An Ontology-Based Framework for Clinical Research Databases

Megan Kong[1], Carl Dahlke[2], Diane Xiang[1], David Karp[4], Richard H. Scheuermann[1,3]

[1]Department of Pathology, [3]Division of Biomedical Informatics, [4]Division of Rheumatology, U.T. Southwestern Medical Center, Dallas, TX, [2] Health Information Systems, Northrop Grumman, Inc., Rockville, MD

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Bioinformatics Integration Support Contract (BISC)

- NIH/NIAID/DAIT would like to:
  - maximize the return on the public investment in basic, translational and clinical research
  - allow investigators to more effectively extract meaningful information from the vast amounts of data generated from advanced research technologies
- Bioinformatics Integration Support Contract (BISC) to support all DAIT-funded programs - basic, translational and clinical research
- Immunology Database and Analysis Portal (***ImmPort***) **-** www.ImmPort.org
  - Archive and manage basic and clinical research data
  - Analyze experiment data
  - Integrate with extensive biological knowledge
  - Analyze integrated data
  - Build upon data generated by previous studies

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

ImmPort Overview - System Components

- **Semi-public web-based database and analysis portal**
  - Multi-level access control
  - Data sharing
- **Data**
  - **Reference data**
    - **Types** - Gene structure, protein function, polymorphisms, metabolic, regulatory, signaling and other networks, protein-protein, gene-gene, host-pathogen interactions
    - **Sources** - NCBI, Uniprot, Swissprot, BIND, Reactome
  - **Experiment data**
    - **Methodologies** – microarrays, SNP genotyping, flow cytometry, ELISA, ELISPOT, qPCR, imaging, etc.
    - **Metadata** (defining how the experiment was performed) - common features of all experiments
    - **Primary results** - from all experiment measurement techniques
    - **Processed results**
      - Interpreted results
      - Analytical metadata
  - **Clinical research data**
    - **Study design** – from clinical protocol
    - **Participant characteristics/phenotype** – from assessments captured in CRF's and laboratory testing
- **Query tools**
  - To support retrieval of reference and experiment data based on specified criteria
  - Pre-defined QBE
  - Customized semantic queries
- **Ontology**
  - Thesaurus function
  - Organize terms and define relationships

- **Analysis tools**
  - **Genetic analysis**
    - LD analysis
    - TagSNP selection
    - Haplotype reconstruction
    - Genotype-phenotype association
  - **Gene expression analysis**
    - Filtering/normalization
    - Clustering
    - Classification
  - **Cell population analysis**
    - Standard FACS statistics
    - Novel population identification based on high dimensional data clustering
  - **Measurement of immune response (**e.g., ELISA, ELISPOT)
    - Statistical analysis of distributions
  - **Biological network analysis**
    - Quantification of topological parameters
    - Module identification
- **Visualization tools**
  - Genome display, including genes, introns, exons, SNPs, tagSNPs, etc.
  - Genetic analysis results
  - Gene expression results
  - Networks, pathways, and molecular interactions
  - Graphing and charting of statistical results

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Initial work

- Population Genetics Analysis Program: Immunity to Vaccines/Infections (PopGen)
- Immune Function and Biodefense in Children, Elderly, and Immunocompromised Populations (Special Pops)
- HLA Region Genetics in Immune-mediated Diseases
- Immune Tolerance Network (ITN)
- Atopic Dermatitis and Vaccinia Network (ADVN)
- NIAID non-human primate colony

ImmPort
Bioinformatics For The Future of Immunology

Varicella:

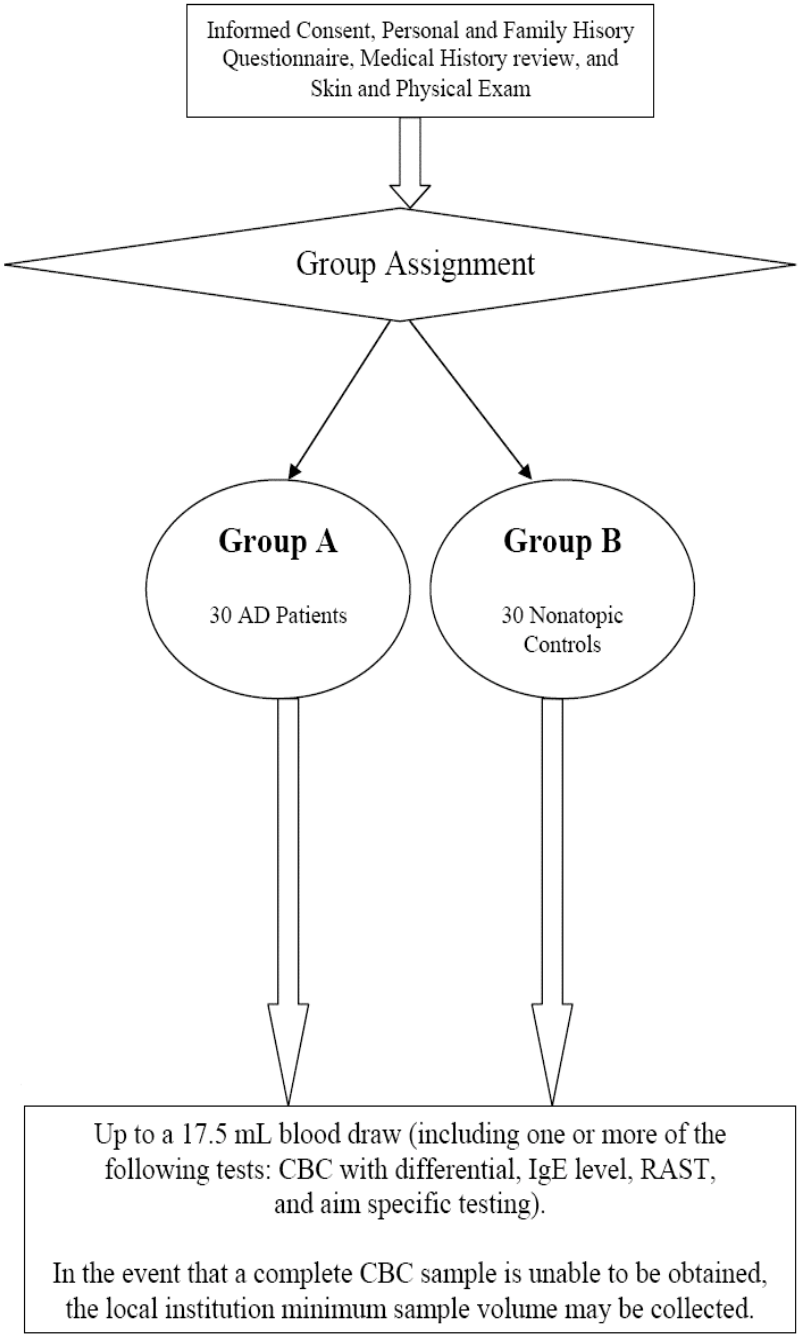
Immune Response to Varicella Vaccination in Subjects with Atopic Dermatitis Compared to Nonatopic Controls

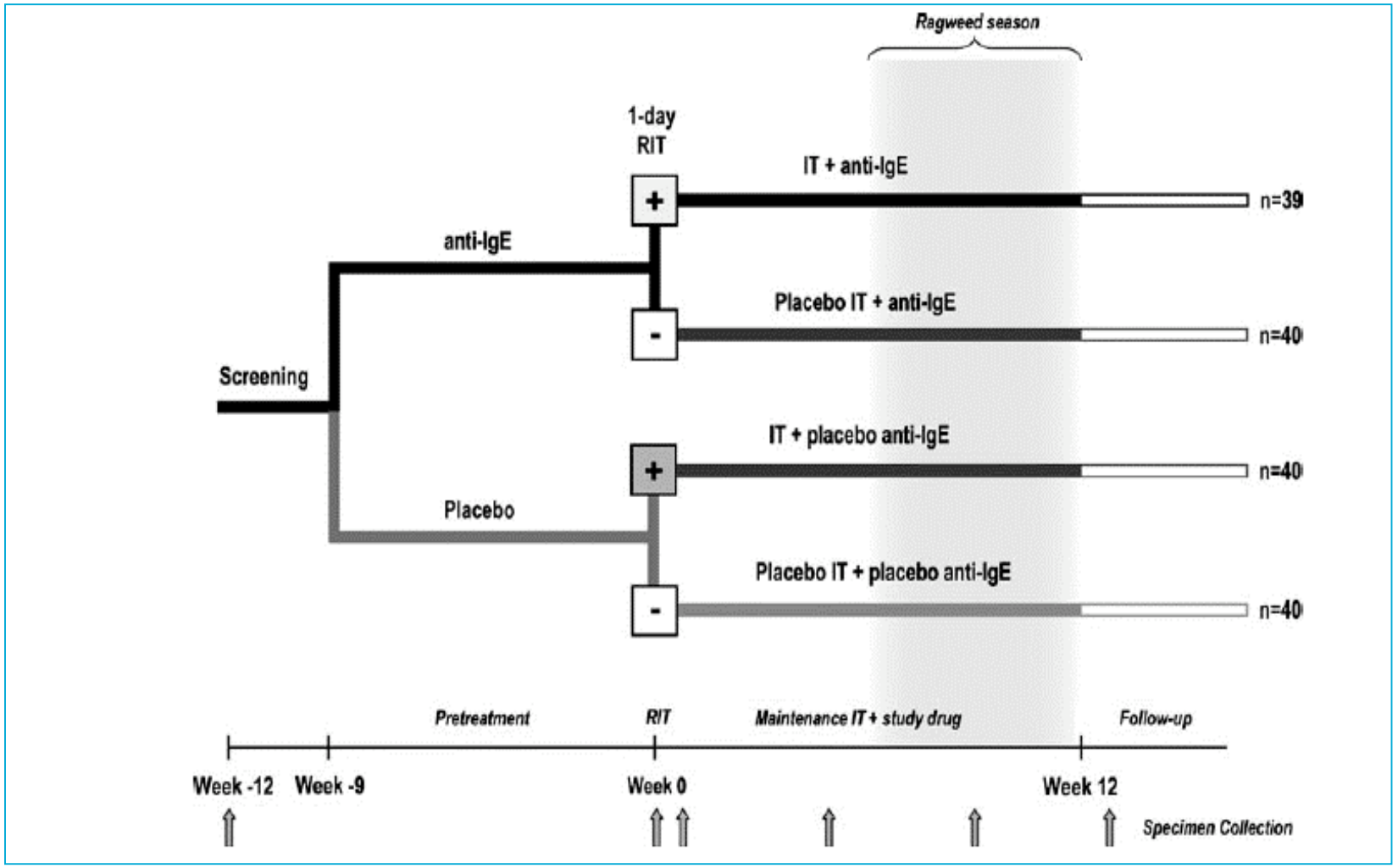
Objectives:

To determine if children with AD have Varicella-specific cell mediated immune (CMI) responses to varicella vaccination that differ from those of nonatopic controls (Th2)

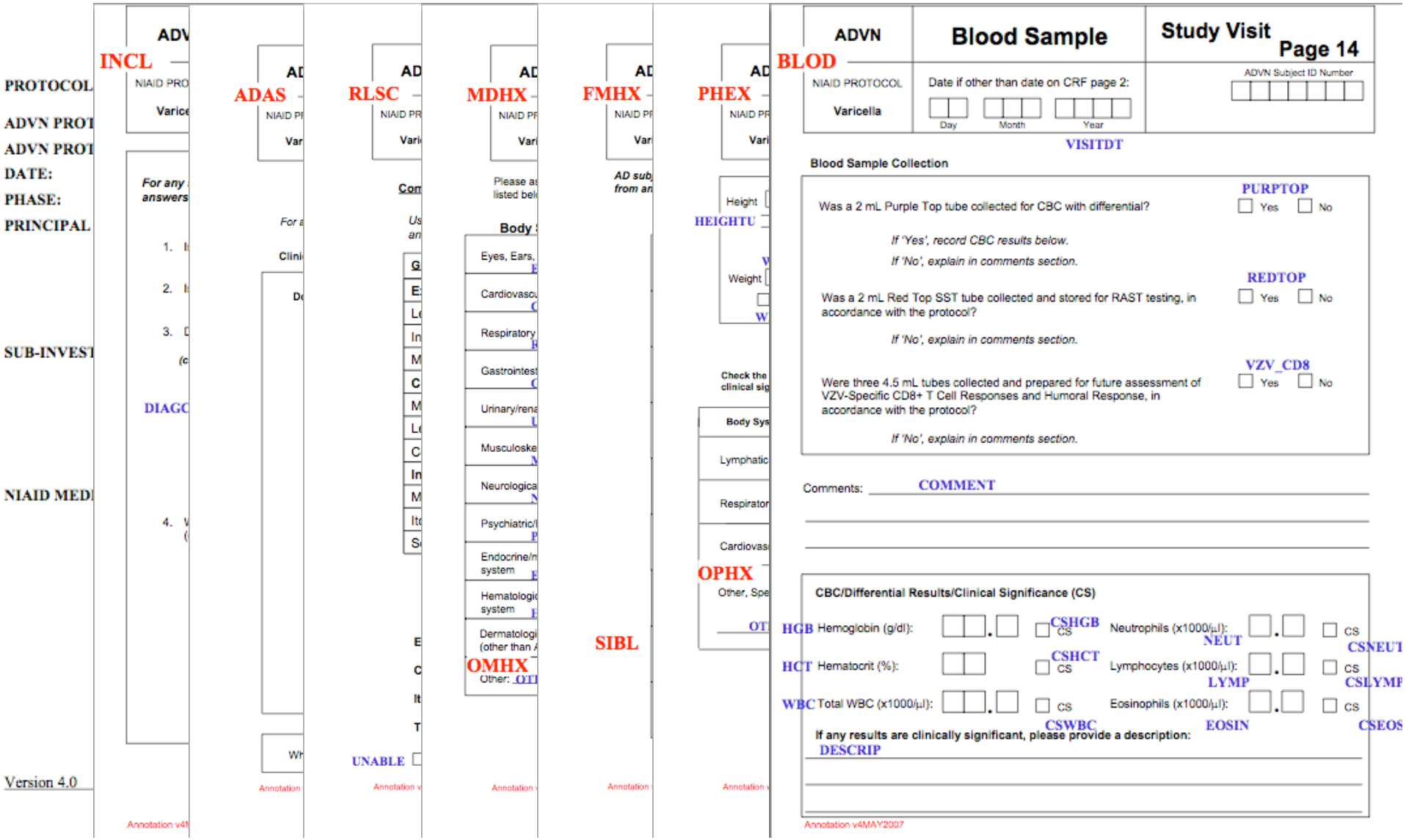
Endpoint:

1. ELISPOT: number of Varicella-specific T-cell
2. ELISA: levels of Varicella-specific antibodies
3. Flow cytometry: perforin generation using 2-color flow cytometry staining for both CD8 and perforin.
   (Perforin levels, which may contribute to defects in CD8+ cytotoxic T lymphocyte function).




Ragweed season
1-day
RIT
IT + anti-IgE
n=39
anti-IgE
Placebo IT + anti-IgE
n=40
Screening
IT + placebo anti-IgE
n=40
Placebo
Placebo IT + placebo anti-IgE
n=40
Pretreatment
RIT
Maintenance IT + study drug
Follow-up
Week -12
Week -9
Week 0
Week 12
Specimen Collection

Driving Projects - ADVN



IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Why OBX?

- Challenges
  - Extensive phenotypic characteristics of interest
  - Complex study designs of treatments, assessments, samplings, etc.
  - Few standards for how they are described or captured
  - Each supported project captures their data in a different format
  - Integration with mechanistic studies of specimens

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Why OBX?

- Challenges
  - Extensive phenotypic characteristics of interest
  - Complex study designs of treatments, assessments, samplings, etc.
  - Few standards for how they are described or captured
  - Each supported project captures their data in a different format
  - Integration with mechanistic studies of specimens
- Evaluation of existing clinical data standards
  - BRIDG – designed for supporting the conduct of clinical trials, not a study repository optimized for analysis
  - CDISC STDM – specification for SAS transport files to the FDA
  - CDISC CDASH – models clinical encounters, but not sequence of events; requires use of EPOCHs (contiguous blocks of time)
  - None designed to manage mechanistic studies

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Why OBX?

- Challenges
  - Extensive phenotypic characteristics of interest
  - Complex study designs of treatments, assessments, samplings, etc.
  - Few standards for how they are described or captured
  - Each supported project captures their data in a different format
  - Integration with mechanistic studies of specimens
- Evaluation of existing clinical data standards
  - BRIDG – designed for supporting the conduct of clinical trials, not a study repository optimized for analysis
  - CDISC STDM – specification for SAS transport files to the FDA
  - CDISC CDASH – models clinical encounters, but not sequence of events; requires use of EPOCHs (contiguous blocks of time)
  - None designed to manage mechanistic studies
- Ontology-Based eXtensible Data Model (OBX)
  - Logical ontology structure (BFO/OBI) => extensible data model
  - Data element value sets derived from ontology hierarchy

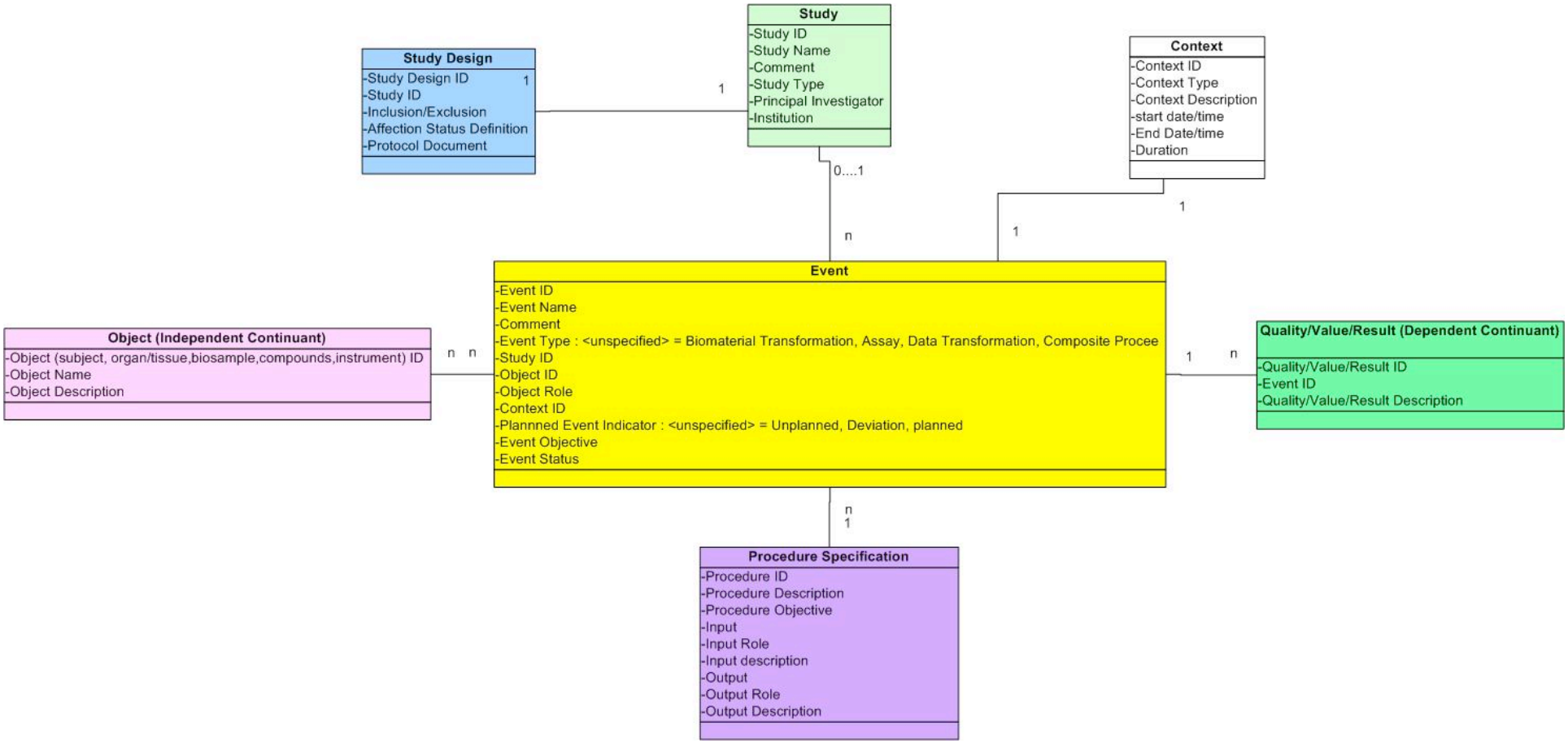
IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

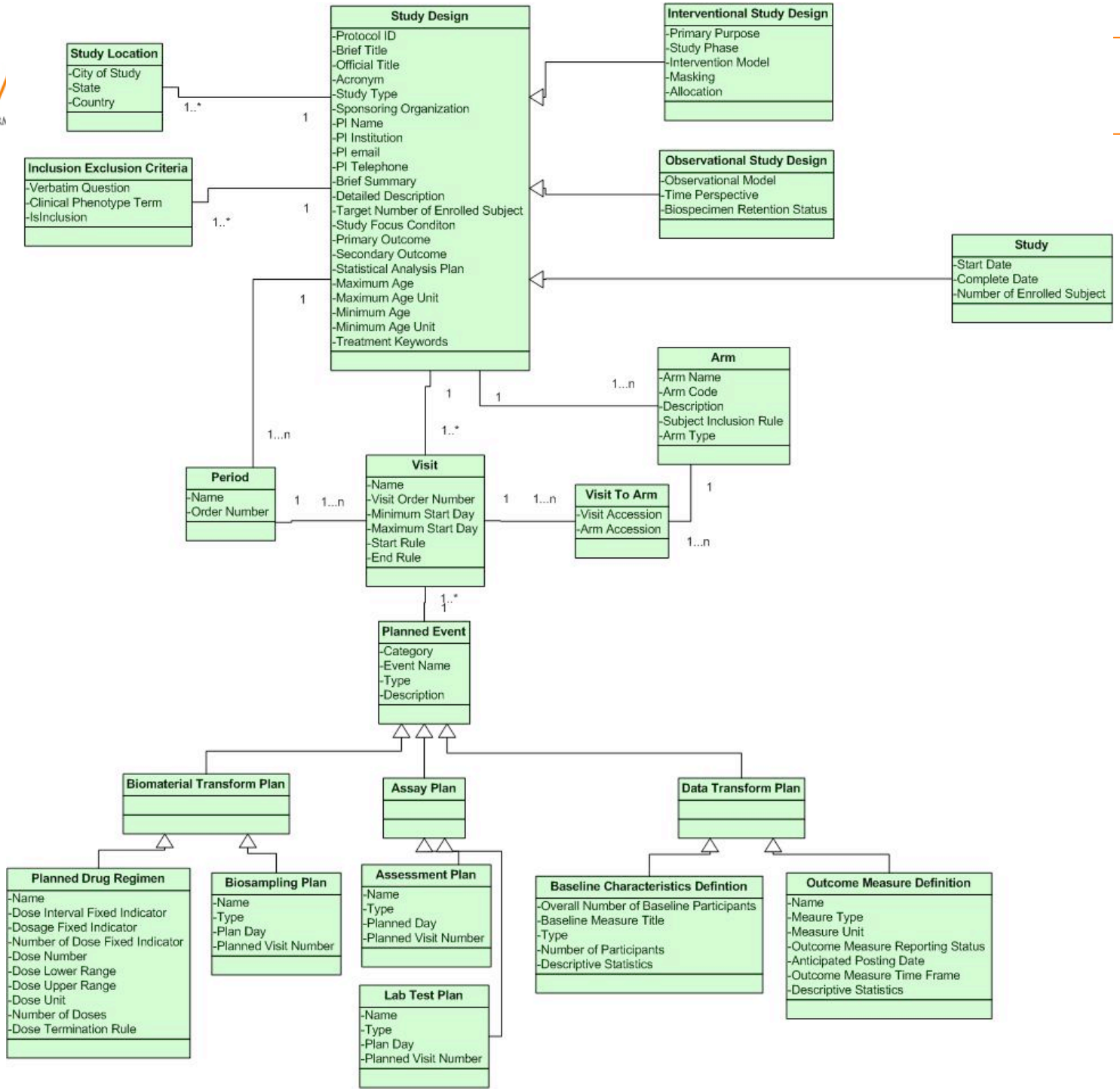
Entities of Interest

- Objects
  - Human subjects
  - Specimens
  - Materials
  - Plans
- Qualities – clinical phenotype
  - Malar rash
  - WBC count
  - Weight
  - Blood glucose levels
  - Atopic dermatitis
- Roles
  - Principal investigator, study coordinator, study participant
  - Reagent, therapeutic
- Processes
  - Recruitment, enrollment, approval
  - Assays, physical exams
  - Blood draw, PBMC isolation
  - Treatment, surgery (interventions)
- Context
  - Time
  - Environment

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

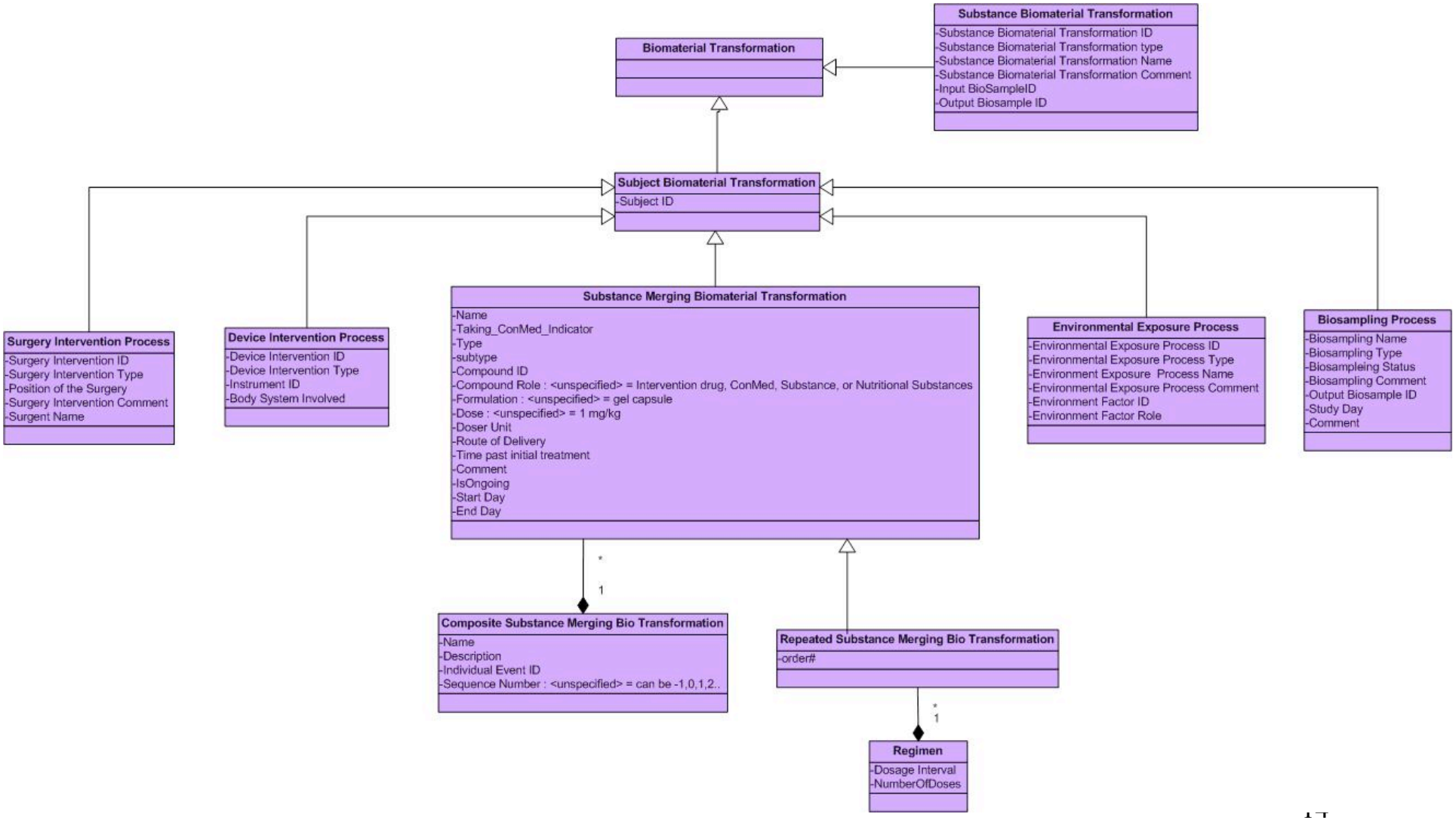
OBX Framework



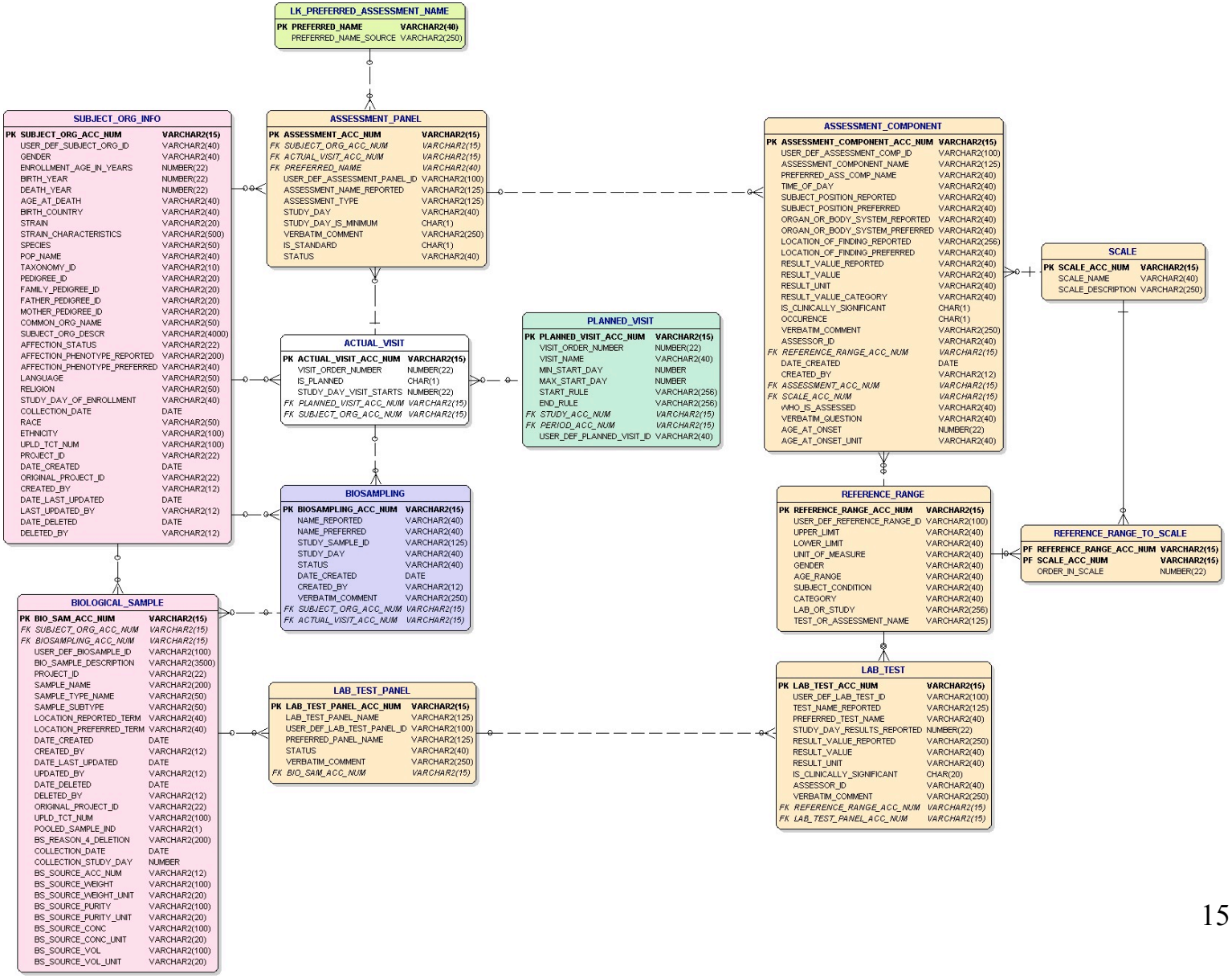

Study Design
-Protocol ID
-Brief Title
-Official Title
-Acronym
-Study Type
-Sponsoring Organization
-PI Name
-PI Institution
-PI email
-PI Telephone
-Brief Summary
-Detailed Description
-Target Number of Enrolled Subject
-Study Focus Conditon
-Primary Outcome
-Secondary Outcome
-Statistical Analysis Plan
-Maximum Age
-Maximum Age Unit
-Minimum Age
-Minimum Age Unit
-Treatment Keywords
Study Location
-City of Study
-State
-Country
Interventional Study Design
-Primary Purpose
-Study Phase
-Intervention Model
-Masking
-Allocation
Observational Study Design
-Observational Model
-Time Perspective
-Biospecimen Retention Status
Inclusion Exclusion Criteria
-Verbatim Question
-Clinical Phenotype Term
-IsInclusion
Study
-Start Date
-Complete Date
-Number of Enrolled Subject
Arm
-Arm Name
-Arm Code
-Description
-Subject Inclusion Rule
-Arm Type
Period
-Name
-Order Number
Visit
-Name
-Visit Order Number
-Minimum Start Day
-Maximum Start Day
-Start Rule
-End Rule
Visit To Arm
-Visit Accession
-Arm Accession
Planned Event
-Category
-Event Name
-Type
-Description
Biomaterial Transform Plan
Assay Plan
Data Transform Plan
Planned Drug Regimen
-Name
-Dose Interval Fixed Indicator
-Dosage Fixed Indicator
-Number of Dose Fixed Indicator
-Dose Number
-Dose Lower Range
-Dose Upper Range
-Dose Unit
-Number of Doses
-Dose Termination Rule
Biosampling Plan
-Name
-Type
-Plan Day
-Planned Visit Number
Assessment Plan
-Name
-Type
-Planned Day
-Planned Visit Number
Lab Test Plan
-Name
-Type
-Plan Day
-Planned Visit Number
Baseline Characteristics Defintion
-Overall Number of Baseline Participants
-Baseline Measure Title
-Type
-Number of Participants
-Descriptive Statistics
Outcome Measure Definition
-Name
-Meaure Type
-Measure Unit
-Outcome Measure Reporting Status
-Anticipated Posting Date
-Outcome Measure Time Frame
-Descriptive Statistics

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

OBX Biomaterial Transformation



Assay Database Schema



Lab test mapping into OBX

Blood Test:

ADVN Fields	ADVN Fields Description	ImmPort Table	ImmPort Attribute
ID	Subject ID	Lab Test	Subject ID
SITE	Study Site	will not capture	
VISITDT	VISIT Date	Lab Test	Visit Date
HGB	Hemoglobin (g/dl):	Lab Test	Type = Blood Test Subtype = Hemoglobin Measurement Test Result Value = HGB Test Result Unit = grams per deciliter Clinical Significance = CSHGB
CSHGB	Hemoglobin Clinical Significant	Lab Test	
HCT	Hematocrit (%):	Lab Test	Type = Blood Test Subtype = Hematocrit Measurement Test Result Value = HCT Test Result Unit = percentage Clinical Significance = CSHCT
CSHCT	Hematocrit Clinical Significant	Lab Test	
WBC	Total WBC (x1000/µl):	Lab Test	Type = Blood Test Subtype = Total White Blood Cell Measurement Test Result Value = HGB Test Result Unit = 1000/µl Clinical Significance = CSHGB
CSWBC	Total WBC Clinical Significant	Lab Test	
NEUT	Neutrophils (x1000/microliter):	Lab Test	Type = Blood Test Subtype = Neutrophils Measurement Test Result Value = NEUT Test Result Unit = 1000/µl Clinical Significance = CSNEUT
CSNEUT	Neutrophils Clinical Significant	Lab Test	
LYMP	Lymphocytes (x1000/µl):	Lab Test	Type = Blood Test Subtype = Lymphocytes Measurement Test Result Value = LYMP Test Result Unit = 1000/µl Clinical Significance = CSLYMP
CSLYMP	Lymphocytes Clinical Significant	Lab Test	
EOSIN	Eosinophils (x1000/µl):	Lab Test	Type = Blood Test Subtype = Eosinophils Measurement Test Result Value = EOSIN Test Result Unit = 1000/µl Clinical Significance = CSEOS
CSEOS	Eosinophils Clinical Significant	Lab Test	

Load File Examples

Lab_Test_Panel_Acc_	Test_Name_Repor	Result_Value	Result_Unit
0101780 blod	Hemoglobin	13.3	g/dl
0315450 blod	Hemoglobin	12.8	g/dl
0315483 blod	Hemoglobin	10.8	g/dl
0315931 blod	Hemoglobin	13.6	g/dl
0316415 blod	Hemoglobin	12.3	g/dl
0101780 blod	Hematocrit	41	%
0315450 blod	Hematocrit	35	%
0315483 blod	Hematocrit	32	%
0315931 blod	Hematocrit	40	%
0316415 blod	Hematocrit	34	%
0101780 blod	Total WBC	12.7	1000/microliter
0315450 blod	Total WBC	5.22	1000/microliter
0315483 blod	Total WBC	4.53	1000/microliter
0315931 blod	Total WBC	8.8	1000/microliter
0316415 blod	Total WBC	11.25	1000/microliter
0101780 blod	Total Neutrophils	3.3	1000/microliter
0315450 blod	Total Neutrophils	0.31	1000/microliter
0315483 blod	Total Neutrophils	1.81	1000/microliter
0315931 blod	Total Neutrophils	1.14	1000/microliter
0316415 blod	Total Neutrophils	3.39	1000/microliter
0101780 blod	Lymphocytes	7.6	1000/microliter
0315450 blod	Lymphocytes	3.92	1000/microliter
0315483 blod	Lymphocytes	1.9	1000/microliter
0315931 blod	Lymphocytes	7.04	1000/microliter
0316415 blod	Lymphocytes	6.22	1000/microliter
0101780 blod	Eosinophils	0.9	1000/microliter
0315450 blod	Eosinophils	0.21	1000/microliter
0315483 blod	Eosinophils	0.18	1000/microliter
0315931 blod	Eosinophils	0	1000/microliter
0316415 blod	Eosinophils	0.98	1000/microliter

Assessment_Compo	Assessment_Co	Location_of_Findi	Organ_or_Body_	Result_Value
0101780 easi	Erythema	Head	Skin	Moderate
0104097 easi	Erythema	Head	Skin	Moderate
0104302 easi	Erythema	Head	Skin	Mild
0104468 easi	Erythema	Head	Skin	Mild
0106229 easi	Erythema	Head	Skin	Mild
0101780 easi	Erythema	Trun	Skin	Absent
0104097 easi	Erythema	Trun	Skin	Moderate
0104302 easi	Erythema	Trun	Skin	Absent
0104468 easi	Erythema	Trun	Skin	Absent
0106229 easi	Erythema	Trun	Skin	Moderate
0101780 easi	Erythema	Arms	Skin	Mild
0104097 easi	Erythema	Arms	Skin	Absent
0104302 easi	Erythema	Arms	Skin	Mild
0104468 easi	Erythema	Arms	Skin	Mild
0106229 easi	Erythema	Arms	Skin	Mild
0101780 easi	Indurati	Head	Skin	Moderate
0104097 easi	Indurati	Head	Skin	Mild
0104302 easi	Indurati	Head	Skin	Mild
0104468 easi	Indurati	Head	Skin	Mild
0106229 easi	Indurati	Head	Skin	Mild
0101780 easi	Lichenif	Head	Skin	Absent
0104097 easi	Lichenif	Head	Skin	Mild
0104302 easi	Lichenif	Head	Skin	Absent
0104468 easi	Lichenif	Head	Skin	Absent
0106229 easi	Lichenif	Head	Skin	Absent

IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Summary

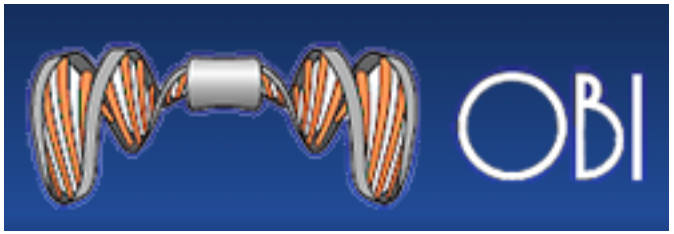
- OBX is an ontology-informed clinical research data model
- Ontological framework incorporated provides for extensibility and straightforward incorporation of ontology terms as value sets
- Ontological framework also facilitated distributed data model development and guides ongoing mapping strategies
- Successfully loaded two large, distinct clinical studies
- Building the user interfaces for data extraction and analysis
- Refinement through use – mapping, loading, extracting, analysis
- Term submission to OBO Foundry ontologies, especially OBI

Acknowledgments

UT Southwestern
Yu (Max) Qian
Jamie Lee
Megan Kong
Jennifer Cai
Jie Huang
Nishanth Marthandan
Diane Xiang
Young Bun Kim
Paula Guidry
Eva Sadat
David Karp

Northrop Grumman
Carl Dahlke
John Campbell
Liz Thompson
Jeff Wiser
Mike Attasi

OBO Foundry
BFO
OBI



Supported by NIH N01AI40076