G-SRS seed data and demo
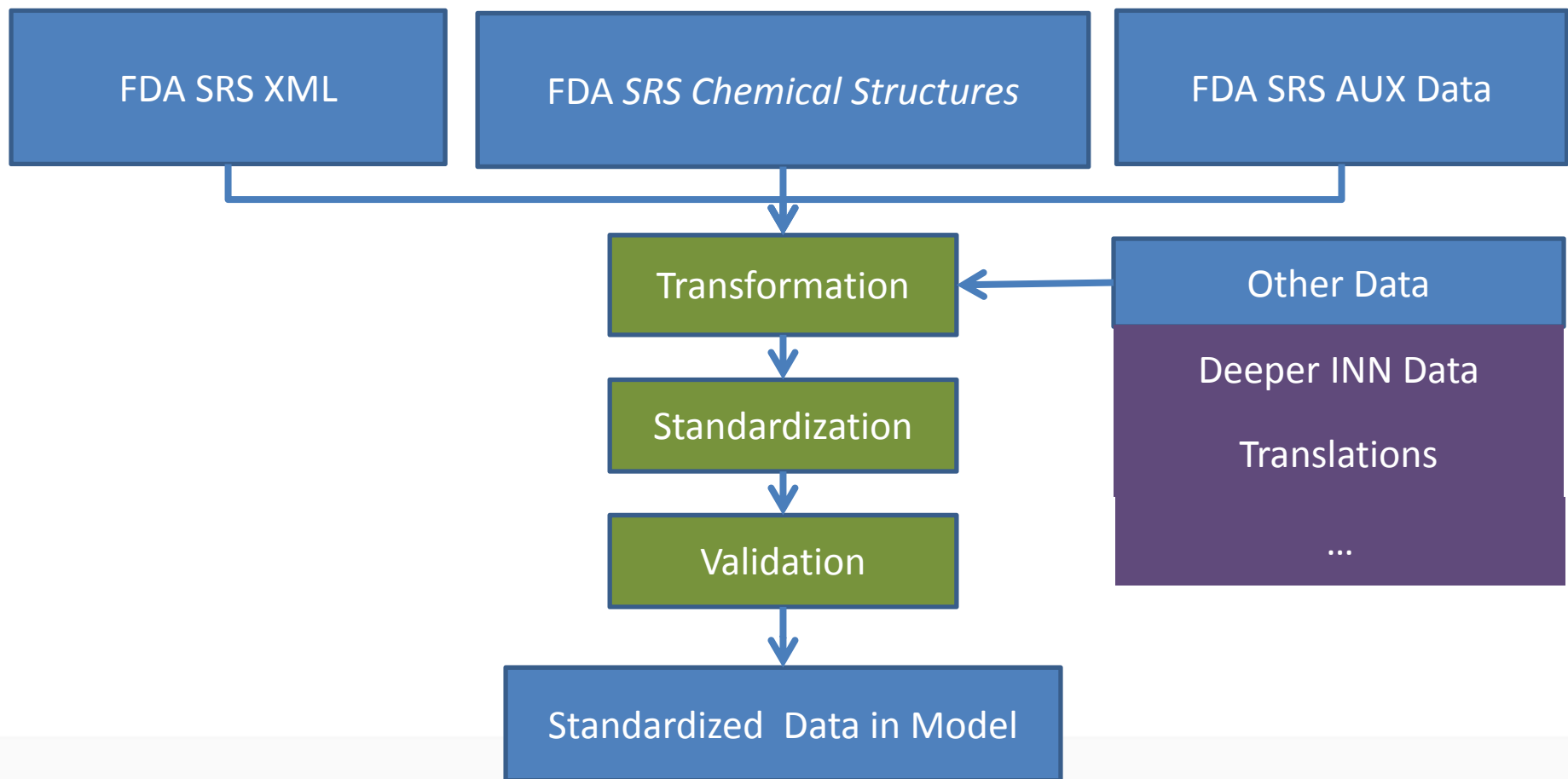
TYLER PERYEA
NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES
NATIONAL INSTITUTES OF HEALTH
7 SEPTEMBER 2015

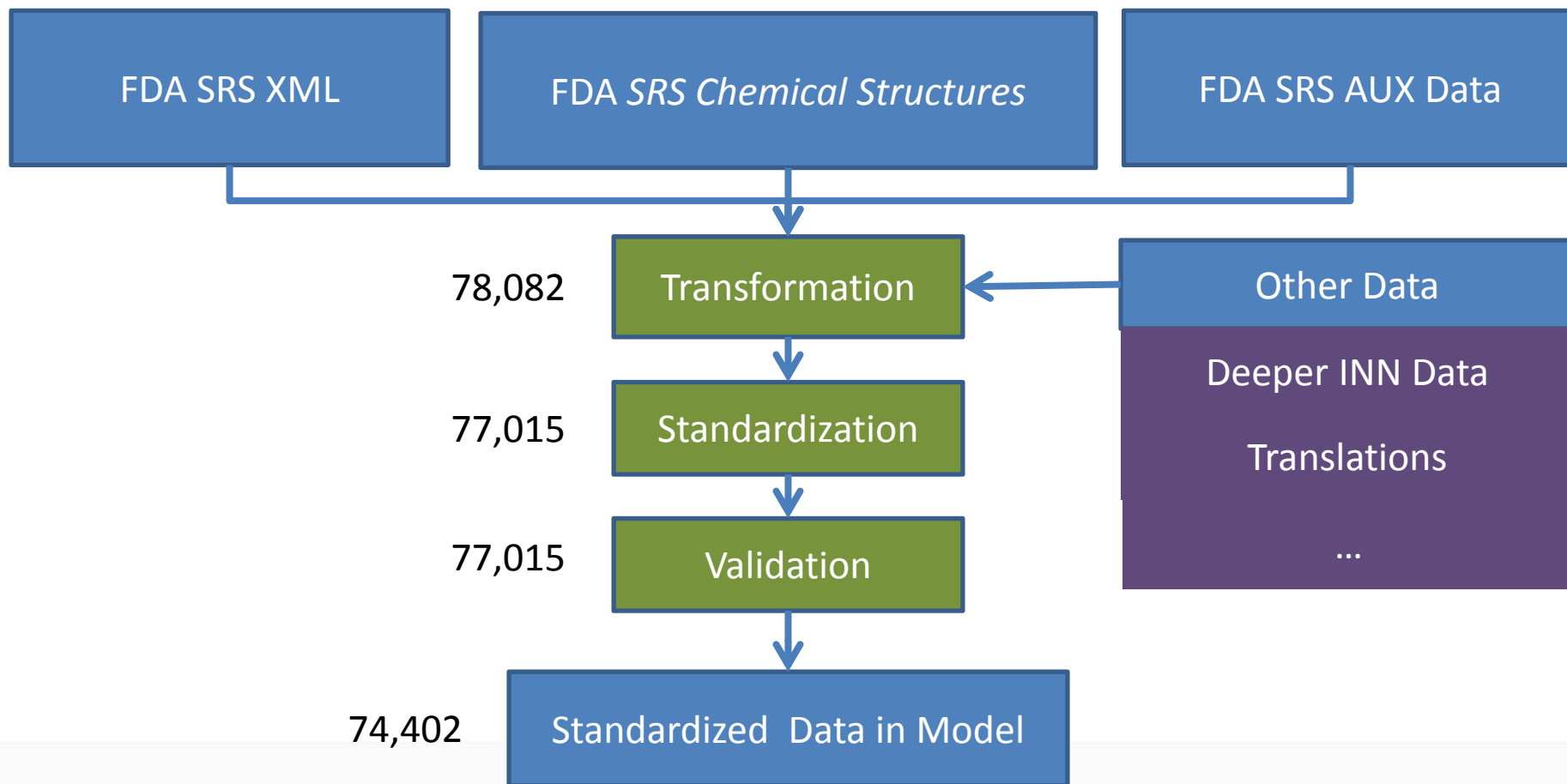NCATS

National Center
for Advancing
Translational Sciences

Outline

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

Substance Seed Data

Transformation

FDA SRS XML	FDA *SRS Chemical Structures*	FDA SRS AUX Data

78,082 → **Transformation** ← Other Data

77,015 → **Standardization**

Deeper INN Data

Translations

...

77,015 → **Validation**

74,402 → Standardized Data in Model

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

Breakdown Success

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

Concepts

Concepts

Seed Data

- Data integrity
  - During this process, we identify known systematic errors and concerns
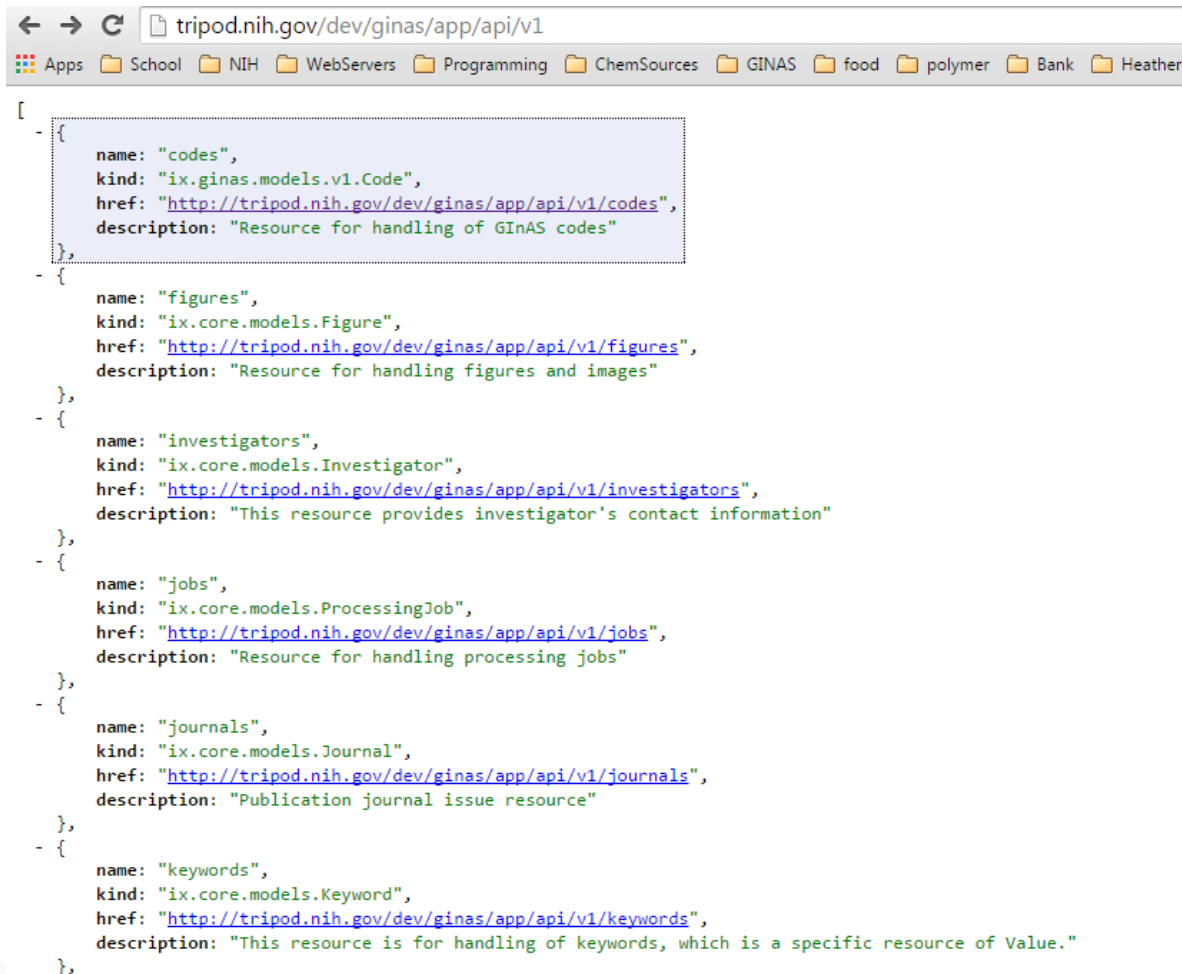
UNII	PT	Structure	Name	Name Structure
MI97BW74XZ	CAPRYLYL GLYCERYL ETHER	CCCCCCCCCOCC(O)CO	CAPRYLYL GLYCERYL ETHER	C(C(O)CO)OC(CCCCCCC)=O
CZ227117JE	CYCLOMETHICONE 4	C[Si]1(C)O[Si](C)(C)O[Si](C)(	2,2,4,4,6,6,8,8,10,10-DECAMETHYL-1,3,	C[Si]1(O[Si](O[Si](O[Si](O1)(C)(
3T4AJR1H24	CHLORODIBROMOMETHANE	ClC(Br)Br	CHLOROBROMOFORM	ClC(Br)(Br)Br
J086PM3RRT	MYRISTAMINE OXIDE	CCCCCCCCCCCCCC[N+](C)(C)	DIMETHYLTETRADECYLAMINE	CN(CCCCCCCCCCCCCC)C
1KEE97570T	DIPHENYL CHLOROPHOSPHONATE	ClP(=O)(OC1=CC=CC=C1)OC2	DIPHENYLPHOSPHORYL CHLORIDE	C1(=CC=CC=C1)P(=O)(C1=CC=CC=C1)C
BI73W105SX	HYDROFURAMIDE	O1C=CC=C1\C=N\C(\N=C\C2=	FURFURAMIDE	C(C1=CC=CO1)(=O)N
AA2L7VX59K	MAGNESIUM PHOSPHATE, MONOBASIC, ANHYDRC	[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			
170BP0DD31	CARBOBENZOXY CHLORIDE			
X3TT5X989E	AQUA REGIA			
P70OUL0YIF	DISODIUM 2-HYDROXY-2-SULFONATOACE			[O-].[Na+].[N
DN57U1925N	AZAN			
61LJ05I15S	GLUTAMIC ACID, DL-			=O)O
WV9CM0O67Z	PANTHENOL			)CO)=O)(=O
025P24OTM5	2-SEC-BUTYLPHENOL			=O
8Z96QXD6UM	TRIETHYL CITRATE			C(=O)O)=O
DO250MG0W6	DIHYDROXYALUMINUM AMINOACETATE			
E7WED276I5	MERCAPTOPURINE			
25IH6R4SGF	EDETATE CALCIUM DISODIUM			CC(=O)[O-])C
8TIF7T48FP	PIPERAZINE PHOSPHATE			
09OXB01F3O	POTASSIUM ALUM ANHYDROUS			.O.S(=O)(=O)
BZ2P21Y2W0	MDE HYDROCHLORIDE, (+)-	Cl.CCN[C@@H](C)CC1=CC2=(	(S)-N-ETHYL-3,4-METHYLENEDIOXYAM	C(C)N[C@@H](C)CC1=CC2=C(C=C1)OC
KVF4175O54	SODIUM PALMITOYL GLUTAMATE	[H+].[Na+].CCCCCCCCCCCC	SODIUM N-PALMITOYL-L-GLUTAMATE	C(CCCCCCCCCCCCCCCC)(=O)N[C@@H](

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

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

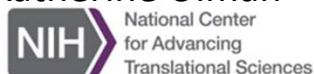Philipp Weyermann

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 Montintoring Centre / WHO

Malin Jakobsson	Malin Fladvad
Martin Strömberg	

National Center for Advancing Translational Sciences