

G-SRS seed data and demo

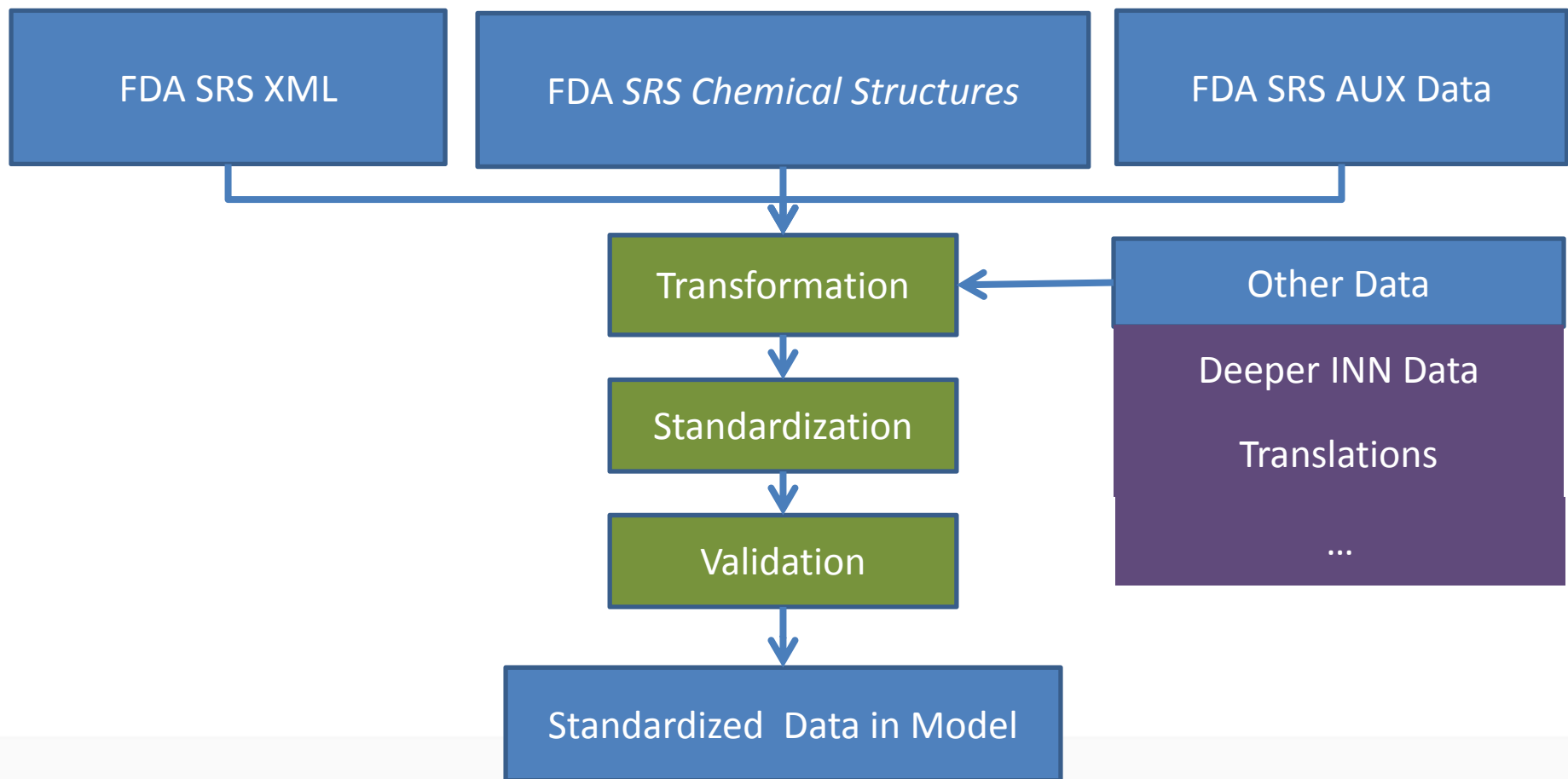
TYLER PERYEA
NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES
NATIONAL INSTITUTES OF HEALTH
7 SEPTEMBER 2015

NCATS

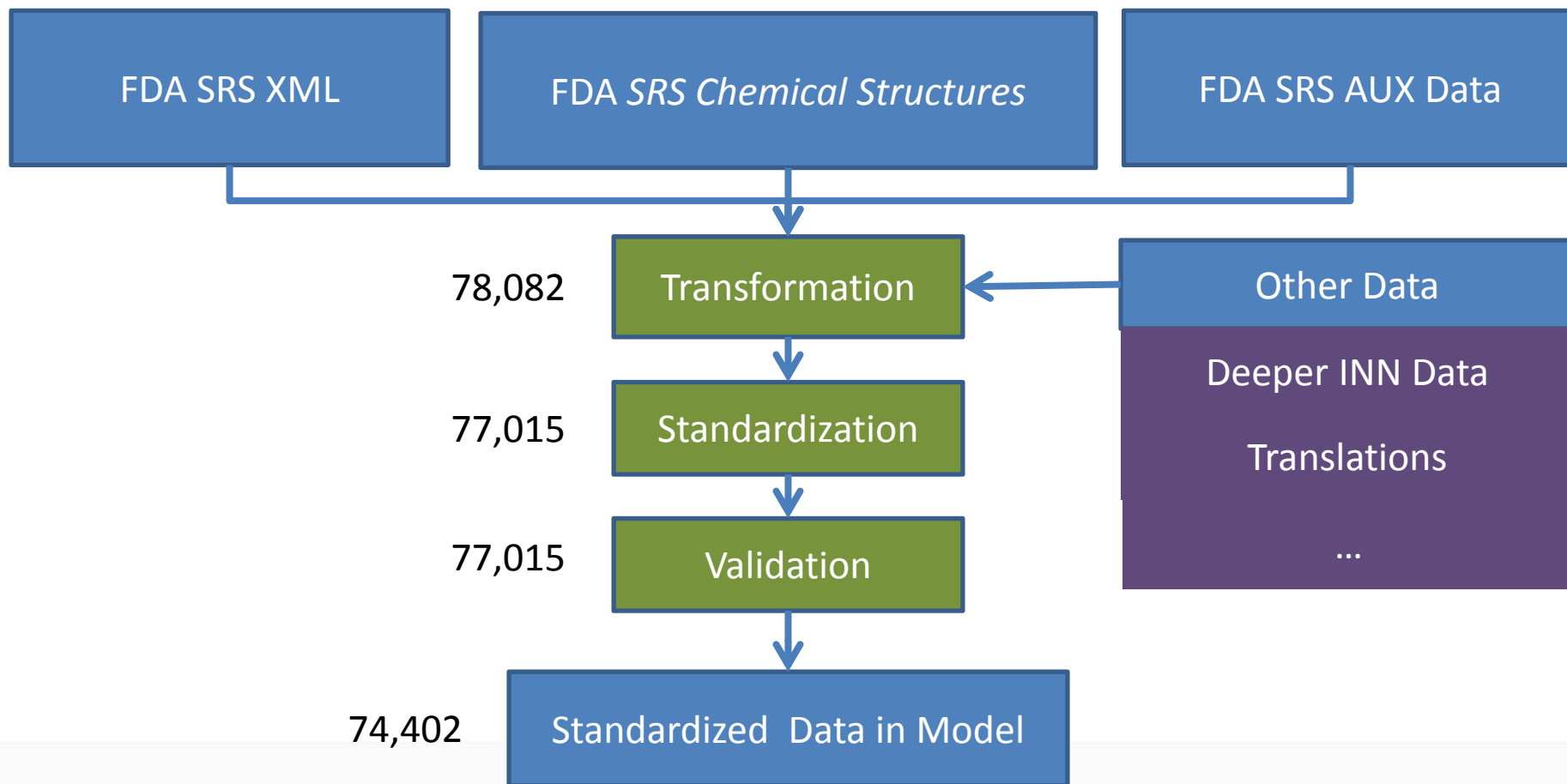
Outline

- Seed data status
- Migration effort
- System Demo
 - Old System
 - New System

Substance Seed Data



Transformation



Breakdown Failures

Problem	Count	Explanation
STEREOCHEMISTRY PROBLEMS	493	Usually marked as achiral, but has stereocenters
WAITING ON REFERENCED SUBSTANCE	445	Details for definition based on non-public or non-validated substance
PARSE ERROR	402	XML not valid
UNKNOWN XML FIELD	397	Legacy XML
INVALID MODIFICATION	206	Usually a non-standard modification agent

Breakdown Success

Substance Class		
<input type="checkbox"/>	Chemical	49423
<input type="checkbox"/>	Structurally Diverse	16712
<input type="checkbox"/>	Concept	4722
<input type="checkbox"/>	Mixture	1568
<input type="checkbox"/>	Polymer	1130
<input type="checkbox"/>	Protein	844



National Center
for Advancing
Translational Sciences

NCATS

Breakdown Success

Substance Class		
<input type="checkbox"/>	Chemical	49423
<input type="checkbox"/>	Structurally Diverse	16712
<input type="checkbox"/>	Concept	4722
<input type="checkbox"/>	Mixture	1568
<input type="checkbox"/>	Polymer	1130
<input type="checkbox"/>	Protein	844

Meaning of concept

Concepts are not necessarily substances but are still important for defining or describing substances.

3 meanings of Concepts

- Specified Substances waiting for implementation
- Slightly more specific terminology / shade of meaning that won't be specified substances
- Non-specific or generalized material "concept"

Seed data

- **Codes**
 - 157,895 codes
 - RN numbers
 - EVMPD codes
- **Names**
 - 485,415 unique names
 - Official
 - Systematic
 - Common names
- **Relationships**
 - 73,150 relationships between substances
 - Metabolites
 - Impurities
 - Active Moiety designations
 - Target / Molecular Interaction data

Concepts

The screenshot displays the FDA Substance Registration System search results for the term 'SALT'. The main page shows the search results for 'SALT', including the preferred substance name 'SODIUM CHLORIDE', the UNII '451W47IQ8X', the formula 'Cl.Na', and the InChIKey 'FAPWRFPIFSIZLT-UH'. A browser window titled 'Synonyms and Mappings - Mozilla Firefox' is overlaid on the right side of the page, showing a list of synonyms and mappings for 'SALT'.

U.S. Food and Drug Administration
Protecting and Promoting Your Health

Substance Registration System - U.S. Food and Drug Administration

Home ▶ Search Results

Search
Substance Registration System

SALT

Go back to previous page.

Search Results

Preferred Substance Name: SODIUM CHLORIDE
UNII: 451W47IQ8X
Formula: Cl.Na
Search Term: SALT
InChIKey: FAPWRFPIFSIZLT-UH

Resources

- ChemIDplus
- DrugPortal
- NCI Thesaurus

Go back to previous page.

About, Contact Us, Copyright, Privacy, Accessibility
U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894
National Institutes of Health, Health & Human Services
Freedom of Information Act
Last updated: Oct 2015

Synonyms and Mappings

- 0.9% NACL
- 0.9% SALINE
- 0.9% SALINE SOLUTION
- 231-598-3
- 7647-14-5
- AQUA MARINA
- AQUA MARINA [HPUS]
- B1655 [LANGUAL]
- COMPOUND SOLUTION OF SODIUM CHLORIDE
- COMPOUND SOLUTION OF SODIUM CHLORIDE [INN]
- FAPWRFPIFSIZLT-UHFFFAOYSA-M
- HALFLYTELY COMPONENT SODIUM CHLORIDE
- HYPERTONIC SALINE
- HYPERTONIC SODIUM CHLORIDE SOLUTION
- ISOTONIC SALINE
- ISOTONIC SODIUM CHLORIDE SOLUTION
- ISOTONIC SODIUM CHLORIDE SOLUTION [U]
- NACL
- NACL 0.9%
- NATRUM MURIATICUM
- NATRUM MURIATICUM [HPUS]
- NORMAL SALINE
- NORMOCARB COMPONENT SODIUM CHLORIDE
- NS
- SALINE
- SALINE SOLUTION
- SALINE SOLUTION, HYPERTONIC

Concepts

ginas alpha version

Search ...

Home Browse Substances Register Substance Download Structure Search Sequence Search

Structure Moieties **2** Names **13** Codes **15** Relationships **10** References **32** **SODIUM CHLORIDE**

Chemical
RECORD STATUS: APPROVED
VERSION: **1**

451W47IQ8X

Structure

Molecular Formula	Molecular Weight	Defined Stereocenters	EZ Centers	Charge	Reference
ClH.Na	58.443	0/0	No	0	2

+Na -Cl

Moieties 2

Concepts



alpha version

Search ...



Home

Browse Substances

Register Substance

Download

Structure Search

Sequence Search

Login

Structure

Moieties 2

Names 13

Codes 18

Relationships 10

References 32

SODIUM CHLORIDE

Variant Concepts 7



SODIUM CHLORIDE INJECTION,
BACTERIOSTATIC



SEA WATER



ISOTONIC SODIUM CHLORIDE
SOLUTION



SALT



National Center
for Advancing
Translational Sciences

NCATS

Seed Data

- Data integrity
 - During this process, we identify known systematic errors and concerns

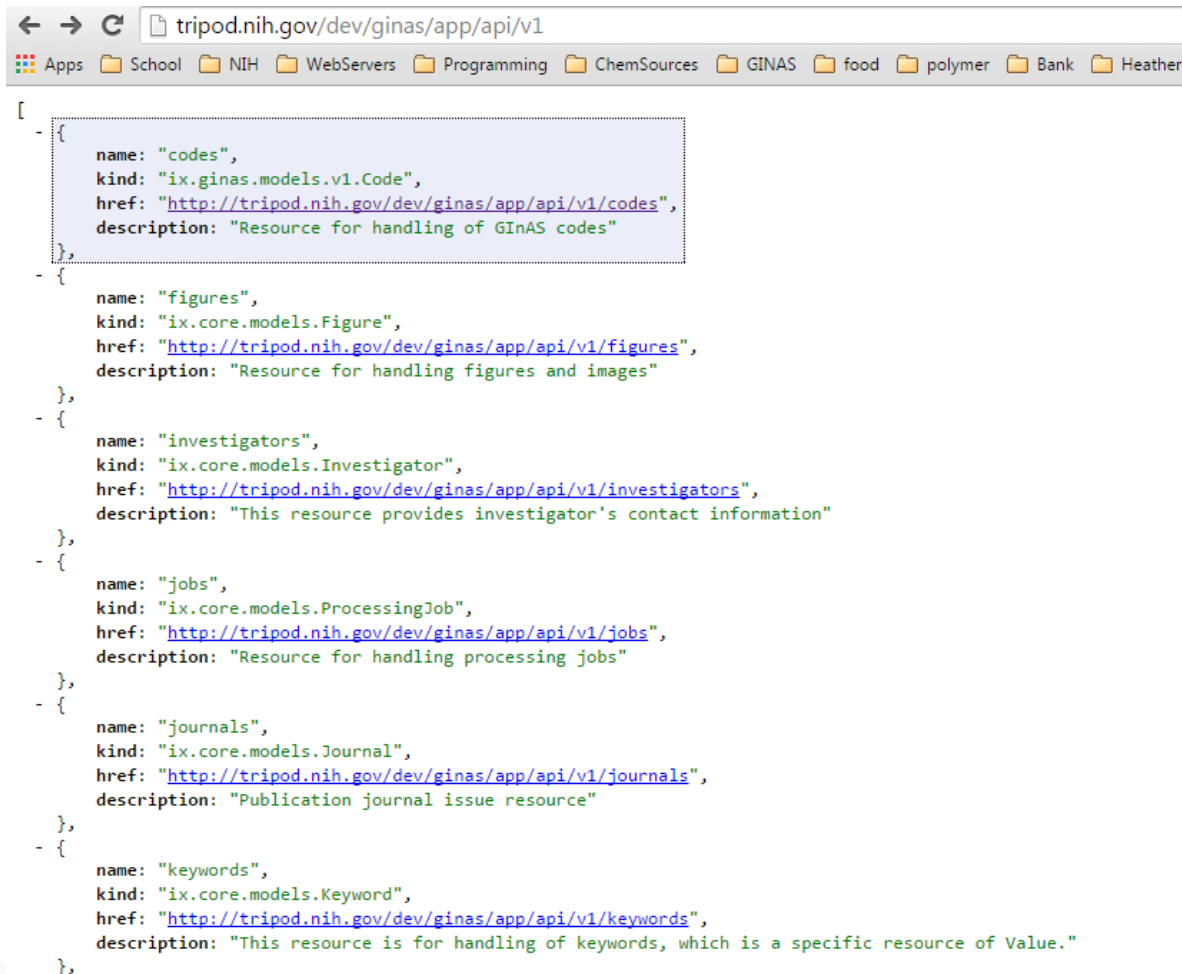
UNII	PT	Structure	Name	Name Structure
MI97BW74XZ	CAPRYLYL GLYCERYL ETHER	CCCCCCCCOCC(O)CO	CAPRYLYL GLYCERYL ETHER	C(C(O)CO)OC(CCCCCC)=O
CZ227117JE	CYCLOMETHICONE 4	C[Si]1(C)[O][Si](C)(C)[O][Si](C)	2,2,4,4,6,6,8,8,10,10-DECAMETHYL-1,3,	C[Si]1([O][Si]([O][Si]([O][Si]([O][Si]([O]
3T4AJR1H24	CHLORODIBROMOMETHANE	ClC(Br)Br	CHLOROBROMOFORM	ClC(Br)(Br)Br
J086PM3RRT	MYRISTAMINE OXIDE	CCCCCCCCCCCC[N+](C)(C)	DIMETHYLTETRADECYLAMINE	CN(CCCCCCCCCCCC)C
1KEE9757OT	DIPHENYL CHLOROPHOSPHONATE	ClP(=O)(OC1=CC=CC=C1)OC2=CC=CC=C2	DIPHENYLPHOSPHORYL CHLORIDE	C1(=CC=CC=C1)P(=O)(C1=CC=CC=C1)
BI73W105SX	HYDROFURAMIDE	O1C=CC=C1/C=N\C(\N=C\C2=FURFURAMIDE	FURFURAMIDE	C(C1=CC=CO1)=(=O)N
AA2L7VX59K	MAGNESIUM PHOSPHATE, MONOBASIC, ANHYDRIC	[Mg++].OP(O)(([O-])=O).OP(O	MONOMAGNESIUM ORTHOPHOSPHAT	P(=O)([O-])([O-])O.[Mg+2]
701EKV9RMN	CALCIUM PHOSPHATE, MONOBASIC, ANH			
3I8OLS586D	ETHYL 2-MERCAPTOPROPIONATE			
170BF0DD31	CARBOBENZOXOXY CHLORIDE			
X3TT5X989E	AQUA REGIA			
P700UL0YIF	DISODIUM 2-HYDROXY-2-SULFONATOACET			
DN57U1925N	AZAN			
61LJ05I15S	GLUTAMIC ACID, DL-			
WW9CM0067Z	PANTHENOL			
025P240TM5	2-SEC-BUTYLPHENOL			
8Z96QXD6UM	TRIETHYL CITRATE			
DO250MG0W6	DIHYDROXYALUMINUM AMINOACETATE			
E7WED276IS	MERCAPTOPYRINE			
25IH6R4SGF	EDETATE CALCIUM DISODIUM			
8TIF7T48FP	PIPERAZINE PHOSPHATE			
090XB01F30	POTASSIUM ALUM ANHYDROUS			
BZ2P21Y2W0	MDE HYDROCHLORIDE, (+)-			
KVF4175054	SODIUM PALMITOYL-L-Glutamate			

Seed Data

- Going forward:
 - Convert concepts waiting for G1SS status to full group 1 specifications.
 - Include Nucleic Acid in seed data
 - Continually curate existing data
 - Force import those that fail validation with warnings
 - Even incomplete / non-conformant substances are worth importing

Demo

API



```
[
- {
  name: "codes",
  kind: "ix.ginas.models.v1.Code",
  href: "http://tripod.nih.gov/dev/ginas/app/api/v1/codes",
  description: "Resource for handling of GINAS codes"
},
- {
  name: "figures",
  kind: "ix.core.models.Figure",
  href: "http://tripod.nih.gov/dev/ginas/app/api/v1/figures",
  description: "Resource for handling figures and images"
},
- {
  name: "investigators",
  kind: "ix.core.models.Investigator",
  href: "http://tripod.nih.gov/dev/ginas/app/api/v1/investigators",
  description: "This resource provides investigator's contact information"
},
- {
  name: "jobs",
  kind: "ix.core.models.ProcessingJob",
  href: "http://tripod.nih.gov/dev/ginas/app/api/v1/jobs",
  description: "Resource for handling processing jobs"
},
- {
  name: "journals",
  kind: "ix.core.models.Journal",
  href: "http://tripod.nih.gov/dev/ginas/app/api/v1/journals",
  description: "Publication journal issue resource"
},
- {
  name: "keywords",
  kind: "ix.core.models.Keyword",
  href: "http://tripod.nih.gov/dev/ginas/app/api/v1/keywords",
  description: "This resource is for handling of keywords, which is a specific resource of Value."
},
]
```

More information

- Project Webpage
 - <http://tripod.nih.gov/ginas>
- New Development Site
 - <http://tripod.nih.gov/ginas/app>
- Old Development Site
 - <http://ginas.hres.ca/ginas/>



Acknowledgements

U.S. Food and Drug Administration

Yulia Borodina	Archana Newatia
Larry Callahan	Vada Perkins
Ta-Jen Chen	Mary-Ann Slack
Ramez Ghazzaoui	Frank Switzer
Elaine Johansen	Alex Welsch

U.S. Pharmacopeial Convention

Fouad Atouf	Andrej Wilk
-------------	-------------

Tina Morris

Federal Institute for Drugs and Medical Devices (Germany)

Thomas Balzer

SwissMedic

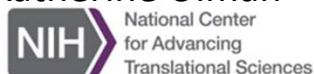
Philipp Weyermann

European Directorate Quality of Medicines

Christopher Jarvis

Dow Corning

Katherine Ulman



NIH/NCATS

Chris Austin	Tyler Peryea
Ajit Jadhav	Tim Sheils
Dac-Trung Nguyen	Tongan Zhao

Medicines Evaluation Board (Netherlands)

Herman Diederik	Burt Kroes
Marcel Hoefnagel	Ciska Matai

Joris Kampmeijer

Health Canada

Vikesh Srivastava

Royal Botanic Gardens, Kew (UK)

Bob Allkin	Elizabeth Dauncey
------------	-------------------

European Medicines Agency

Paolo Alcini	Telonis Pangiotis
Sabine Brosch	Ilaria Del Seppia

Uppsala Monitoring Centre / WHO

Malin Jakobsson	Malin Fladvad
-----------------	---------------

Martin Strömberg