



Informatics/Terminology

Summary of NCS@NIH Presentations

Michael G. Kahn MD, PhD
University of Colorado Department of Pediatrics

Presented to NCS Federal Advisory Committee
19 October 2011 1:15P-2:00P

Dana Dabelea, MD, PhD
Principal Investigator
University of Colorado Denver
Colorado School of Public Health

Colorado Study Center 

Presentations Summarized



- Michael Kahn (Colorado): Neonatal Research Networks Terminology (NRNT) Project
- Warren Kibbe (Northwestern): Open Source Software: Laying the Foundation
- William Hogan (Arkansas): Meeting the Informatics Challenges of the National **Children's Study**



The one line summaries



- Michael Kahn (Colorado): Neonatal Research Networks Terminology (NRNT) Project
 - A project that harmonized the data collection terms across three existing national neonatal research networks and linked the harmonized terms to SNOMED, a license-free terminology identified by the federal government as a national standard for EHRs



The one line summaries



- Warren Kibbe (Northwestern): Open Source Software: Laying the Foundation
- **Northwestern's effort to create a** comprehensive clinical research management system using an open-source multi-institutional consortium development model and the challenges of balancing flexibility versus complexity



The one line summaries



- William Hogan (Arkansas): Meeting the Informatics Challenges of the National Children's Study
 - The Arkansas experience with implementing and integrating **multiple** open-source software applications developed by other organizations to support NCS data collection needs.



In what way are these three talks similar?



- Michael Kahn (Colorado): Neonatal Research Networks Terminology (NRNT) Project
- Warren Kibbe (Northwestern): Open Source Software: Laying the Foundation
- William Hogan (Arkansas): Meeting the Informatics Challenges of the National **Children's Study**



Future-proofing the NCS findings



- Over the 20+ year lifespan of NCS, technology will change in totally unanticipated ways
 - Current technology lifecycle < 18 months and decreasing
 - **NCS data must be “liquid”**
 - Balancing innovation with stability is living on a knife’s edge
 - Clinical research informatics is a very young field



Future-proofing the NCS findings

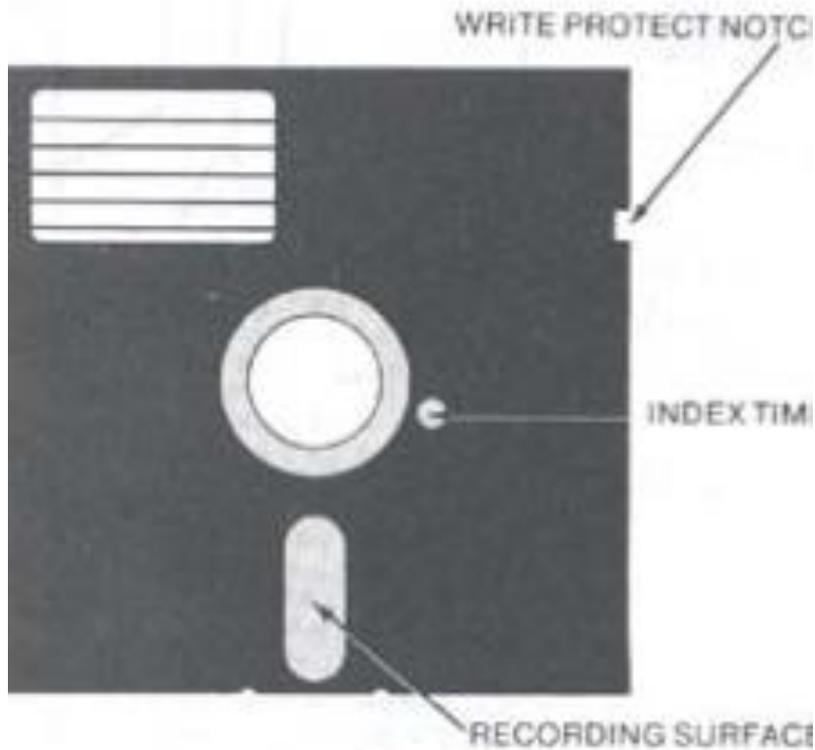


What technologies were you using in 1991?

- Microsoft releases MS DOS 5.0 / Windows 3.0
Apple releases System 7
- Internet is made available to unrestricted commercial use and number of computers on the net reaches 1 million
- The World Wide Web is launched to the public August 6, 1991. Linux is introduced by Linus Torvald in August 25 1991
- Intel introduces the 486x chip, selling for \$258.00



Do you have the ability to read your data from these storage devices?



What does “Data are Liquid” mean?



- Mandl KD, Kohane IS. No small change for the health information economy. NEJM 2009; 360: 1278-1281 (March 26, 2009)
“The platform and its applications should reduce impediments to the transfer of data, in an agreed-upon form, from one system to another.”
- Kish, L. Liquid data and the health information economy: Is 2011 finally the year?. Opensource.com Posted 23 May 2011.
“To get to a true health information economy, health info has to travel from its vast untapped repositories to **where it’s needed. Once it’s liberated, data will flow to** help patients and physicians make better choices and continue learning while technologists use that data to provide better solutions.”





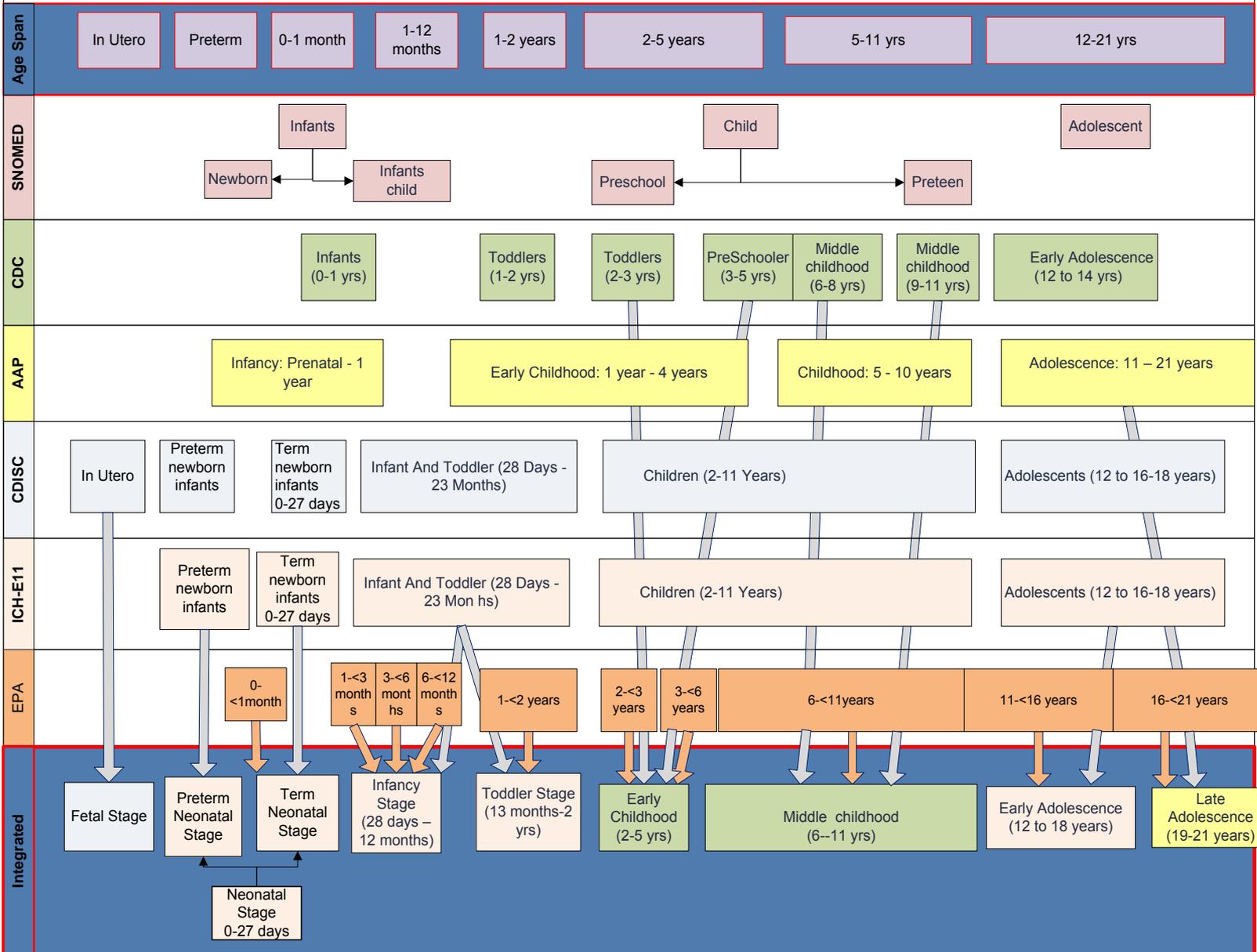
Neonatal Research Networks Terminology (NRNT) Project

Michael G. Kahn MD, PhD
University of Colorado Department of Pediatrics
24 August 2011

Dana Dabelea, MD, PhD
Principal Investigator
University of Colorado Denver
Colorado School of Public Health

Colorado Study Center 

Integrated child-life stages for NICHD Pediatric Terminology as mapped to existing medical terminologies



AAP: American Academy of Pediatrics

CDC: Centers for Disease Control and Prevention

CDISC: Clinical Data Interchange Standards Consortium

EPA: Environmental Protection Agency

ICH-E11: International Conference on Harmonisation

SNOMED: Systematized Nomenclature of Medicine

NRNT: Goals & Objectives



- Create a harmonized neonatal terminology that:
 - Builds upon accepted neonatal research data collection needs
 - Harmonizes across the stakeholders to provide a common data view
 - Aligns with an accepted international clinical care terminology standard
 - Does not require participating networks to change current data collection and analysis practices
- Develop a set of terminology harmonization procedures, methods, and tools that can be reused in other areas of pediatrics research



NRNT Participants



- Neonatal networks
 - Vermont-Oxford
 - **Children's Hospital Neonatal Consortium**
 - NICHD Neonatal Research Network
- Domain experts
 - **CCHMC, CHOP, Texas Children's, NICHD, UC-Denver, VON**
- Terminology technologies experts
 - Apelon, Booz Allen Hamilton
- International participation
 - GRIP Network (Adolf Valls i Soler)
 - Canadian Neonatal Network (Prakesh Shah)



Results Thus Far: Counts by Body System*



Body System	NRNT Concepts	Stakeholder Terms	SNOMED CT Mappings	NRNT-Specific Concepts
CNS	249	162	197	50
Cardiovascular	209	141	155	50
Endocrinology	80	40	76	4
Genetics	163	97	153	9
Gastrointestinal	158	119	135	23
Genitourinary	89	56	79	10
Hematology	125	66	117	8
Metabolic	112	67	110	2
Musculoskeletal	70	44	66	4
Oncology	33	13	31	2
Ophthalmology	83	41	58	25
Respiratory	114	94	84	29
Skin	105	38	97	8

* The same concept will be counted in multiple rows if it is included in multiple body systems



What has NRNT produced?



- Robust pediatric terminology that covers data collection requirement for neonatal research networks
 - Links to an international terminology standard
 - Retains grouping concepts for data reporting and aggregation
 - Supports cross-network comparison and aggregation
- An initial core pediatric terminology with known extensions to SNOMED CT
 - Extensions will be promoted to NLM and IHTSDO





Open Source Software Laying the foundation

Warren A. Kibbe, PhD
Northwestern University Health and Biomedical Informatics
24 August 2011

Jane Holl, MD
Principal Investigator
Feinberg School of Medicine
Northwestern University

Greater Chicago Study Center



Clinical Research Informatics Challenges



- Make protocols computable
- Make software only as complex (or as simple) as it needs to be
- Change only what needs to change, but change it everywhere.
DRY (Don't Repeat Yourself)
- Identify and enforce a single source of truth

Everything should be kept as simple as possible, but not simpler. Albert Einstein



About open source



- Open source is an approach to delivery software solutions based on community input, resources, and needs
- Done right, open source projects offer a lightweight framework for enabling flexible **collaboration and support 'microunits'** of contribution
- Open source does not mean lack of control. A control board typically oversees standards, training, review and hands out the ability to **'commit' (submit code and documents)** to the project

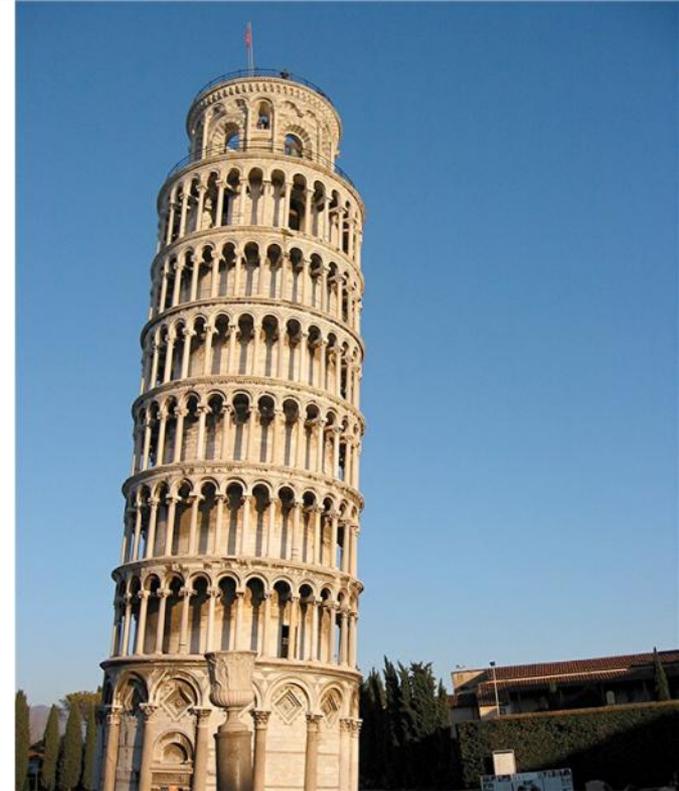


Supporting Research with Software

What is the problem?



- Protocols are exception driven and hard to make computable
- Flexibility in design usually makes complexity in operation
- Change is good, necessary, but hard to manage without computable representation



Providing a robust model for clinical research with the 'right' level of flexibility is difficult



NCS Navigator



- An open source set of tools for managing the NCS study
- Uses modern software development practices
- The NCS Navigator Consortium comprises seven study sites located at seven institutions. The institutions are:

Northwestern University, Emory University, Johns Hopkins University, Maine Medical Center, University of Colorado Denver, University of Minnesota, University of Pittsburgh



NCS Navigator and the MDES



- A key component of the NCS Navigator is the NCS Navigator MDES Warehouse. MDES is the Master Data Element Specification. It is a computable representation of the data definitions, rules and collection instruments available in the MDES.
- The MDES Warehouse versions and accessions submissions based on the then current version of the MDES and migrates existing data to newer MDES schemas with as little human intervention as possible



Next Steps



- Provide an open source, collaborative framework for collection instrument life cycle management
- Provide an open source, collaborative framework for versioning and sharing changes to the data elements, rules and definitions in the MDES
- Provide tools for automating the creation of validated data collection instruments from the MDES definitions

The goal is on demand central data submissions, data exchange between sites, and 'future proofing' the data repository





Meeting the Informatics Challenges of the National Children's Study

William R. Hogan, MD, MS
NCS Research Day
August 24th, 2011

Meeting the Challenges: Point of Embarkation



- No adequate software systems available “out of the box”
 - Must therefore assemble a suite
 - Lack of standards → integration is hard
 - Must adapt components to an ambitious protocol
- Minimal functionality for:
 - Data quality
 - Reporting
 - Extraction of data for submission



A Few Guiding Principles for the Journey



- Open-source systems are:
 - Easier to adapt
 - Easier to integrate
 - **Easier to “future proof”**
- Division of labor and collaborative intelligence will get us there faster
 - Software
 - Instruments
 - Standards, common definitions, etc.
 - Documentation
 - *Innovation*



The Journey So Far



- We and others have:
 - Scoured the landscape for appropriate open-source applications
 - Started integration and adaptation efforts
 - Worked to build communities that make faster progress than in insolation
- The Program Office has:
 - Paid careful attention to standardization
 - Developed methods to analyze quality of data it receives



In Arkansas



- Open-source applications in use
 - caBIG Central Clinical Participant Registry (C3PR) – *participant registration and consent information*
 - LimeSurvey – *instruments*
 - Open sYStem for Entity Resolution (OYSTER) – *resolving duplicate address records*
- Significant adaptations to C3PR
 - Move participants back and forth among epochs
 - Address/participant entity resolution to resolve duplicate records
 - **Restructuring of consents, “checkboxes”, refusals to participate, etc.**
- One FTE required to build and update instruments in LimeSurvey



Arkansas Experience Demonstrates That...



- The use of open-source software has indeed enabled more rapid progress
- Modifications of software to meet NCS requirements has benefited other research, and vice versa
- We have benefited from development done by other study centers



Take Home Messages for the NCS Advisory Committee



- Research informatics is a young field that is evolving rapidly.
 - The answers to the NCS needs are not available off the shelf
- Complexity is a killer. Simple is harder than complex
 - Just because something can be done does not mean it should be done
 - **It takes more leadership to say "No" than "Yes"**
- The future is uncertain. Avoid lock-in at every step
 - Community-based development is messy and slower but it is the only way to ensure broad acceptance





I am available to help in ANY capacity.
Thank you for this opportunity.

Michael.Kahn@ucdenver.edu

