PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Building an easy-to-explore graph database for smoke and/or aerosol constituents, enriched with publicly available toxicological and flavorant properties data

June 14th, 2016

Pavel Pospisil
Philip Morris Products SA
(part of Philip Morris International group of companies)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Reduced-Risk Products ("RRPs") is the term the company uses to refer to products with the potential to reduce individual risk and population harm in comparison to smoking cigarettes.

PMI's RRPs are in various stages of development and commercialization, and we are conducting extensive and rigorous scientific studies to determine whether we can support claims for such products of reduced exposure to harmful and potentially harmful constituents in smoke, and ultimately claims of reduced disease risk, when compared to smoking cigarettes.

Before making any such claims, we will rigorously evaluate the full set of data from the relevant scientific studies to determine whether they substantiate reduced exposure or risk. Any such claims may also be subject to government review and authorization, as is the case in the United States today.

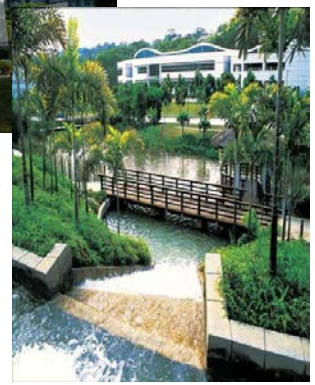
Outlook

- Introducing Philip Morris International R&D
- Collecting data - smoke/aerosol complex matrices
- Organizing data - Graph-based KnowledgeBase
- Exploring data - Advanced graphical visualization
- Graphical Read-Across
- Demonstration

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Philip Morris International R&D

Neuchatel, Switzerland



Singapore

- Over 300 scientists and engineers in the fields of biology, physics, chemistry, electronic engineering, mathematics, computational science, medicine and pharmacy.
- The main focus is to have a positive impact on public health by developing innovative and acceptable RRPs, which are supported by compelling evidence that they are likely to have a positive impact on the health of current adult smokers.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Reduced Risk Products

>800°C

Distillation / Pyrolysis / Combustion

Smoke

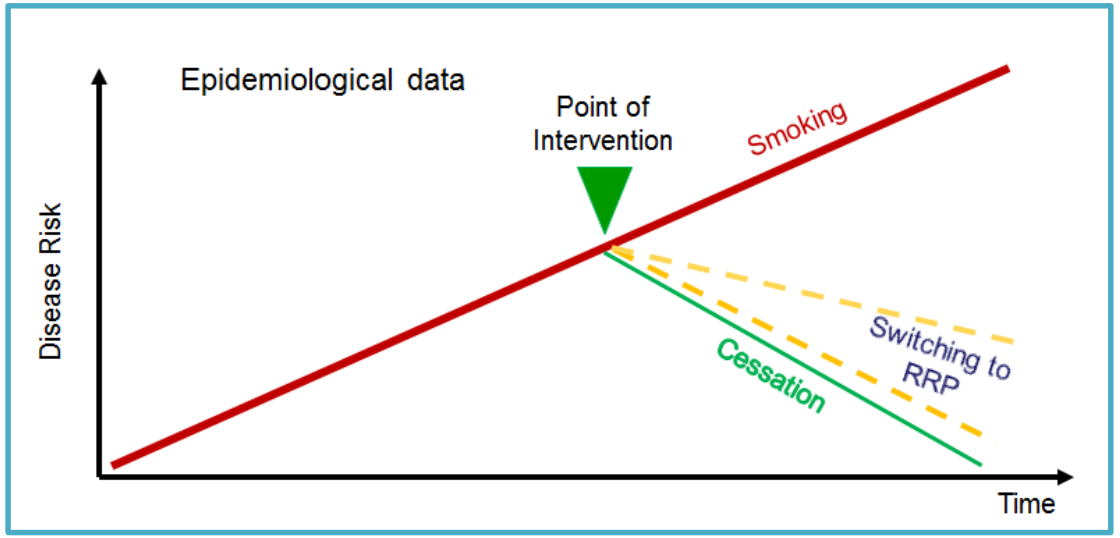
<300°C

Heating not Burning

Aerosol

Some of these constituents are categorized as harmful and potentially harmful (HPHC).

	Platform 1	Platform 2
Heat-not-burn products		
	Platform 3	Platform 4
E-Vapor products		



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

RRPs Aerosol Characterization

Aerosol

Smoking machines generate aerosols that are trapped and collected.

Analytical chemistry

Aerosols are analyzed for HPHCs (regulatory) or assessed using a non-targeted approaches.

Structural Identification

Computational chemistry

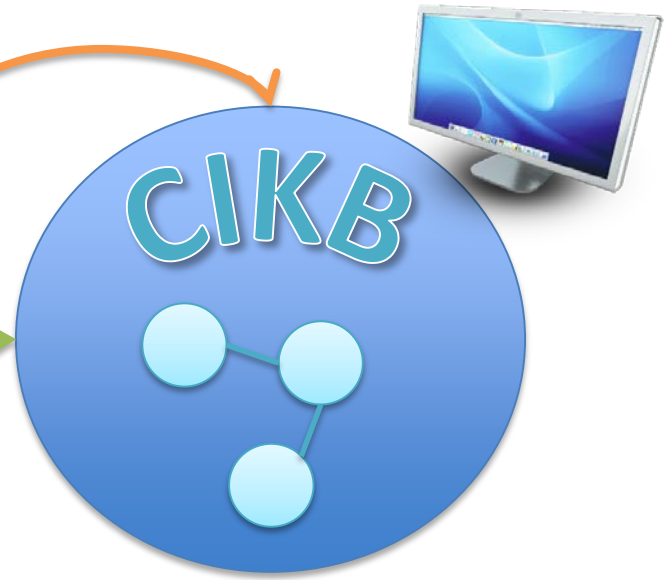
Automation and acceleration of structural identification. Databanking, QSAR modeling and clustering.

Chemical and Spectra Registry

Public Domain

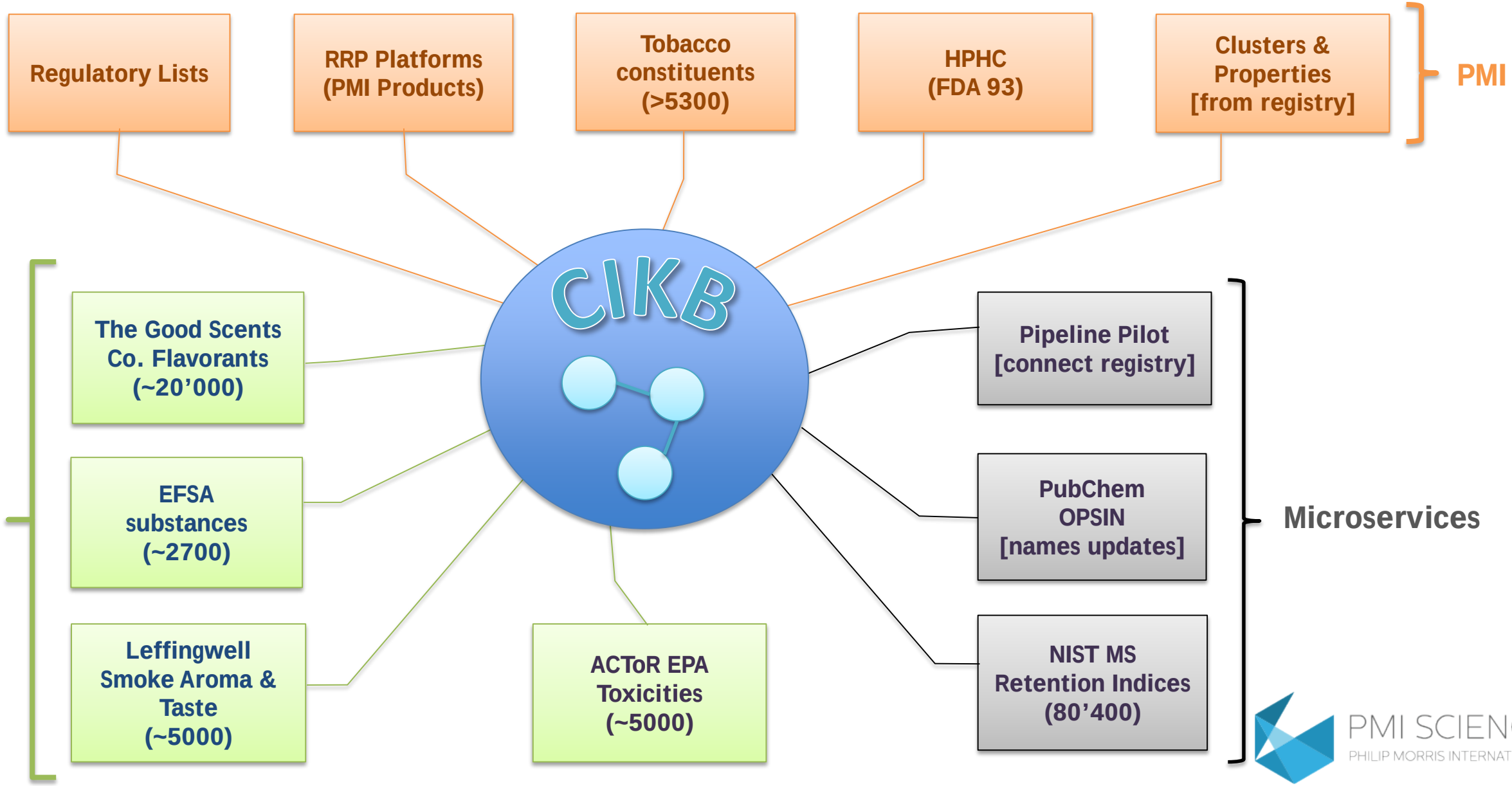
Systems Toxicology, Clinics

Studies to advance scientific and medical knowledge. Clinical and toxicological assessment.



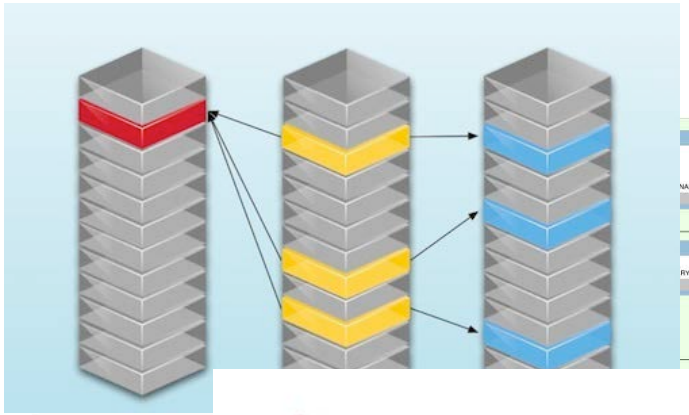
ChemoInformatics KnowledgeBase (CIKB)

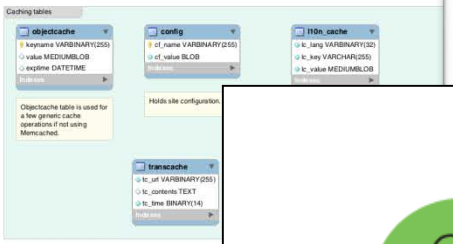
ChemoInformatics KnowledgeBase Data

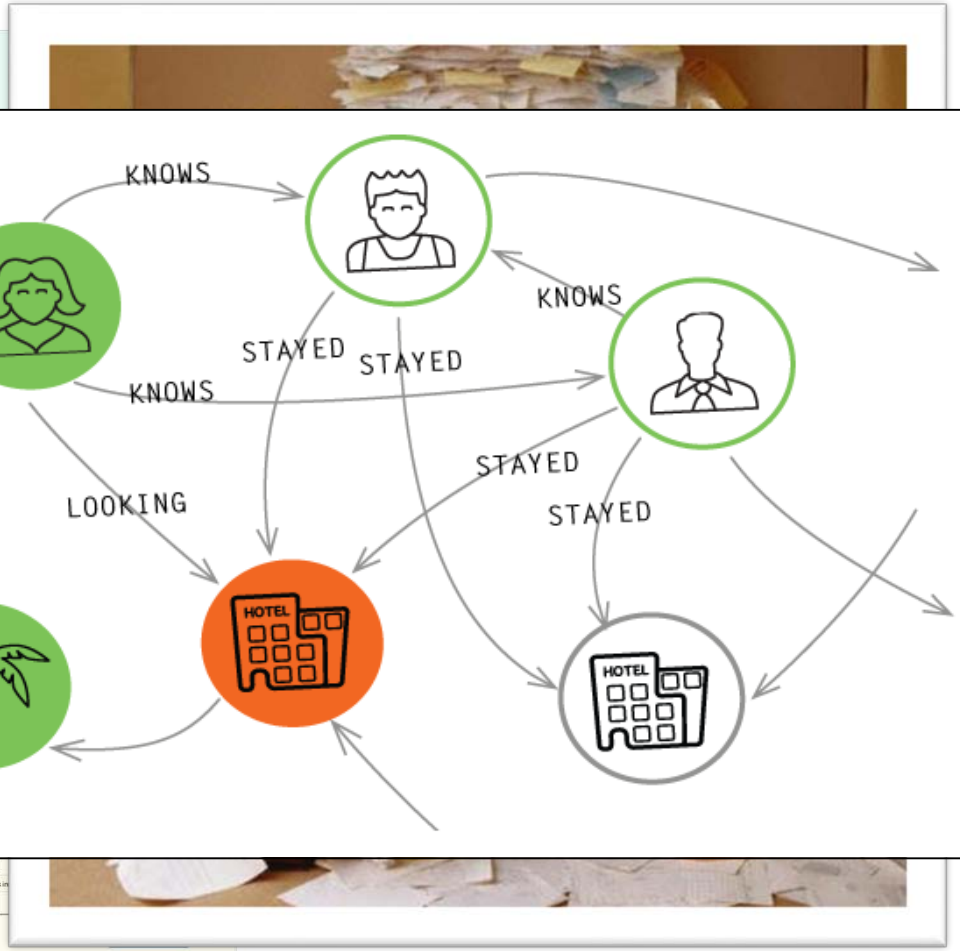


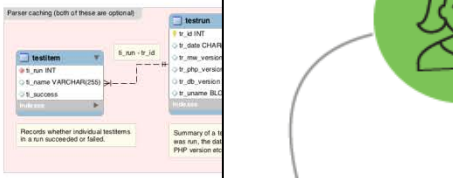
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

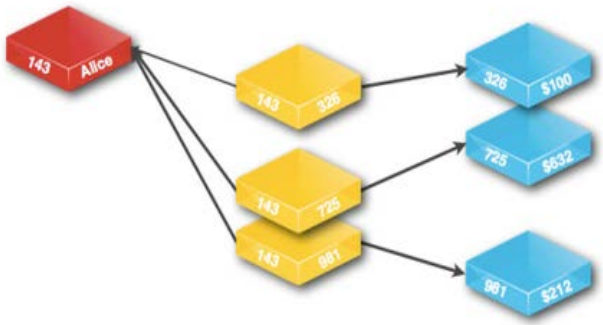
Data Too Complex to Model

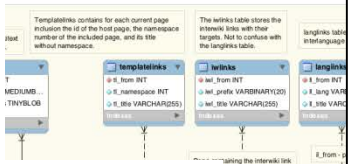


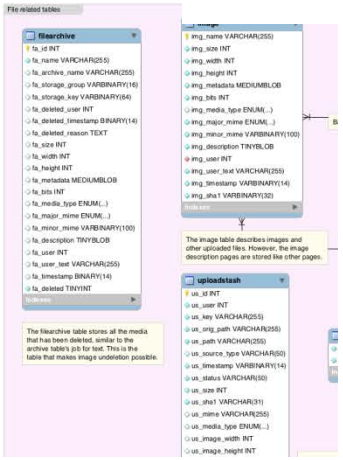


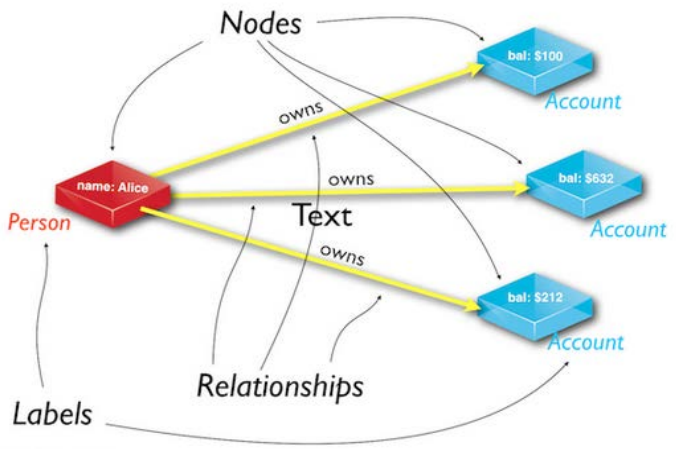


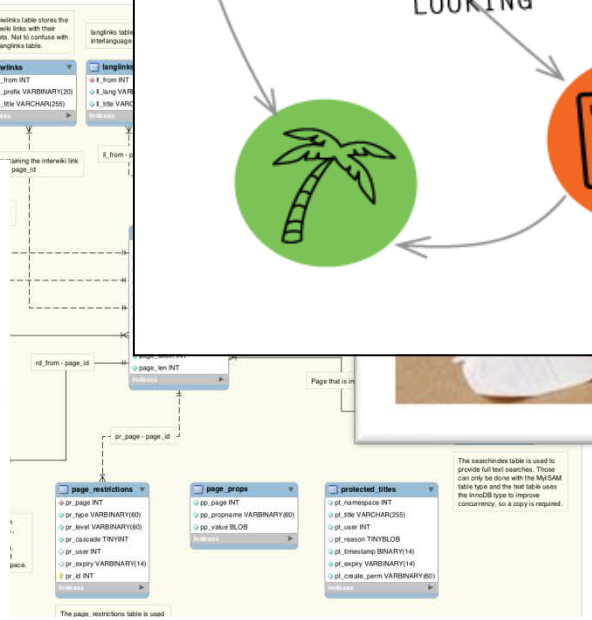








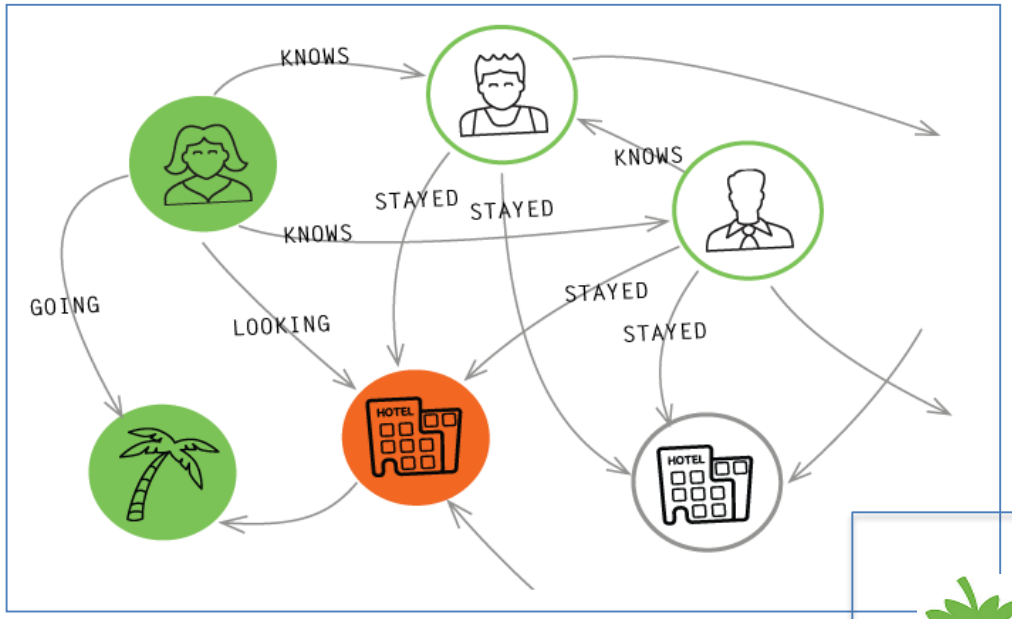


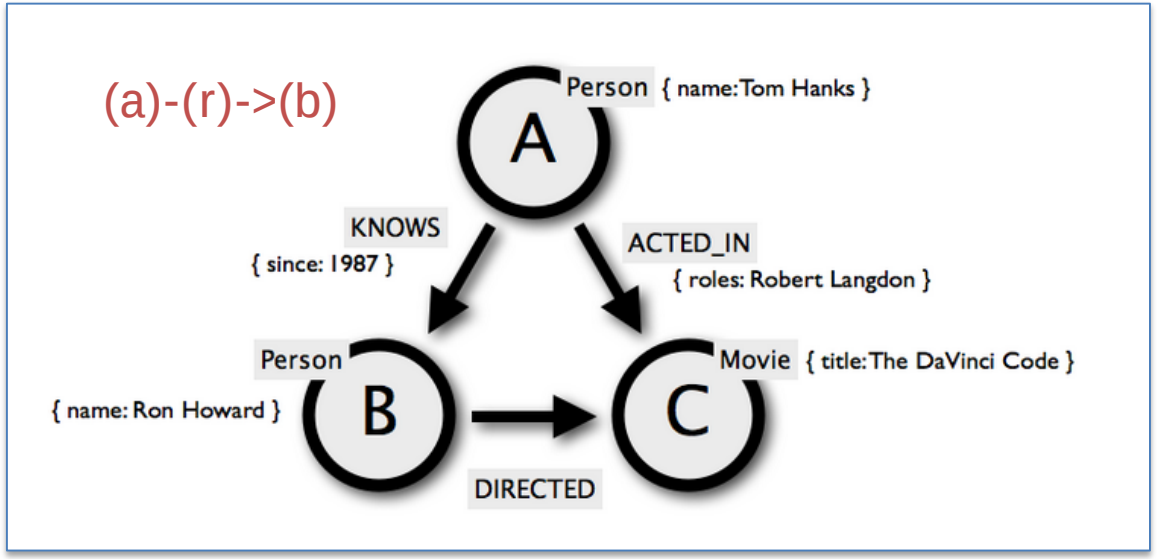


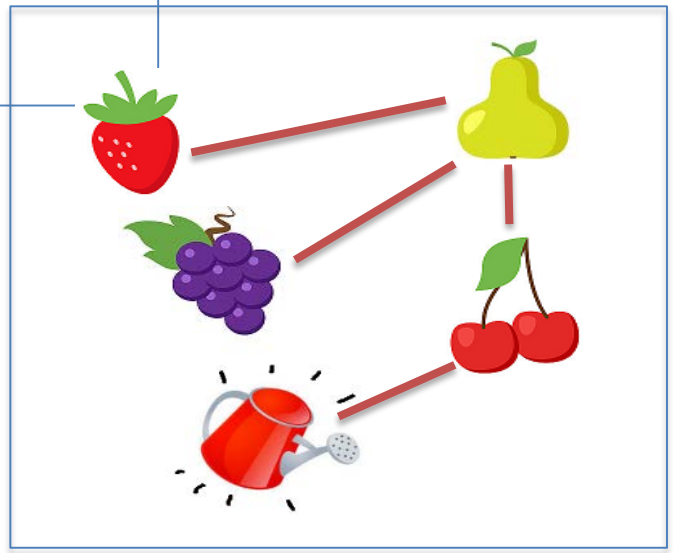
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Graph Database

- Flexible data structure, can change in the future, can create new nodes, new relationships
- Close match to business logic







- Can be queried by Cypher language

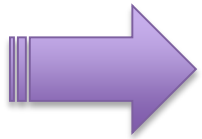
MATCH (n:Person)-(r:ACTED_IN)-(m)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Graphical User Interface

Technology built on AngularJS framework and microservices, and D3.js libraries.

JAX-RS (RESTful web services)
JSON responses



OSGi bundle as a web service

Browser client, Jscript

Model / View / Controller

UI Graphs

D3 Documents

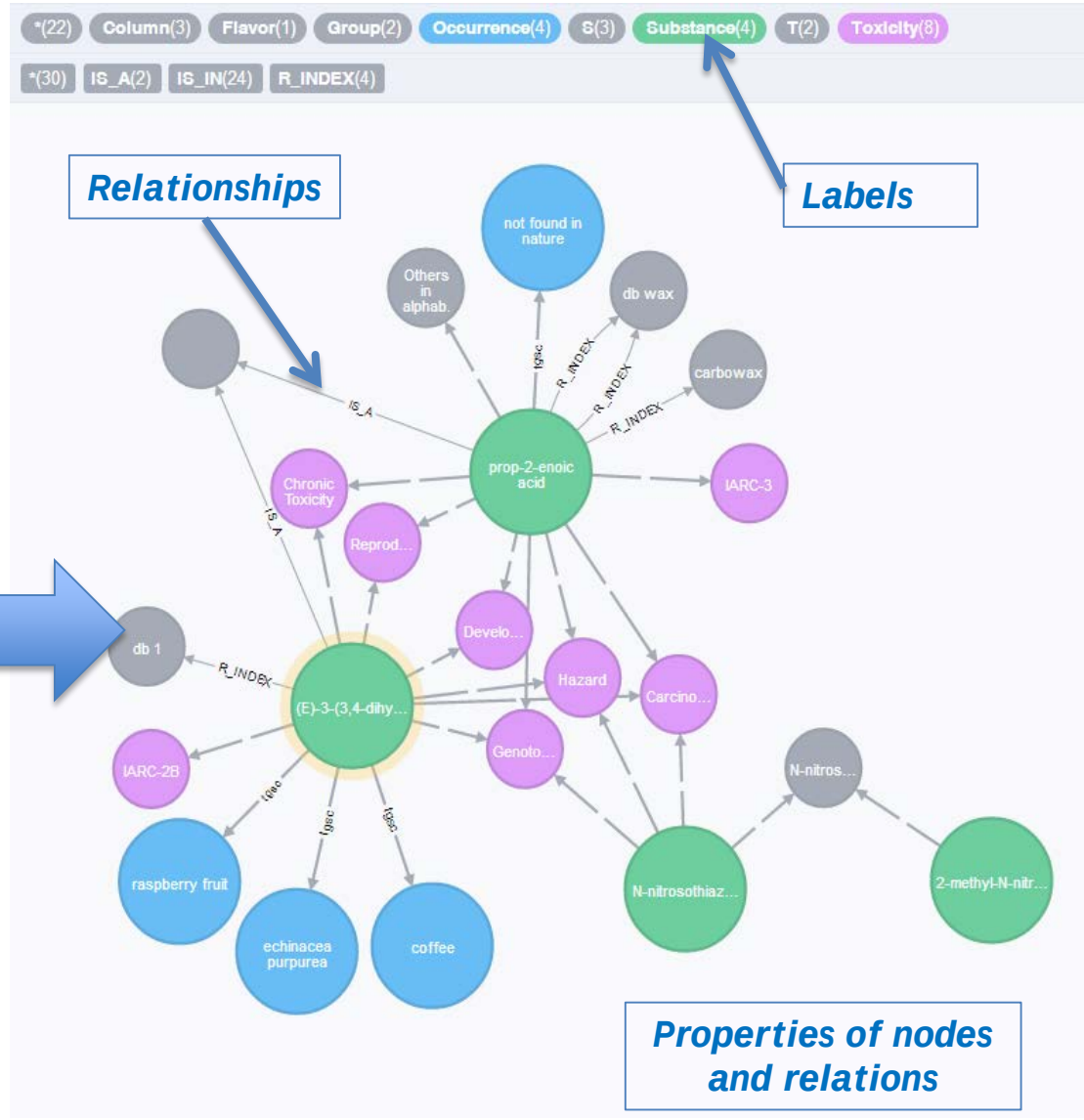
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Exploring the KnowledgeBase

- Example 1. Exploring EFSA

G7 {"name":"IARC-3","description":"Not classifiable as to its carcinogenicity to humans"}

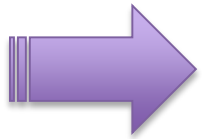
	A	B	C	D	E	F	G	H	I
1	n_id	n_labels	n_prop	r_prop	r_id	r_type	m_prop	m_id	m_labels
2	176	["Substance","S"]	116","uscd_date":1453218731713}	{}	293	IS_IN	{"name":"N-nitrosamines"}	161	["Group"]
3	178	["Substance","T"]	052","uscd_date":1453287539439}	{}	610726	IS_IN	{"name":"Genotoxicity"}	98	["Toxicity"]
4	178	["Substance","T"]	052","uscd_date":1453287539439}	{}	605784	IS_IN	{"name":"Carcinogenicity"}	97	["Toxicity"]
5	178	["Substance","T"]	052","uscd_date":1453287539439}	{}	588662	IS_IN	{"name":"Hazard"}	95	["Toxicity"]
6	178	["Substance","T"]	052","uscd_date":1453287539439}	{}	295	IS_IN	{"name":"N-nitrosamines"}	161	["Group"]
7	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	626305	IS_IN	{"name":"IARC-3","description":"No	167133	["Toxicity"]
8	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	619505	IS_IN	{"name":"Reproductive Toxicity"}	100	["Toxicity"]
9	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	614354	IS_IN	{"name":"Developmental Toxicity"	99	["Toxicity"]
10	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	610788	IS_IN	{"name":"Genotoxicity"}	98	["Toxicity"]
11	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	605850	IS_IN	{"name":"Carcinogenicity"}	97	["Toxicity"]
12	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	601958	IS_IN	{"name":"Chronic Toxicity"}	96	["Toxicity"]
13	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	588730	IS_IN	{"name":"Hazard"}	95	["Toxici
14	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	ameter":"0.25 mm"}	432224	R_INDEX	{"code":"carbowax20m","synonyms	52069	["Colun
15	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	ameter":"0.25 mm"}	273578	R_INDEX	{"code":"dbwax","synonyms":[],"na	28503	["Colun
16	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	ameter":"0.53 mm"}	273577	R_INDEX	{"code":"dbwax","synonyms":[],"na	28503	["Colun
17	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	h":"","origin":"tgsc"}	6448	IS_A	{"name":""}	1372	["Flavor"]
18	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{"origin":"tgsc"}	6447	IS_IN	{"name":"not found in nature"}	528	["Occurrence"]
19	255	["Substance","S"]	lic acid, [waste]","acrylic acid; pr	{}	437	IS_IN	{"name":"Others in alphab. order"	251	["Group"]
20	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{}	626528	IS_IN	{"name":"IARC-2B","description":"P	121	["Toxicity"]
21	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{}	619511	IS_IN	{"name":"Reproductive Toxicity"}	100	["Toxicity"]
22	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{}	614362	IS_IN	{"name":"Developmental Toxicity"	99	["Toxicity"]
23	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{}	610795	IS_IN	{"name":"Genotoxicity"}	98	["Toxicity"]
24	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{}	605856	IS_IN	{"name":"Carcinogenicity"}	97	["Toxicity"]
25	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{}	601964	IS_IN	{"name":"Chronic Toxicity"}	96	["Toxicity"]
26	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{}	588738	IS_IN	{"name":"Hazard"}	95	["Toxicity"]
27	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	ameter":"0.25 mm"}	214443	R_INDEX	{"code":"db1","synonyms":[],"name	28515	["Column"]
28	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	h":"","origin":"tgsc"}	35404	IS_A	{"name":""}	1372	["Flavor"]
29	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{"origin":"tgsc"}	35403	IS_IN	{"name":"raspberry fruit"}	1522	["Occurrence"]
30	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{"origin":"tgsc"}	35402	IS_IN	{"name":"echinacea purpurea"}	5047	["Occurrence"]
31	263	["Substance","S","T"]	rop-2-enoic acid","2-propenoic a	{"origin":"tgsc"}	35401	IS_IN	{"name":"coffee"}	1757	["Occurrence"]



Graphical User Interface

Technology built on AngularJS framework and microservices, and D3.js libraries.

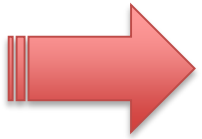
JAX-RS (RESTful web services)
JSON responses



OSGi bundle as a web service

Browser client, Jscript

Model / View / Controller



UI Graphs

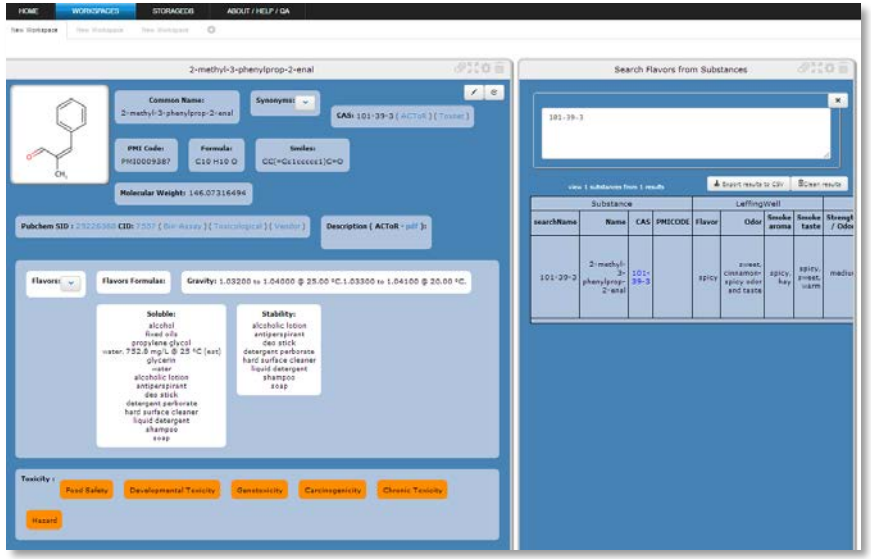
D3 Documents

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Demonstration – GDB Exploration

- Example 1. Exploring EFSA

- Example 2. Searching flavorants by specific attributes



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Graphical User Interface

Technology built on AngularJS framework and microservices, and D3.js libraries.

JAX-RS (RESTful web services)
JSON responses

neo4j

OSGi bundle as a web service

Browser client, Jscript

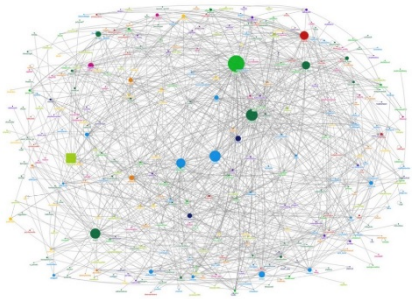
Model / View / Controller

ANGULARJS by Google

UI Graphs

D3

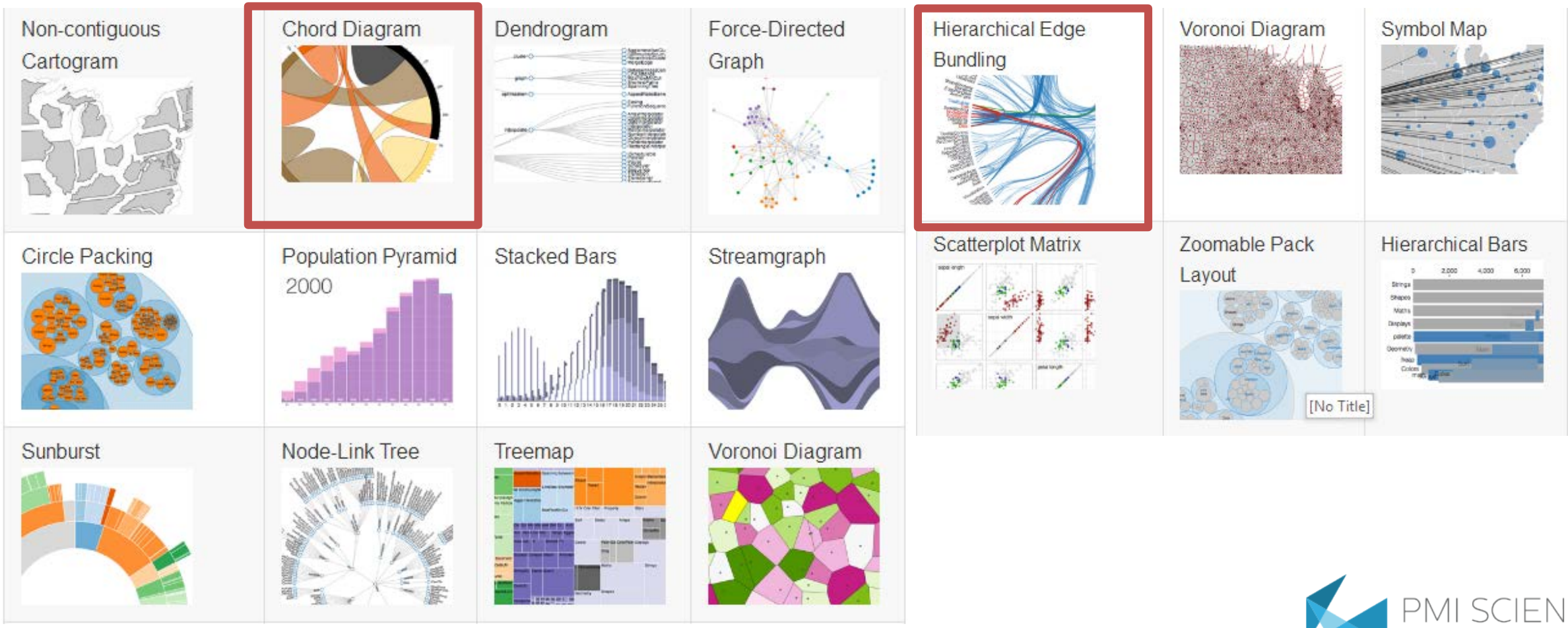
D3 Documents



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

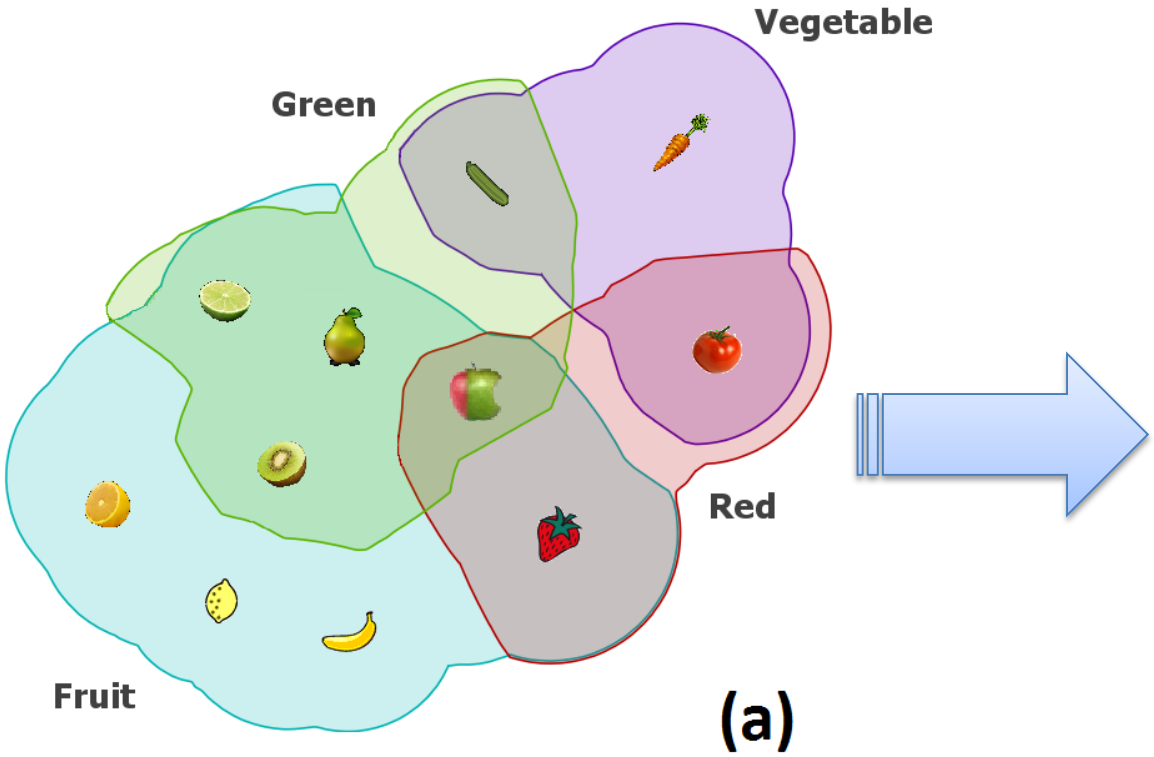
Dynamic Data Visualization Using D3

- D3.js (Data Driven Documents) is highly dynamic graphical visualization library



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Radial Set

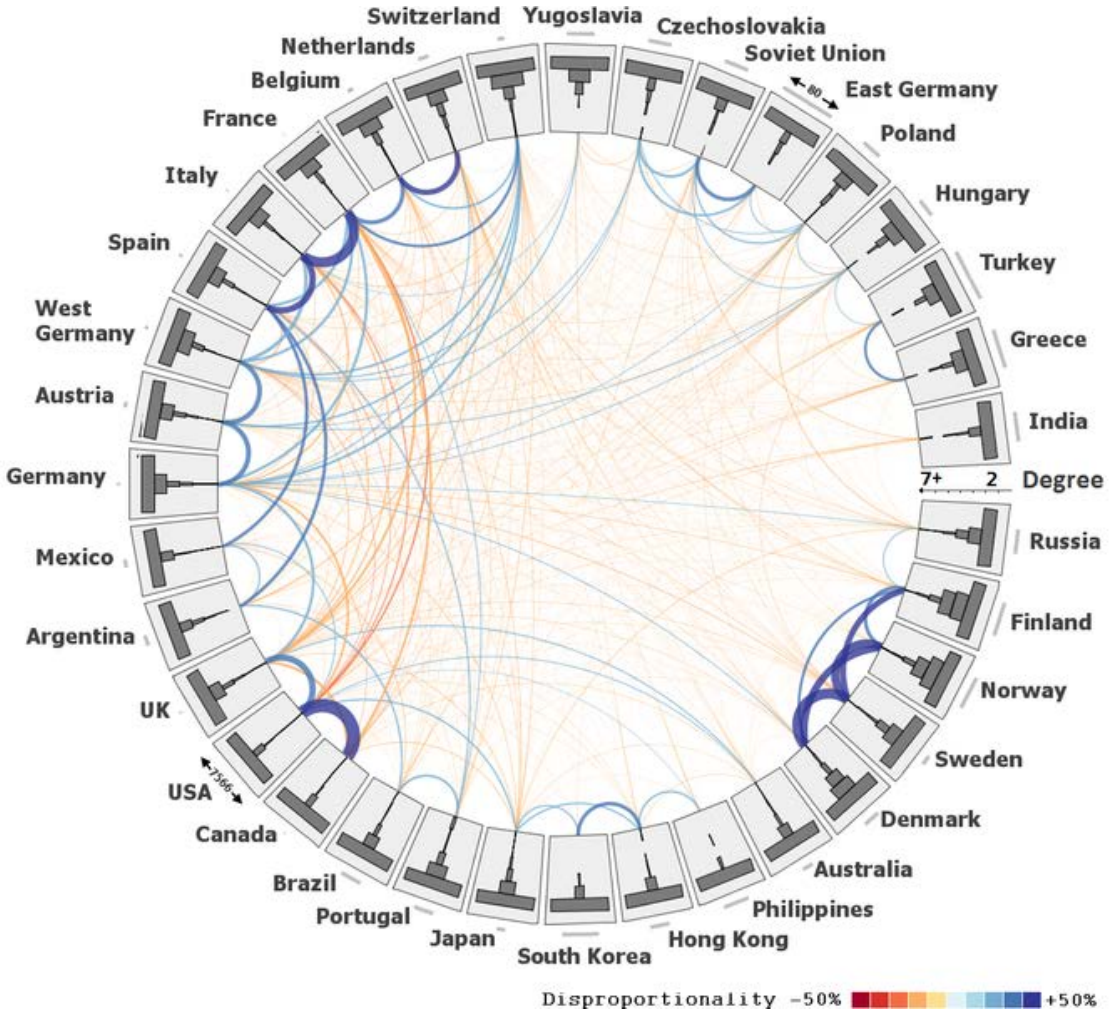


Alsallakh et al., "Radial Sets: Interactive Visual Analysis of Large Overlapping Sets", IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis), 19(12):2496-2505, 2013.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Exploring Radial Sets

- Example 3. Graphical visualization



- Radial sets reveal, and enable the analysis of, a multitude of overlapping patterns between large sets
- Frequency-based representation
- Overlaps, over- or under-represented elements
- The grouping of sets based on attributes (not only predefined sets)
- Potential to 'relate unrelated' – e.g. clusters -> toxicity values -> regulatory registry…

➔ **Graphical Read Across**

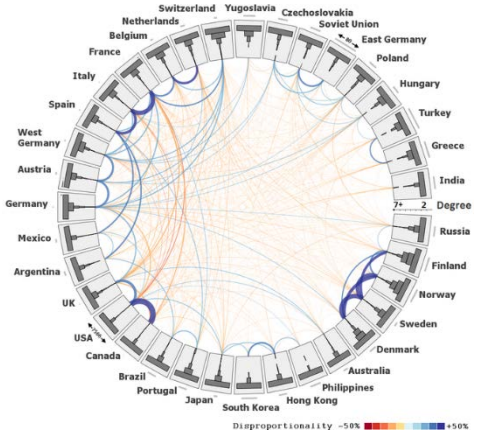
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Demonstration

- Example 1. Exploring EFSA

- Example 2. Searching flavorants by specific attributes

- Example 3: Visualizing set-attributes relationships using Radial Set
  (Graphical Read Across)



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Conclusion

- At PMI, we analyze complex chemical matrices related to smoke and aerosols
- Chemoinformatics team accelerates compounds identification and categorization

- Completely novel concept of databanking is based on:
  - Chemocentricity – the substance is the central element
  - Neo4j Graph database technology – flexible, evolving architecture
  - Microservices keeping database up-to-date by public data repositories
  - Customizable browser (Angular) and state-of-the-art D3 visualization

- Shown cases: Digitized flavorant sets - TGSC and EFSA, Radial Sets

- Proper KnowledgeBase that allows Storing – Exploring – Discovering

  Store large sets – Explore from any angle – Discover the unexpected

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Thank you to Antonio Castellon and Elyette Martin

and the Complex Matrix Analysis team of Mark Bentley.

Building an easy-to-explore graph database for smoke and/or aerosol constituents, enriched with publicly available toxicological and flavorant properties data

June 14th, 2016

Pavel Pospisil
Philip Morris International R&D