
DENVER, CO

CASE NO 99X LABORATORY US EPA NEIC CONTRACT SMYTH
SAS NO TEST CO. REPORTED 02/17/90 12:21 CUSTOMER REGION99

LAB SAMP ID SAMPLE ID S-01
LAB QC ID 120489 TYPE
DATA RELEASE AUTHORIZED BY

FILE NAME 99X01551 RECEIVED 12/07/89 METHOD 8270
TUNE DFTPP010390 EXTRACTED 12/10/89 FRACTION SV
STANDARD XSV0103C ANALYZED 01/03/90 21:20 INST X5100
BLANK VERIFIED ANALYST LANG
TAPE/POS BOTTLE

% MOISTURE pH LEVEL MED
(DECANTED) CLEANUP MATRIX SOIL
DIL FACTOR 1 EXTRACT METHOD SOX UNITS MG/KG

SAMPLE: SEMIVOLATILES
CONDITIONS: 30M 0.25U PTE-5, 50C-2MIN, 8/MIN TO 290C, 8 PSIG
99X01551;

CODE	CAS NO	COMPOUND	TYPES	CONC	FLAGS
C450	91-20-3	Naphthalene		1.69	
C470	91-57-6	2-Methylnaphthalene		0.734	
C550	83-32-9	Acenaphthene		0.237	J
C565	132-64-9	Dibenzofuran		0.359	
C590	86-73-7	Fluorene		0.190	J
C615	86-30-6	N-Nitrosodiphenylamine (1)		0.448	J
C640	85-01-8	Phenanthrene		1.01	
C645	120-12-7	Anthracene		0.253	J
C655	206-44-0	Fluoranthene		1.29	
C715	129-00-0	Pyrene		1.17	
C740	218-01-9	Chrysene		0.758	J
C765	205-99-2	Benzo(b and/or k)Fluoranthene		1.42	J
CS20		Nitrobenzene-d5		1.92	
CS25		2-Fluorobiphenyl		3.58	
CS30		Terphenyl-d14		4.32	
CS45		Phenol-d5		2.85	
CS50		2-Fluorophenol		2.70	
CS55		2,4,6-Tribromophenol		3.86	

Figure 1.

flash point less than 60 °C (140 °F), as determined by a Pensky–Martens Closed Cup Tester, using the test method specified in ASTM Standard D-93-79 or D-93-80, or a Setaflash Closed Cup Tester, using the test method specified in ASTM Standard D-3278-78, or as determined by an equivalent test method approved by the Administrator.

The results of the analyses are presented in Figure 3.

The data indicate that certain compounds such as toluene, xylenes, and ethylbenzene were present in the storm drain at

ENVIRONMENTAL PROTECTION AGENCY
National Enforcement Investigations Center

STATION NUMBER	S-01	S-02	S-03	S-04	S-05	S-06	S-07	LOD
SAMPLING DATE	120489	120489	120489	120489	120489	120489	120489	
SAMPLING TIME								
TYPE								
LOD FACTOR	1	50	100	100	100	100	100	
COMPOUND NAME (MG/KG)								
Naphthalene	1.00	186.00	15.00	622.00	103.00	68.00	10.00	2.00
2-Methylnaphthalene	1.00	146.00	7.00	32.00	139.00	35.00	21.00	4.00
Acenaphthene	ND	239.00	17.00	127.00	290.00	74.00	109.00	4.00
Dibenzofuran	ND	154.00	15.00	68.00	173.00	48.00	81.00	4.00
Fluorene	ND	144.00	16.00	83.00	149.00	44.00	80.00	2.00
N-Nitrosodiphenylamine (1)	ND	ND	ND	10.00	ND	ND	ND	10.00
Phenanthrene	1.00	199.00	27.00	116.00	283.00	201.00	153.00	4.00
Anthracene	ND	37.00	6.00	33.00	79.00	52.00	ND	4.00
Carbazole	ND	9.00	ND	ND	16.00	18.00	ND	2.00
Fluoranthene	ND	100.00	20.00	170.00	270.00	160.00	80.00	10.00
Pyrene	ND	70.00	20.00	120.00	200.00	120.00	70.00	10.00
Benzo(a)Anthracene	ND	20.00	10.00	60.00	70.00	40.00	30.00	10.00
Chrysene	ND	20.00	10.00	80.00	70.00	40.00	30.00	10.00
Benzo(b and/or k)Fluoranthene	ND	20.00	10.00	120.00	90.00	70.00	20.00	10.00
Benzo(k and/or b)Fluoranthene	ND	10.00	ND	ND	ND	ND	20.00	10.00
Benzo(a)Pyrene	ND	10.00	ND	50.00	50.00	50.00	20.00	10.00
Indeno(1,2,3-cd)Pyrene	ND	ND	ND	50.00	ND	ND	ND	10.00
Benzo(g,h,i)Perylene	ND	ND	ND	ND	40.00	ND	ND	10.00
Nitrobenzene-d5	2.00	ND	3.00	ND	ND	ND	ND	1.00
2-Fluorobiphenyl	2.00	ND	4.00	ND	ND	ND	ND	1.00
Terphenyl-d14	3.00	ND	4.00	ND	ND	ND	ND	1.00
Phenol-d5	2.00	ND	3.00	ND	ND	ND	ND	1.00
2-Fluorophenol	2.00	ND	3.00	ND	ND	ND	ND	1.00
2,4,6-Tribromophenol	4.00	ND	6.00	ND	ND	ND	ND	1.00

a) LOD is the calculated limit of detection for undiluted samples. Multiply the LOD factor times the list LOD to obtain the sample LOD.
b) ND means not detected.

Figure 2.

ENVIRONMENTAL PROTECTION AGENCY
National Enforcement Investigations Center

STATION NUMBER	02	07	15	16	34	41A	418	
SAMPLING DATE	081188	081188	081188	081188	081188	081188	081188	
TYPE	DRUM	DRUM	DRUM	DRUM	MANHOLE	MANHOLE	MANHOLE	SOLUBILITY
LOD FACTOR	1000	1000	1000	1000	1000	1000	1000	IN WATER
COMPOUND NAME (MG/KG)								
Toluene	300000.	520000.	470000.	330000.	490000.	480000.	520000.	.05%
Methyl ethyl ketone	140000.	140000.	140000.	110000.	360.	620.	320.	27%
Methyl isobutyl ketone	100000.	100000.	100000.	77000.	40000.	40000.	40000.	1.9%
Methanol	32000.	14000.	26000.	51000.	ND	ND	ND	100%
Xylenes	16000.	13000.	42000.	26000.	38000.	36000.	37000.	.00003%
Isopropanol	12000.	5100.	6000.	ND	ND	ND	ND	100%
Ethylbenzene	6000.	5000.	7000.	5000.	9000.	7000.	8000.	.015%
Acetone	5800.	1800.	2400.	4000.	ND	ND	ND	100%
Isobutanol	890.	1400.	3500.	1900.	ND	ND	ND	8.7%
n-Butanol	2500.	1200.	1600.	2900.	ND	ND	ND	7.7%
Ethanol	840.	780.	1600.	1700.	ND	ND	ND	100%

Figure 3.

levels equal to or greater than in drums stored at the plant. One compound, methyl ethyl ketone, decreased significantly. Methanol and isopropyl alcohol was not detectable in the manhole samples. Although the ratio of the solvents did not match exactly, this could be due to the different solubilities of the various compounds in water. Compounds that are highly soluble in water may have been carried away by water running through the storm drain, while other compounds that are not as soluble in water were left behind in the same or higher concentrations.

Presentation of the analytical data in a matrix format could have made this explanation relatively simple and more com-

```
STATION  EXBLANKA
SDATE    062190
STIME    1400              CONC.  0.830   0.084    .008
FILENAME H21BLA            LOD.  10.000  10.000  10.000
LEVEL    M                 DILU.      1      10     100
MATRIX   SOIL
UNITS    MG/KG
COMPCODE C580
COMPNAME Diethylphthalate          AVERAGE  0.82
FLAG                               PER RSD  2.53%
TYPE     BLANK

+ = Next, - = Previous, Q = Quit
> = Next Station, < = Previous Station
```

Figure 4.

prehensible to a scientifically unsophisticated audience in the courtroom.

SOFTWARE

OARS is divided into two parts. The first part of OARS extracts data from the Formaster Data System's QUAN Report. The second part of the system processes the data into a matrix report format.

Part one of OARS is written in Microsoft FORTRAN[1] and is made up of four FORTRAN programs run in a batch.

Table I

field no.	QUAN label	NEIC use	field size	example
*01	SAMPLE ID	station number	12	D4
*02	LAB QA ID	sample date	6	41789
*03	LAB SAMP ID	sample time	4	1200
04	FILE NAME	file name	12	63TD4L2C
05	LEVEL	level	1	M
06	MATRIX	matrix	6	SOIL
*07	DIL FACTOR	dilution	9	4000
08	UNITS	units	6	mg/Kg
*09	TYPE	type	8	AR1
10	FRACTION	fraction	4	VO
*11	CODE	compound code	6	C244
*12	COMPOUND	compound name	30	1,2,4-...
13	CONC	concentration	13	860000
14	FLAG	flags	1	

OARS uses certain key fields from the QUAN report. These Key fields are shown in Table I.

Before Oars can process the data, an ASCII flat file version of the Formaster QUAN report is needed. This can be generated with the Formaster command

```
J(X):QR_ONE
        LISTNAME/L,TESTCODE/N,XFER.PRN
```

In this example, the file XFER.PRN is written in the C:\J\JTM subdirectory. The 14 data fields are extracted from the file using four FORTRAN programs given in the sup-

ENVIRONMENTAL PROTECTION AGENCY
National Enforcement Investigations Center

STATION NUMBER	S-01	S-02	S-03	S-04	S-05	S-06	S-07	LOD
SAMPLING DATE	120489	120489	120489	120489	120489	120489	120489	
SAMPLING TIME								
TYPE	D1-FT1	D50	D10	D100	D100	D100	D100	
LOD FACTOR	0	0	0	0	0	0	0	
COMPOUND NAME (MG/KG)								
Naphthalene	1.00	186.00	15.00	622.00	103.00	68.00	10.00	5.00
2-Methylnaphthalene	1.00	146.00	7.00	32.00	139.00	35.00	21.00	5.00
Acenaphthylene	.	.	.	.	.	.	.	10.00
Acenaphthene	.	239.00	17.00	127.00	290.00	74.00	109.00	5.00
Dibenzofuran	.	154.00	15.00	68.00	173.00	48.00	81.00	5.00
Diethylphthalate	.	.	.	.	.	.	.	10.00
Fluorene	.	144.00	16.00	83.00	149.00	44.00	80.00	5.00
N-Nitrosodiphenylamine (1)	.	.	.	10.00	.	.	.	50.00
Phenanthrene	1.00	199.00	27.00	116.00	283.00	201.00	153.00	5.00
Anthracene	.	37.00	6.00	33.00	79.00	52.00	.	5.00
Carbazole	.	9.00	.	.	16.00	18.00	.	5.00
Di-n-Butylphthalate	.	.	.	.	.	.	.	30.00
Fluoranthene	.	100.00	20.00	170.00	270.00	160.00	80.00	10.00
Pyrene	.	70.00	20.00	120.00	200.00	120.00	70.00	10.00
Butylbenzylphthalate	.	.	.	.	.	.	.	30.00
Benzo(a)Anthracene	.	20.00	10.00	60.00	70.00	40.00	30.00	30.00
bis(2-Ethylhexyl)Phthalate	.	.	.	.	.	.	.	30.00
Chrysene	.	20.00	10.00	80.00	70.00	40.00	30.00	30.00
Di-n-Octyl Phthalate	.	.	.	.	.	.	.	30.00
Benzo(b and/or k)Fluoranthene	.	20.00	10.00	120.00	90.00	70.00	20.00	30.00
Benzo(k and/or b)Fluoranthene	.	10.00	.	.	.	.	20.00	30.00
Benzo(a)Pyrene	.	10.00	.	50.00	50.00	50.00	20.00	30.00
Indeno(1,2,3-cd)Pyrene	.	.	.	50.00	.	.	.	30.00
Dibenz(a,h)Anthracene	.	.	.	.	.	.	.	30.00
Benzo(g,h,i)Perylene	.	.	.	.	40.00	.	.	30.00

a) LOD is the calculated limit of detection for undiluted samples. Multiply the LOD factor times the list LOD to obtain the sample LOD.
b) ND means not detected.

Figure 5.

plementary material (Appendix A). This first program, INCOS1.FOR, extracts compound number, compound name, concentration, and flag. The second program, INCOS2.FOR, extracts date, station number, filename, level, matrix, dilution, and type. The third program, INCOS3.FOR, merges the data from the first two programs. And the fourth program, LOD.FOR, adds the limit of detection from an external table.

The last file created by this part of OARS is INCOS2.DB2. This file is fixed-format, delimited by double quotes. The file is ready for direct importing into Dbase III plus.

The second part of OARS is written in Dbase III plus[2] and is comprised of five separate Dbase programs given in the supplementary material (Appendix B).

The first Dbase program is a Dbase importation to get data into a file named INCOS. This program deletes compounds that have flag = "U", eliminating compounds looked for but not found. As can be seen in Figure 2, compounds that are detected in some samples but not in others are treated differently. This is essential to characterizing a waste site.

The second Dbase program adds results from the acid and base fractions. Many EPA methods for the analysis of semivolatile compounds in water samples call for sequential extraction of basic and acidic compounds. Some compounds may appear in both fractions. Hence the results are summed to give the concentration in the sample.

The third Dbase program merges samples that were analyzed at different dilutions; it can accommodate up to three different dilutions.

The fourth program averages the non-zero concentrations and calculates the percent relative difference or percent relative standard deviation. The analyst is then allowed to review the data on the screen and make corrections as necessary. Manual correction normally entails removing sample concentrations that were diluted below the instrument's limit of detection. Figure 4 shows an edit screen that is used at this stage.

The last program then produces a final report in the matrix format. An example of this final report is shown in Figure 5.

CONCLUSION

There is a difference between chemical data and chemical information. Data becomes information when the observer obtains knowledge. The NEIC OARS is a mechanism by which NEIC scientists can present chemical data to the nonscientist and the nonscientist acquires chemical information.

Supplementary Material Available: Four FORTRAN programs (Appendix A) and five Dbase programs (Appendix B) used in the OARS system (27 pages). Ordering information is given on any current masthead page.

REFERENCES AND NOTES

(1) Microsoft and Microsoft FORTRAN are registered trademarks of Microsoft Corp.
(2) Dbase and Dbase III plus are trademarks of Ashton-Tate.

Computer-Assisted Knowledge Acquisition System for Synthesis Planning

TAKASHI NAKAYAMA

Department of Information Science, Faculty of Science, Kanagawa University, Tsuchiya, Hiratsuka-shi, Kanagawa, 259-12 Japan

Received January 3, 1991

A computer-assisted knowledge acquisition system for synthesis planning (KASP) is described and can be used by organic chemists to generate transforms interactively using a graphics user interface. KASP extracts transforms from a specific reaction database which contains about 30 000 records and some of its features are as follows: (1) It is a semiautomatic transform acquisition system. (2) It has a graphical and interactive language which will facilitate its use by chemists. (3) It has a generic-term dictionary that is helpful to describe chemical structures.

INTRODUCTION

Beginning with OCCS in 1967 by Corey,[1] a number of computer-based synthesis planning systems have been developed during the last two decades. They are divided into two categories, although several ideas have been proposed so far. One category is a logic-oriented approach, and the other is an empirical knowledge-oriented approach. For instance, EROS,[2] SYNGEN,[3], etc. belong to the former category, while LHASA,[4] SECS,[5] etc. belong to the latter one. Some of them are said to be in practical use at several companies. Knowledge-oriented systems in particular are said to be very prospective.[6] Specifically this type of system is based on a knowledge base consisting of empirical reaction rules called transforms. However, it seems that there are several problems impeding the development of a really practical system: First, the system must have enough processing speed; second, it must give results that are as precise as possible; and third, it must be able to process as many different targets as possible. The first problem is concerned with the combinatorial explosions that result from the opportunistic application of the rules (transforms). A second one is concerned with relevancy of the applied transforms. This suggests the importance of the quality of transforms and the selection mechanism. A third is concerned with the quantity of transforms. The second and third problems lead to the building of a knowledge base (transform base) with sufficient quality and quantity. In order to build a transform base, some ideas and practical systems have been proposed and implemented. Two of the practical systems are the well-known CHMTRN[4] and ALCHEM[5] which employ an English-like artificial language as a coding tool for transforms. Some favorable results have been reported concerning the utility of those languages.[7] In fact, however, it is not so easy to build a transform base on a practical scale, and this work is still continuing.[8-12]

In our laboratory, a computer-assisted synthesis planning system based on empirical knowledge (SPEK) is under development, using a reaction database which contains about 30 000 specific reaction data coded from *Organic Syntheses*

0095-2338/91/1631-0495$02.50/0