

From compound words to digital codes, how IUPAC's past and present enable chemistry's future

IUPAC 100 Congress
9 July 2019

Slides posted at evanheplersmith.com/page-cv

Evan Hepler-Smith
Duke University
Department of History

evan.heplersmith@gmail.com
evanheplersmith.com
@ehepler

Thursday, Room 352A: Digital Chemistry & the Lab of the Future

11 JULY 2019

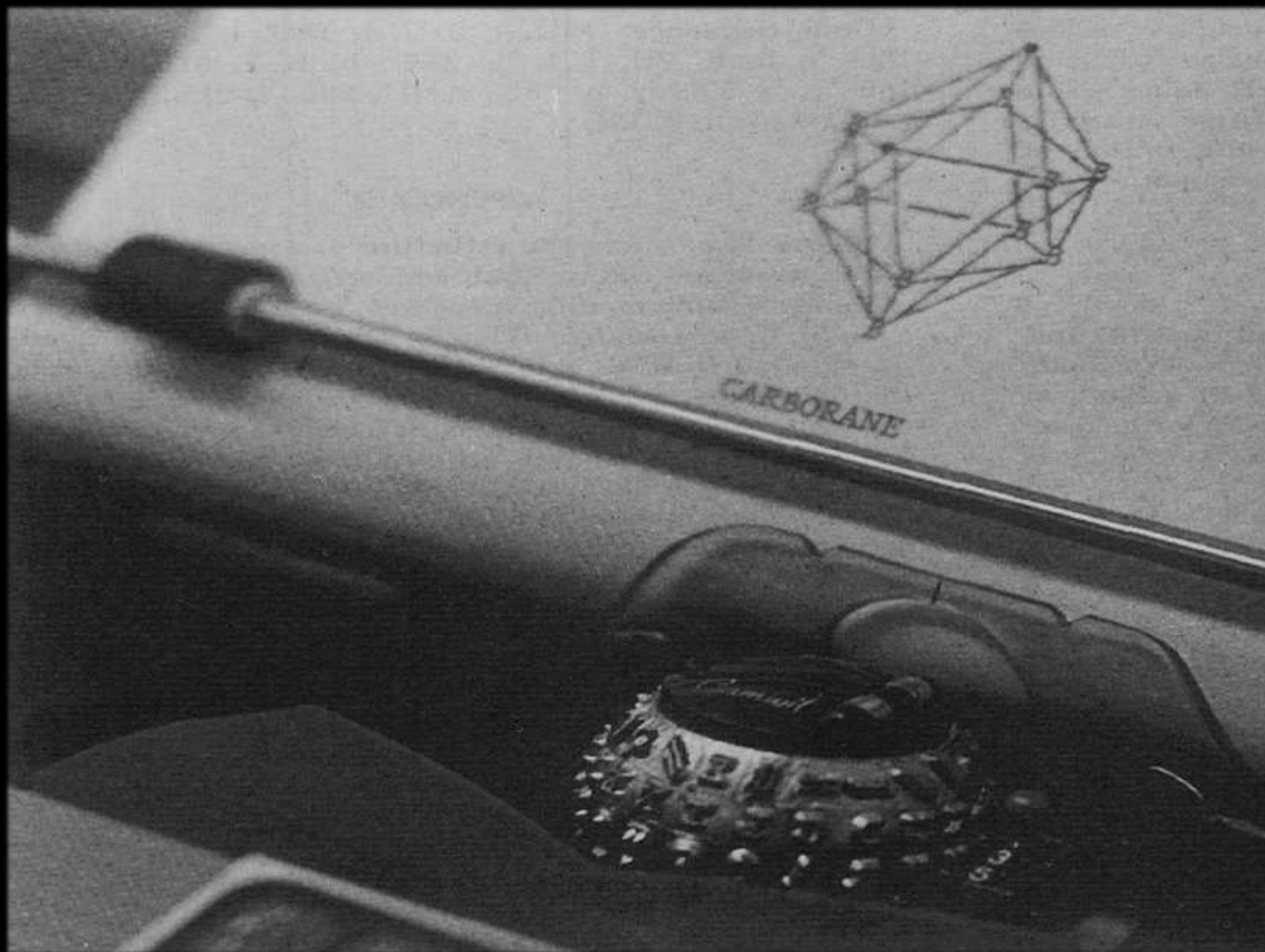
DIGITAL CHEMISTRY AND THE LAB OF THE FUTURE , SYMPOSIUM

S-8.2.1 • 08:30 > 12:30 • ADVANCING FRONTIERS IN DIGITAL CHEMISTRY •

THURSDAY, JULY 11, 2019

DIGITAL CHEMISTRY AND THE LAB OF THE FUTURE , SYMPOSIUM

S-8.2.2 • 2:00 PM > 6:00 PM • QUALITY DATA FOR QUALITY CHEMISTRY •



Ronald Gottardi, "A Modified Dot-Bond Structural Formula Font with Improved Stereochemical Notation Abilities," *Journal of Chemical Documentation* 10 (1970): 75–81, on 79, <https://doi.org/10.1021/c160037a002>.

TABEAU DE LA NOMENCLATURE CHIMIQUE!
PROPOSÉE PAR MM. DE MORVEAU, LAVOISIER, BERTHOLET ET DE FOURCROY.

I.		II.		III.		IV.		V.	
SUBSTANCES NON DÉCOMPOSÉES.		MISES À L'ÉTAT DE GAZ PAR LE CALORIQUE.		COMBINÉES AVEC L'OXYGÈNE.		OXYGÉNÉES GAZEUSES.		OXYGÉNÉES AVEC BASES.	
NOM NOUVEAU, OU ADOPTÉ.	NOM ANCIEN.	NOM NOUVEAU, OU ADOPTÉ.	NOM ANCIEN.	NOM NOUVEAU, OU ADOPTÉ.	NOM ANCIEN.	NOM NOUVEAU, OU ADOPTÉ.	NOM ANCIEN.	NOM NOUVEAU, OU ADOPTÉ.	NOM ANCIEN.
Lumière.									
Calorique.	Chaleur latente, ou mesure de la chaleur.								
Oxygène.	Base de l'air vital.	Gaz oxygène. Noms. Il paraît que la lumière contribue à le mettre en état de gaz.	Air déphlogistiqué, ou air vital.						
Hydrogène.	Base du gaz inflammable.	Gaz hydrogène.	Gaz inflammable.	Eau.	Eau.				
Azote, ou Radical nitrique.	Base de l'air phlogistiqué, ou de la matière atmosphérique.	Gaz azotique.	Air phlogistiqué, ou matière atmosphérique.	Base du gaz nitreux. Acide nitrique. Et avec excès d'oxygène, Acide nitreux.	Base du gaz nitreux. Acide nitreux oléant. Acide nitreux fumant.	Gaz nitreux. Gaz acide nitreux.		Nitrate de potasse, de soude, &c. Nitre de potasse.	Nitre commun. Nitre calciné.
Carbone, ou Radical carbonique.	Charbon pur.			Acide carbonique.	Air fixe, ou Acide crayeux.	Gaz acide carbonique.	Air fixe, air méphitique.	de chaux. Carbonate de potasse, &c. de soude, &c.	Craie. Alcalis effervescens. Rouille de fer, &c.
Soufre, ou Radical sulfurique.				Acide sulfurique. Et avec moins d'oxygène, Acide sulfureux.	Acide viriolique. Acide sulfureux.			Sulfate de potasse, de soude, de chaux, d'alumine, de baryte, de fer, &c.	Tartré viriolé. Sel de Glauber. Sélénite. Alun. Spasme pesant. Vitriol de fer.
						Gaz acide sulfureux.	Gaz acide sulfureux.	Sulfure de potasse, &c.	Sel sulfureux de Stahl.



https://commons.wikimedia.org/wiki/File:Marx_and_Engels.jpg

Karl Marx (1818-1883)
Friedrich Engels (1820-1895)



https://commons.wikimedia.org/wiki/File:Marx_and_Engels.jpg

Karl Marx (1818-1883)

Friedrich Engels (1820-1895)

“Every new aspect of a science involves a revolution in the technical terms of that science. This is best shown by chemistry, where the whole of the terminology is radically changed about once in twenty years, and where you will hardly find a single organic compound that has not gone through a whole series of different names.”

Friedrich Engels, preface to Marx’s *Capital*

Karl Marx, *Capital*, vol. 1, trans. Samuel Moore & Edward B. Aveling (London: S. Sonnenschein, Lowrey, 1887), on 19, <https://www.marxists.org/archive/marx/works/1867-c1/index.htm>



RÉSOLUTIONS

PRÉSES PAR LA

COMMISSION INTERNATIONALE

POUR LA RÉFORME DE LA NOMENCLATURE CHIMIQUE

réunie à Genève du 19 au 22 avril 1892

1. A côté des procédés habituels de nomenclature, il sera établi pour chaque composé organique un nom *officiel* permettant de le retrouver sous une rubrique unique dans les tables et dictionnaires.

La Commission exprime le vœu que les auteurs prennent l'habitude de mentionner dans leurs mémoires, entre parenthèses, le nom officiel à côté du nom choisi par eux.

2. On décide de ne s'occuper, pour le moment, que de la nomenclature des composés de constitution connue, et de remettre à plus tard la question des corps de constitution inconnue.

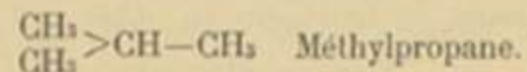
Resolution 1: "An official name will be established for each organic compound, so that it can be looked up under a unique heading in indexes and reference works."

I. Hydrocarbures

3. La désinence *-ane* est adoptée pour tous les hydrocarbures saturés.

4. Les noms actuels des quatre premiers hydrocarbures normaux saturés (*méthane, éthane, propane, butane*) sont conservés; on emploiera les noms tirés des nombres grecs pour ceux qui ont plus de quatre atomes de carbone.

5. Les hydrocarbures à chaîne arborescente sont regardés comme dérivés des hydrocarbures normaux, et on rapportera leur nom à la chaîne normale la plus longue qu'on puisse établir dans leur formule, en y ajoutant la désignation des chaînes latérales.



6. Lorsqu'un radical hydrocarboné est introduit dans une chaîne latérale, on emploiera *métho-, étho-, etc.*, au lieu de *méthyl-, éthyl-,*

RÉSOLUTIONS

PRESES PAR LA

COMMISSION INTERNATIONALE

POUR LA RÉFORME DE LA NOMENCLATURE CHIMIQUE

réunie à Genève du 19 au 22 avril 1892

1. A côté des procédés habituels de nomenclature, il sera établi pour chaque composé organique un nom *officiel* permettant de le retrouver sous une rubrique unique dans les tables et dictionnaires.

La Commission exprime le vœu que les auteurs prennent l'habitude de mentionner dans leurs mémoires, entre parenthèses, le nom officiel à côté du nom choisi par eux.

2. On décide de ne s'occuper, pour le moment, que de la nomenclature des composés de constitution connue, et de remettre à plus tard la question des corps de constitution inconnue.

“What is important is that the official name be the faithful translation of the constitution of the compound, representing it just as the structural formula does.”

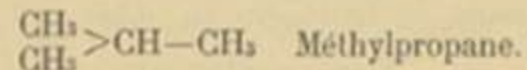
Amé Pictet, “Le congrès international de Genève pour la réforme de la nomenclature chimique,” *Les archives des sciences physiques et naturelles* 27 (May 1892): 485–520, on 491.

I. Hydrocarbures

3. La désinence *-ane* est adoptée pour tous les hydrocarbures saturés.

4. Les noms actuels des quatre premiers hydrocarbures normaux saturés (*méthane, éthane, propane, butane*) sont conservés; on emploiera les noms tirés des nombres grecs pour ceux qui ont plus de quatre atomes de carbone.

5. Les hydrocarbures à chaîne arborescente sont regardés comme dérivés des hydrocarbures normaux, et on rapportera leur nom à la chaîne normale la plus longue qu'on puisse établir dans leur formule, en y ajoutant la désignation des chaînes latérales.



6. Lorsqu'un radical hydrocarboné est introduit dans une chaîne latérale, on emploiera *métho-, étho-,* etc., au lieu de *méthyl-, éthyl-,*

RÉSOLUTIONS

PRÉSES PAR LA

COMMISSION INTERNATIONALE

POUR LA RÉFORME DE LA NOMENCLATURE CHIMIQUE

réunie à Genève du 19 au 22 avril 1892

1. A côté des procédés habituels de nomenclature, il sera établi pour chaque composé organique un nom *officiel* permettant de le retrouver sous une rubrique unique dans les tables et dictionnaires.

La Commission exprime le vœu que les auteurs prennent l'habitude de mentionner dans leurs mémoires, entre parenthèses, le nom officiel à côté du nom choisi par eux.

2. On décide de ne s'occuper, pour le moment, que de la nomenclature des composés de constitution simple, et de laisser en retard la question des corps de constitution complexe.

I. Hydrocarbures

3. La désinence *-ane* est adoptée pour tous les hydrocarbures saturés.

4. Les noms actuels des quatre premiers hydrocarbures normaux saturés (*méthane, éthane, propane, butane*) sont conservés; on emploiera les noms tirés des nombres grecs pour ceux qui ont plus de quatre atomes de carbone.

The PRESIDENT said that we must keep in mind that such systematic names as had been suggested were really names of formulæ rather than names of substances. Thus, a name had been suggested for *acetic acid* on the assumption that *Medicus's formula* correctly represented

6. Lorsqu'un radical hydrocarboné est introduit dans une chaîne latérale, on emploiera *métho-, étho-, etc.*, au lieu de *méthyl-, éthyl-,*

Alexander Crum Brown, quoted in *Proceedings of the Chemical Society*, 16 June 1892.

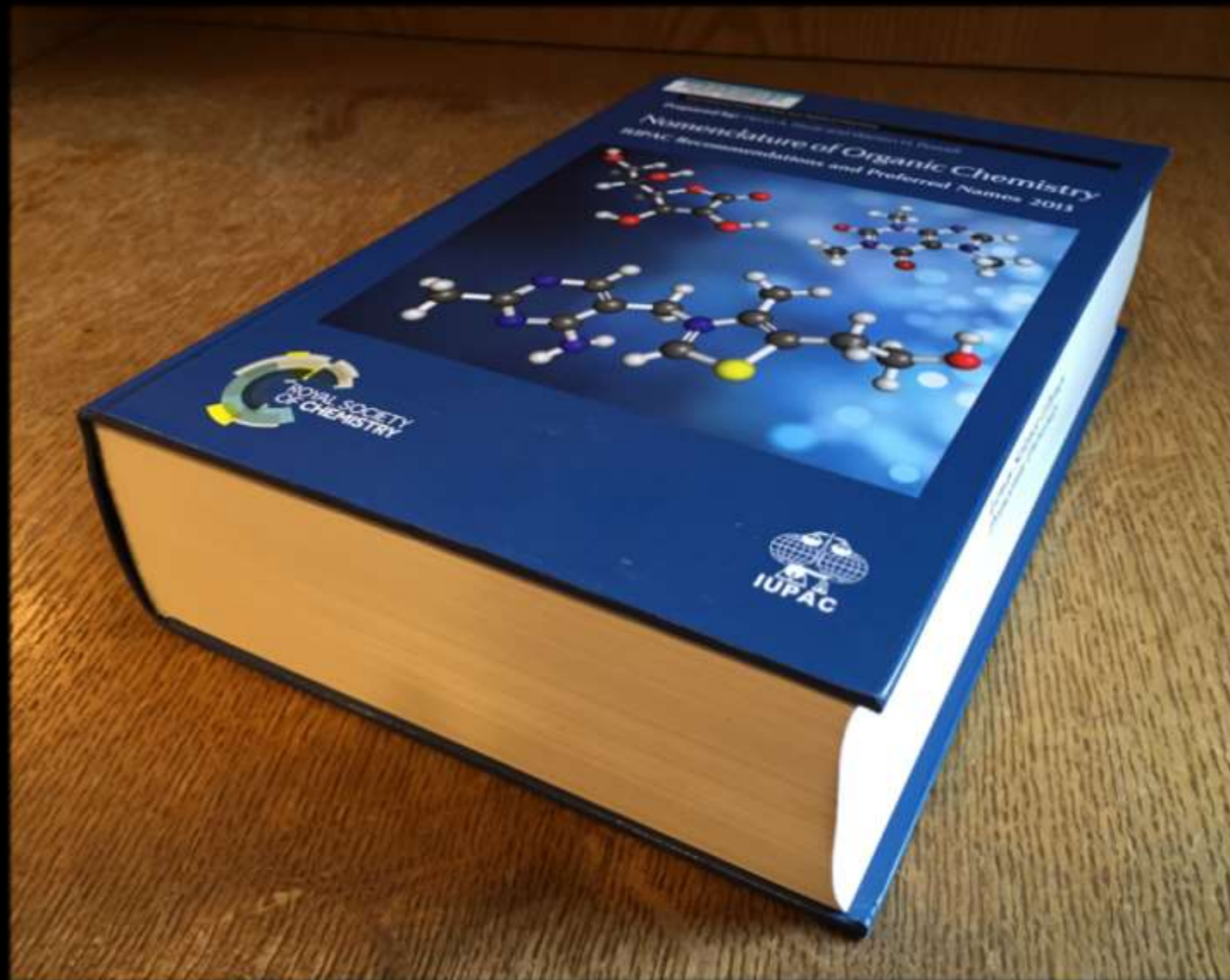
Toast by Charles Moureu, founding president of IUPAC, in April 1919:

“Among all the instruments the chemist uses in his work, there is nothing more indispensable than good books and good bibliographic literature. The chemist can no more do without these than without the balance or the thermometer.”



“La conférence interalliée de la chimie, 14-15-16 Avril,” *Chimie & Industrie* 2 (1919): 569.

**2013 “Blue Book”
(IUPAC Nomenclature
of Organic Chemistry)**





G. Malcolm Dyson (1902-1978)

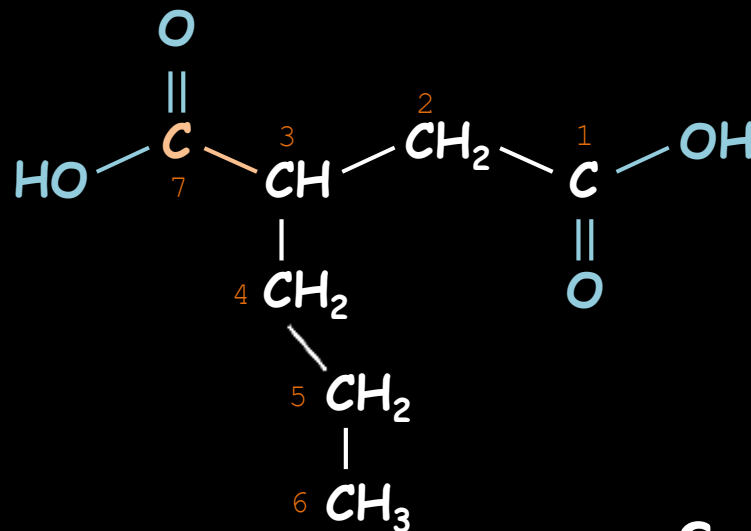
“George Malcolm Dyson,” National
Portrait Gallery, NPG x71342,
<https://www.npg.org.uk/>



**G. Malcolm Dyson
(1902-1978)**

“George Malcolm Dyson,” National
Portrait Gallery, NPG x71342,
<https://www.npg.org.uk/>

Dyson's notation



C6 . 3C . 1, 7X

Compare:

3-methyl-1,3¹-hexanedioic acid

propylsuccinic acid

1,2-pentanedicarboxylic acid

2-propylbutanedioic acid

INTERNATIONAL UNION OF CHEMISTRY

(UNION INTERNATIONALE DE CHIMIE)

Bureau (Comité d'Action)

Paris, 19 & 20 Mars 1947

February 27, 1947

Professor Dr. R. Delaby
4, Avenue de l'Observatoire
Paris (6), France

Dear Mr. Secretary General

This is to let you know that the organization of a new international Commission on Codification, Ciphering, Punched Cards, etc., has been approved, under the presidency of Dr. G. Malcolm Dyson of "Highfields", Ashby Road, Loughborough, Leicestershire, England. He is proceeding immediately to

Marston Bogert to Raymond Delaby, 27 Feb 1947, International Union of Pure and Applied Chemistry (IUPAC) Archive, Othmer Library of Chemical History, Science History Institute, Philadelphia, PA, Box 94.

Robert V. Williams, "Madeline M. Henderson: From Chemical Information Science Pioneer to Architect of the New Information Science," *Libraries & the Cultural Record* 45 (2010): 167-84, on 170.

Meeting of IUPAC Commission on Codification, Ciphering, and Punched-Card Techniques, 1951





**G. Malcolm Dyson
(1902-1978)**

"George Malcolm Dyson," National
Portrait Gallery, NPG x71342,
<https://www.npg.org.uk/>

Dyson on the Dyson cipher:

"A new notation which may go far towards solving those difficulties of chemical nomenclature that, as chemistry advances, become more apparent each year, particularly in classification and indexing."

"A new 'Lexicon of Organic Chemistry' is needed which contains the information of *Beilstein* brought up-to-date and arranged in an immediately accessible form..."

"Such mechanical handling would, of course, be essential to the expeditious compilation of a new 'Lexicon.'

"The cipher has been constructed to be amenable to mechanical computation and sorting."

G. Malcolm Dyson, *A New Notation and Enumeration System for Organic Compounds* (New York: Longmans, Green, 1947), on iii, 5, 7.

**Friedrich Richter,
editor-in-chief of *Beilstein***

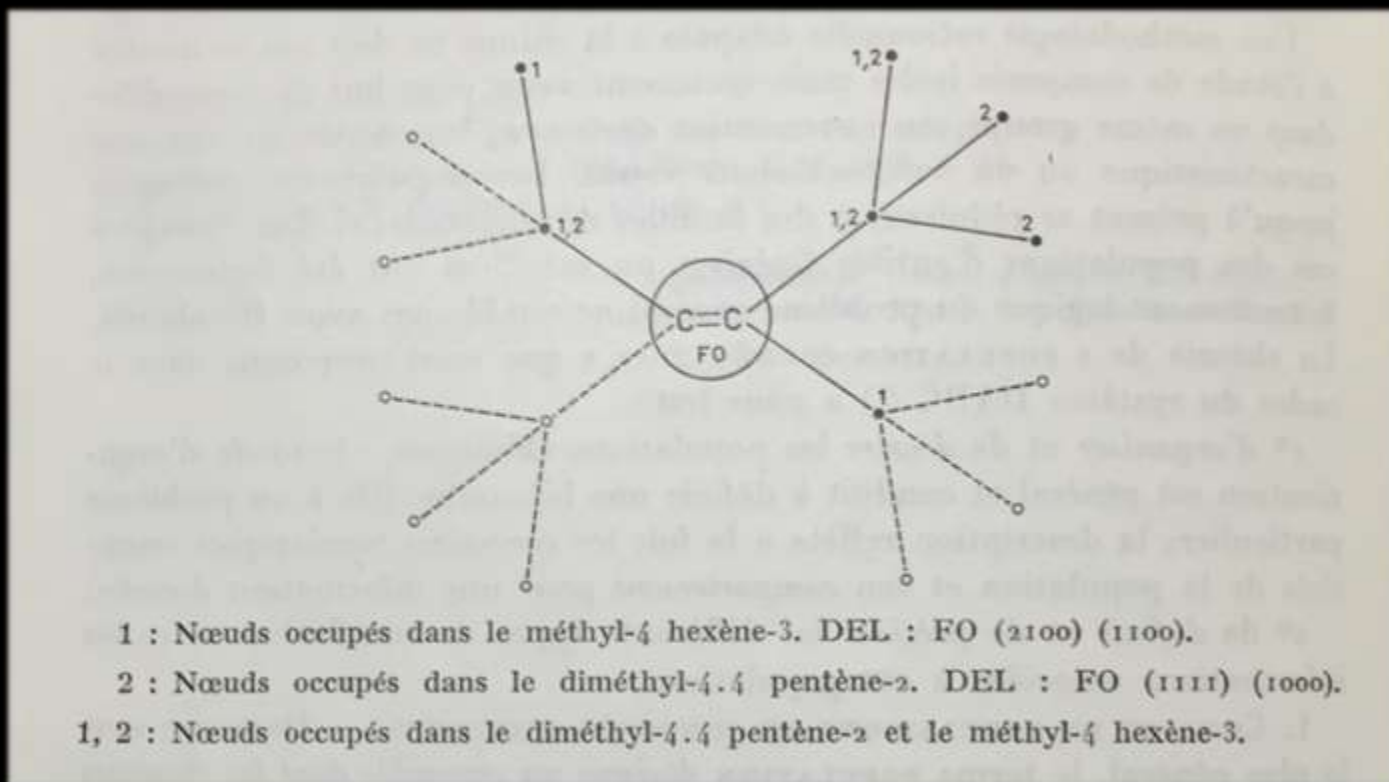


Richter on the Dyson cipher:

“Certain ciphering systems, by giving linear representations ("notations") of structures in terms of (mostly familiar) partial structures, are, to a certain measure, only nomenclature further abbreviated.”

“Any future modification of nomenclature should be in the direction of fortification rather than destruction of such logic as there is in present nomenclature.”

Friedrich Richter, “Basic Features of Nomenclature in Organic Chemistry,” in *Chemical Nomenclature* (American Chemical Society, 1953), 65–74, on <http://dx.doi.org/10.1021/ba-1953-0008.ch008>.



**Structure, substructure, “hyperstructure”:
 “new concepts ... which meet the requirements
 of computer-based modern chemistry.”**

**Jacques-Émile Dubois
 (1920–2005)**

MACHINE DOCUMENTATION IN THE CHEMICAL FIELD
Report to the Bureau of IUPAC

B. The first task of the commission should be directed towards the machine handling of chemical structures and the computer generation of nomenclature.

Byron Riegel, "Machine documentation in the chemical field: Report to the bureau of IUPAC," 4 July 1969, IUPAC Archive (addenda), Science History Institute, Box 91.

B. The first task of the commission should be directed towards the machine handling of chemical structures and the computer generation of nomenclature.

Dubois on committee goals

B. The first task of the commission should be directed towards the machine handling of chemical structures and the computer generation of nomenclature.

Representations Levels	Topological	Fragmentation	Registry Number	Nomenclature
Abstracting and Indexing Center	×		×	×
Analysis Center Data Bank	×	×		
Dissemination Center	×	×		

PROPOSAL FOR

"MINIMAL REQUIREMENTS FOR MACHINE HANDLING OF
CHEMICAL STRUCTURES "

Human Environment

N_s (Systematic Nomenclature)

R (Registry Number)

Computer Environment

T (Topology)

graphic representation

N_c (Ciphred Notation)

Fragment or segment
representation

Jacques-Émile Dubois, "Proposal for minimum requirements for machine handling of chemical structures," January 1971, IUPAC Archive (addenda), Science History Institute, Box 91.

Fred Tate (Chemical Abstracts Service) on committee goals

B. The first task of the commission should be directed towards the machine handling of chemical structures and the computer generation of nomenclature.

October 6, 1971

Tasks of the Inter-Divisional Committee on
Machine Documentation in the Chemical Field

The first task of the committee should be directed towards the machine handling of chemical structures and the computer generation of nomenclature.

There is a real need for a unique definition of chemical structure in natural language which is understandable on the printed page and yet logical and unambiguous to a computer program. The natural language representation

Fred Tate, "Tasks of the Inter-Divisional Committee," 6 Oct 1971, IUPAC Archive (addenda), Science History Institute, Box 91.

“InChI is based on a canonical algorithm that notates chemical structure information in a layered format, the InChI string, with the formula and connectivity at the core. This standard form, based on a normalized structure, enables interoperability between databases.”

Leah McEwen, “InChI’ng Forward: Community Engagement in IUPAC’s Digital Chemical Identifier,” *Chemistry International* 40, no. 1 (2018): 27–31, on 27 <https://doi.org/10.1515/ci-2018-0109>.

“InChI is based on a canonical algorithm that notates chemical structure information in a layered format, the InChI string, with the formula and connectivity at the core. This standard form, based on a normalized structure, enables interoperability between databases.”

Leah McEwen, “InChI’ng Forward: Community Engagement in IUPAC’s Digital Chemical Identifier,” *Chemistry International* 40, no. 1 (2018): 27–31, on 27 <https://doi.org/10.1515/ci-2018-0109>.

CHEMIN(CHI)FORMATICS: IUPAC’S ROLE IN THE ENVIRONMENTAL MONITORING REVOLUTION

CHEMISTRY FOR THE ENVIRONMENT

1. Monitoring Chemicals for a Safer Environment

E SCHYMANSKI ^{1,*}

“High resolution mass spectrometry, informatics and “big data” are upon us and have revolutionized the way we can monitor chemicals in our environment.... InChI and InChIKey contributed, perhaps unintentionally, towards this revolution.”

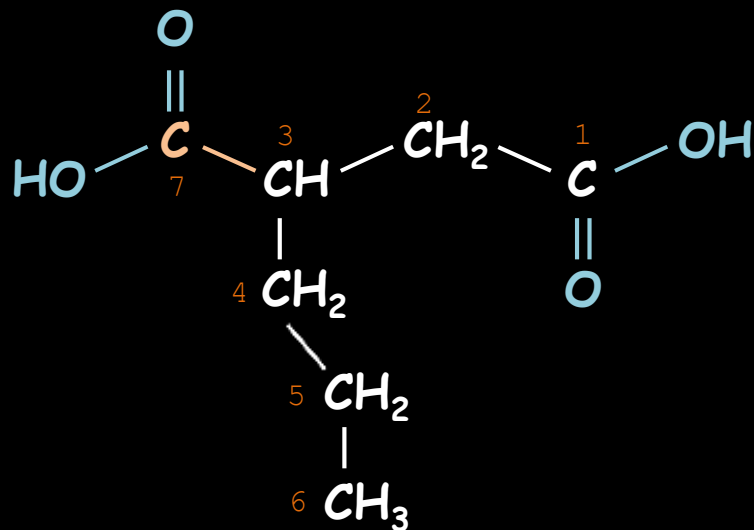
From compound words to digital codes, how IUPAC's past and present enable chemistry's future

IUPAC 100 Congress
9 July 2019

Slides posted at evanheplersmith.com/page-cv

Evan Hepler-Smith
Duke University
Department of History

evan.heplersmith@gmail.com
evanheplersmith.com
@ehepler



C6 . 3C . 1, 7X

InChI=1S/C7H12O4/c1-2-3-
5(7(10)11)4-6(8)9/h5H,2-
4H2,1H3,(H,8,9)(H,10,11)

InChI Key:
QLTZBYGZXPKHLF-UHFFFAOYSA-N

Compare:

3-methyl-1,3¹-hexanedioic acid

propylsuccinic acid

1,2-pentanedicarboxylic acid

2-propylbutanedioic acid