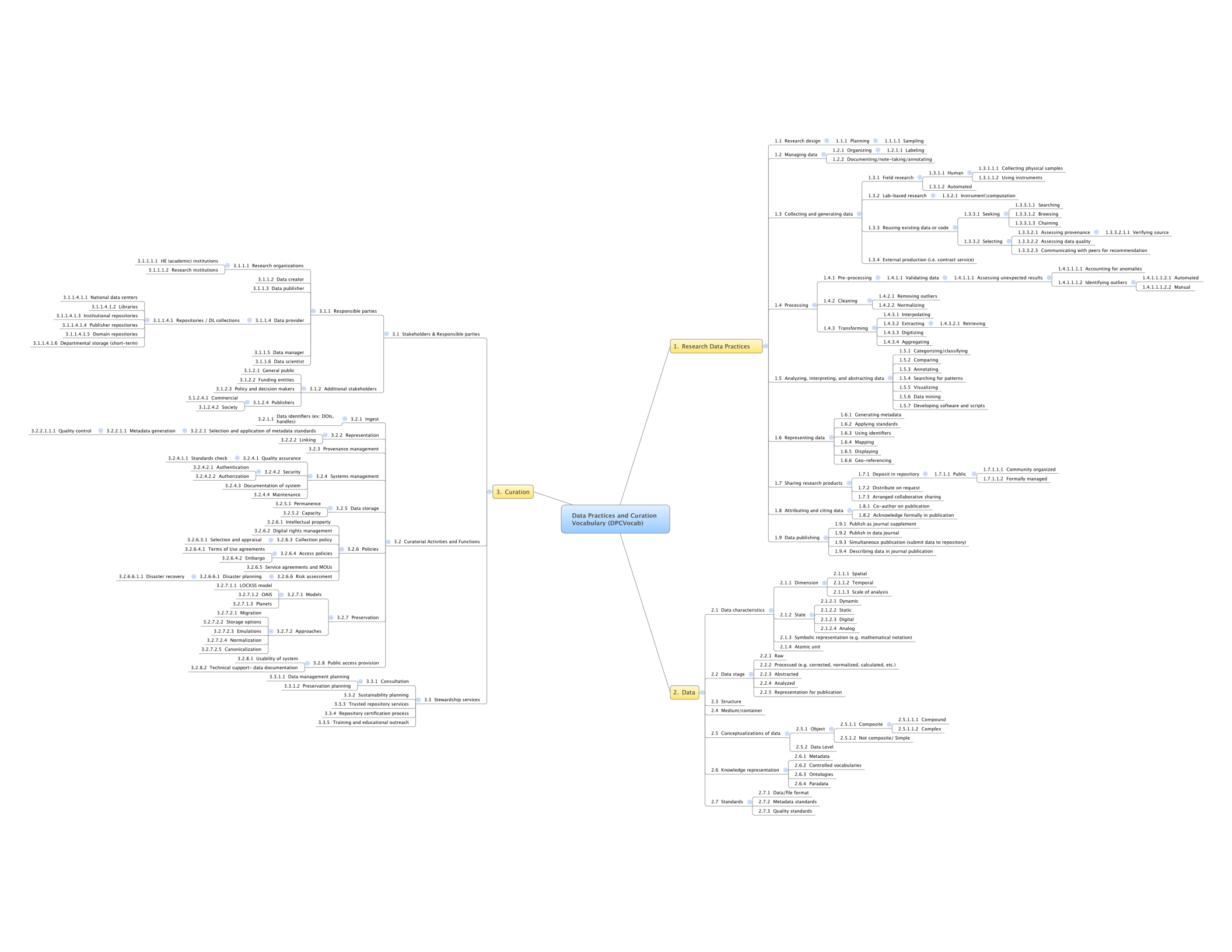
Data Practices and Curation Vocabulary (DPCVocab)
1. Research Data Practices
1.1 Research design
1.1.1 Planning
1.1.1.1 Sampling
1.2 Managing data
1.2.1 Organizing
1.2.1.1 Labeling
1.2.2 Documenting/note-taking/annotating
1.3 Collecting and generating data
1.3.1 Field research
1.3.1.1 Human
1.3.1.1.1 Collecting physical samples
1.3.1.1.2 Using instruments
1.3.1.2 Automated
1.3.2 Lab-based research
1.3.2.1 Instrument\computation
1.3.3 Reusing existing data or code
1.3.3.1 Seeking
1.3.3.1.1 Searching
1.3.3.1.2 Browsing
1.3.3.1.3 Chaining
1.3.3.2 Selecting
1.3.3.2.1 Assessing provenance
1.3.3.2.1.1 Verifying source
1.3.3.2.2 Assessing data quality
1.3.3.2.3 Communicating with peers for recommendation
1.3.4 External production (i.e. contract service)
1.4 Processing
1.4.1 Pre-processing
1.4.1.1 Validating data
1.4.1.1.1 Assessing unexpected results
1.4.1.1.1.1 Accounting for anomalies
1.4.1.1.1.2 Identifying outliers
1.4.1.1.2.1 Automated
1.4.1.1.2.2 Manual
1.4.2 Cleaning
1.4.2.1 Removing outliers
1.4.2.2 Normalizing
1.4.3 Transforming
1.4.3.1 Interpolating
1.4.3.2 Extracting
1.4.3.2.1 Retrieving
1.4.3.3 Digitizing
1.4.3.4 Aggregating
1.5 Analyzing, interpreting, and abstracting data
1.5.1 Categorizing/classifying
1.5.2 Comparing
1.5.3 Annotating
1.5.4 Searching for patterns
1.5.5 Visualizing
1.5.6 Data mining
1.5.7 Developing software and scripts
1.6 Representing data
1.6.1 Generating metadata
1.6.2 Applying standards
1.6.3 Using identifiers
1.6.4 Mapping
1.6.5 Displaying
1.6.6 Geo-referencing
1.7 Sharing research products
1.7.1 Deposit in repository
1.7.1.1 Public
1.7.1.1.1 Community organized
1.7.1.1.2 Formally managed
1.7.2 Distribute on request
1.7.3 Arranged collaborative sharing
1.8 Attributing and citing data
1.8.1 Co-author on publication
1.8.2 Acknowledge formally in publication
1.9 Data publishing
1.9.1 Publish as journal supplement
1.9.2 Publish in data journal
1.9.3 Simultaneous publication (submit data to repository)
1.9.4 Describing data in journal publication
2. Data
2.1 Data characteristics
2.1.1 Dimension
2.1.1.1 Spatial
2.1.1.2 Temporal
2.1.1.3 Scale of analysis
2.1.2 State
2.1.2.1 Dynamic
2.1.2.2 Static
2.1.2.3 Digital
2.1.2.4 Analog
2.1.3 Symbolic representation (e.g. mathematical notation)
2.1.4 Atomic unit
2.2 Data stage
2.2.1 Raw
2.2.2 Processed (e.g. corrected, normalized, calculated, etc.)
2.2.3 Abstracted
2.2.4 Analyzed
2.2.5 Representation for publication
2.3 Structure
2.4 Medium/container
2.5 Conceptualizations of data
2.5.1 Object
2.5.1.1 Composite
2.5.1.1.1 Compound
2.5.1.1.2 Complex
2.5.1.2 Not composite/ Simple
2.5.2 Data Level
2.6 Knowledge representation
2.6.1 Metadata
2.6.2 Controlled vocabularies
2.6.3 Ontologies
2.6.4 Paradata
2.7 Standards
2.7.1 Data/file format
2.7.2 Metadata standards
2.7.3 Quality standards
3. Curation
3.1 Stakeholders & Responsible parties
3.1.1 Responsible parties
3.1.1.1 Research organizations
3.1.1.1.1 HE (academic) institutions
3.1.1.1.2 Research institutions
3.1.1.2 Data creator
3.1.1.3 Data publisher
3.1.1.4 Data provider
3.1.1.4.1 Repositories / DL collections
3.1.1.4.1.1 National data centers
3.1.1.4.1.2 Libraries
3.1.1.4.1.3 Institutional repositories
3.1.1.4.1.4 Publisher repositories
3.1.1.4.1.5 Domain repositories
3.1.1.4.1.6 Departmental storage (short-term)
3.1.1.5 Data manager
3.1.1.6 Data scientist
3.1.2 Additional stakeholders
3.1.2.1 General public
3.1.2.2 Funding entities
3.1.2.3 Policy and decision makers
3.1.2.4 Publishers
3.1.2.4.1 Commercial
3.1.2.4.2 Society
3.2 Curatorial Activities and Functions
3.2.1 Ingest
3.2.1.1 Data identifiers (ex: DOIs, handles)
3.2.2 Representation
3.2.2.1 Selection and application of metadata standards
3.2.2.1.1 Quality control
3.2.2.1.1.1 Metadata generation
3.2.2.2 Linking
3.2.3 Provenance management
3.2.4 Systems management
3.2.4.1 Quality assurance
3.2.4.1.1 Standards check
3.2.4.2 Security
3.2.4.2.1 Authentication
3.2.4.2.2 Authorization
3.2.4.3 Documentation of system
3.2.4.4 Maintenance
3.2.5 Data storage
3.2.5.1 Permanence
3.2.5.2 Capacity
3.2.6 Policies
3.2.6.1 Intellectual property
3.2.6.2 Digital rights management
3.2.6.3 Collection policy
3.2.6.3.1 Selection and appraisal
3.2.6.4 Access policies
3.2.6.4.1 Terms of Use agreements
3.2.6.4.2 Embargo
3.2.6.5 Service agreements and MOUs
3.2.6.6 Risk assessment
3.2.6.6.1 Disaster planning
3.2.6.6.1.1 Disaster recovery
3.2.7 Preservation
3.2.7.1 Models
3.2.7.1.1 LOCKSS model
3.2.7.1.2 OAIS
3.2.7.1.3 Planets
3.2.7.2 Approaches
3.2.7.2.1 Migration
3.2.7.2.2 Storage options
3.2.7.2.3 Emulations
3.2.7.2.4 Normalization
3.2.7.2.5 Canonicalization
3.2.8 Public access provision
3.2.8.1 Usability of system
3.2.8.2 Technical support- data documentation
3.3 Stewardship services
3.3.1 Consultation
3.3.1.1 Data management planning
3.3.1.2 Preservation planning
3.3.2 Sustainability planning
3.3.3 Trusted repository services
3.3.4 Repository certification process
3.3.5 Training and educational outreach