What we talk about when we talk about

DataTags

Internal Project Presentation, May 2014

IQSS

Data Privacy Lab

Dataverse Network PROJECT

Privacy Tools
for Sharing Research Data

Mission Statement

DataTags will provide dataset owners with simple data handling prescriptions that comply with the numerous regulations and data use agreements that may apply to a given dataset.

…and do it in an easy, researcher-friendly way

Mission Statement

DataTags will provide dataset owners with simple data handling prescriptions that comply with the numerous regulations and data use agreements that may apply to a given dataset.

…and do it in an easy, researcher-friendly way

Prescriptions are human-readable and machine-actionable

components

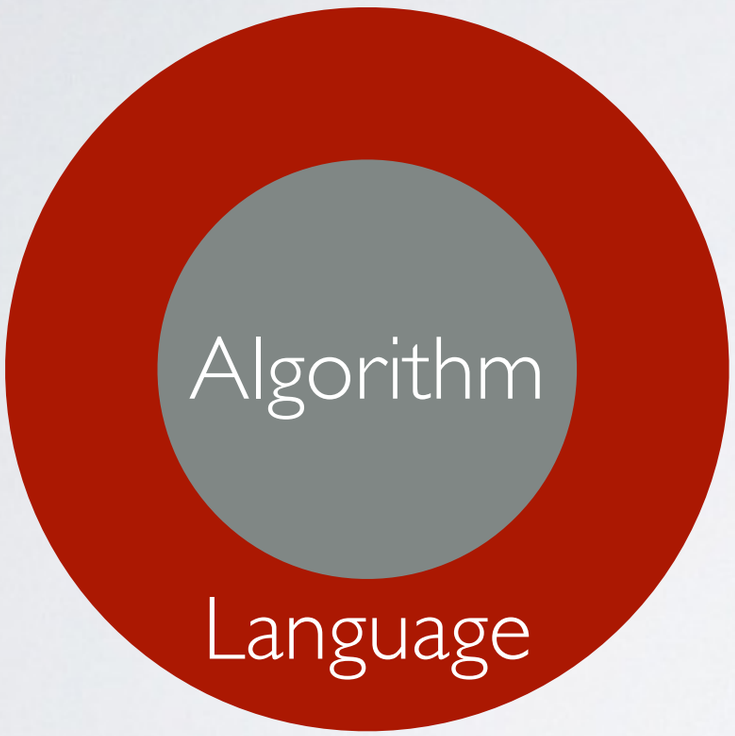
Algorithm

Algorithm

- "Harmonizes law and technology"
- Consists of a tag ontology and an interview process
- Composed by legal and technological experts
- Currently Supports HIPAA, FERPA, CIPSEA and Privacy Act
  Berkman, DPL
- Next Up: DUA

components

Language



- .dtDef
  ontology definition language
- .dtf
  Interview and coding language
- Future: inference based language

Language

dtf

- Define an interview and coding process
  - Ask Questions
  - Set values to the tags (coding)
- Operations: Ask, Set, Call, End, Todo, Reject
  - future: add “if”/“when” (really, “ask yourself”)
- Supports closed-ended questions. Implicit answers for easier interview creation.
- Entities can have persistent references, to allow other systems to work with the interview structure
  e.g. localizations, reference
- Supports any closed-ended questionnaire. DataTags is a private case of this.
- Bare-metal parts working. Missing important niceties such as comments, namespaces.

Language

dtf

```
(>medical-start< ask:
    (text: Person-specific. Does your data include personal information?)
    (terms:
        (data: 0s and 1s in some structured way)
        (personal information: as defined in HIPAA))
    (no:
        (set: code=green, storage=clear, transit=clear, auth=none,
            basis=notApplicable, identity=notPersonSpecific,
            harm=negligible)
        (end)
))
(>ec< ask:
(text: Explicit Consent. Did each person whose information appears in the
    data give explicit permission to share the data?)
  (yes:
    (set: basis=consent)
    (ask:
        (text: Did the consent have any restrictions on data sharing?)
        (no:  (set: code=green, storage=clear, transit=clear, auth=none))
        (yes: (call: dua)))
    (end)
))
```

Language

dtDef

- Define tagging ontology
- Allows atomic (simple), aggregate and compound values
- The tagging process involves a single, top-level (and normally compound) value.

Language

dtDef

```
DataTags: code, basis, Handling, DataType, DUA, IP, identity, FERPA, CIPSEA.
TODO: IP.
code: one of
  blue    (Non-confidential information that can be stored and shared freely.),
  green   (Potentially identifiable but not…),
  yellow  (Potentially harmful personal information…),
  orange  (May include sensitive, identifiable personal information…),
  red     (Very sensitive identifiable personal information…),
  crimson (Requires explicit permission for each transaction…)
.
Handling: storage, transit, authentication, auth.
storage: one of clear, encrypt, doubleEncrypt.
standards: some of HIPAA, FERPA, ElectronicWiretapping, CommonRule, CIPSEA.
```

Value Types: Atomic

Aggregate

Compound

Language

dtDef

```
DataTags: code, basis, Handling, DataType, DUA, IP, identity, FERPA, CIPSEA.
TODO: IP.
code: one of
  blue    (Non-confidential information that can be stored and shared freely.),
  green   (Potentially identifiable but not…),
  yellow  (Potentially harmful personal information…),
  orange  (May include sensitive, identifiable personal information…),
  red     (Very sensitive identifiable personal information…),
  crimson (Requires explicit permission for each transaction…)
.
Handling: storage, transit, authentication, auth.
storage: one of clear, encrypt, doubleEncrypt.
standards: some of HIPAA, FERPA, ElectronicWiretapping, CommonRule, CIPSEA.
```

Value Types: Atomic

Aggregate

Compound

Language

dtDef

```
DataTags: code, basis, Handling, DataType, DUA, IP, identity, FERPA, CIPSEA.
TODO: IP.
code: one of
  blue    (Non-confidential information that can be stored and shared freely.),
  green   (Potentially identifiable but not…),
  yellow  (Potentially harmful personal information…),
  orange  (May include sensitive, identifiable personal information…),
  red     (Very sensitive identifiable personal information…),
  crimson (Requires explicit permission for each transaction…)
.
Handling: storage, transit, authentication, auth.
storage: one of clear, encrypt, doubleEncrypt.
standards: some of HIPAA, FERPA, ElectronicWiretapping, CommonRule, CIPSEA.
```

Value Types: Atomic

Aggregate

Compound

Language

dtDef

```
DataTags: code, basis, Handling, DataType, DUA, IP, identity, FERPA, CIPSEA.
TODO: IP.
code: one of
  blue    (Non-confidential information that can be stored and shared freely.),
  green   (Potentially identifiable but not…),
  yellow  (Potentially harmful personal information…),
  orange  (May include sensitive, identifiable personal information…),
  red     (Very sensitive identifiable personal information…),
  crimson (Requires explicit permission for each transaction…)
.
Handling: storage, transit, authentication, auth.
storage: one of clear, encrypt, doubleEncrypt.
standards: some of HIPAA, FERPA, ElectronicWiretapping, CommonRule, CIPSEA.
```

Value Types: Atomic

Aggregate

Compound

Language

dtDef

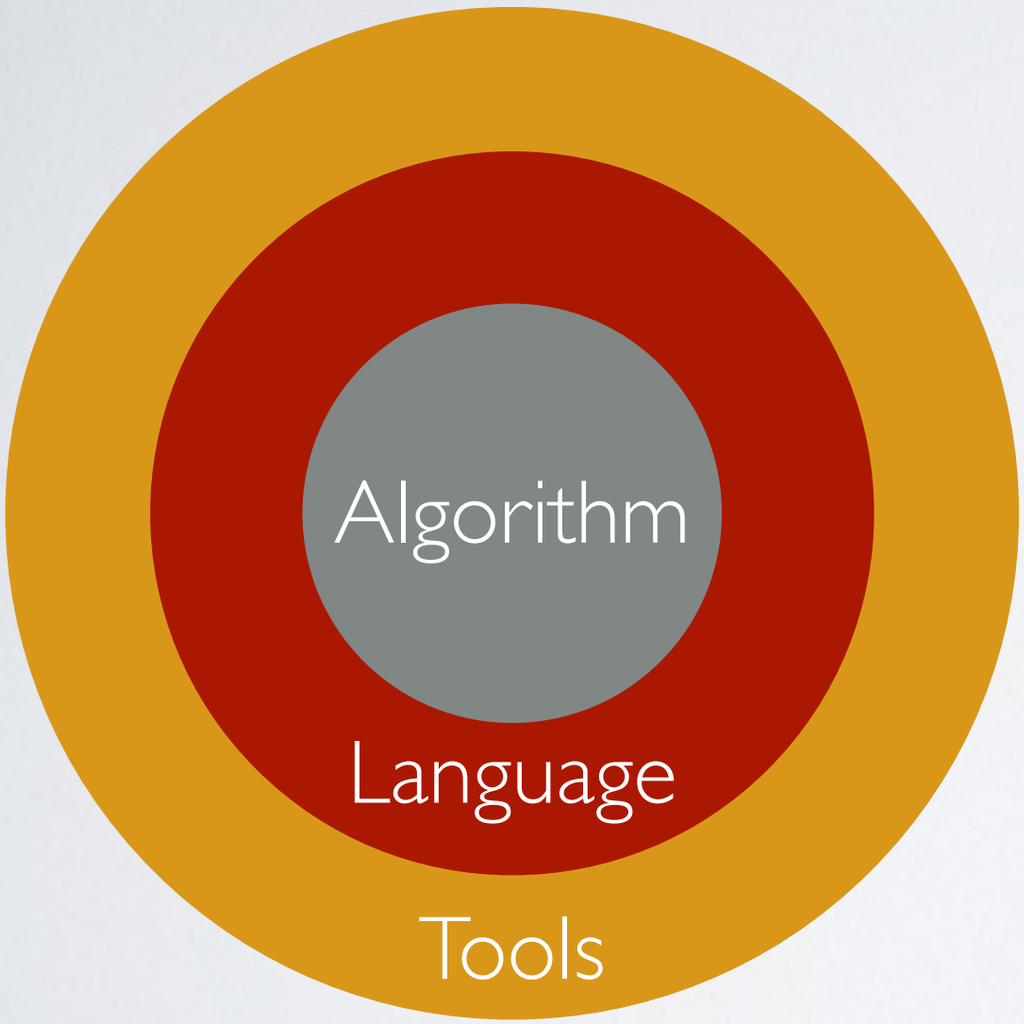
```
DataTags: code, basis, Handling, DataType, DUA, IP, identity, FERPA, CIPSEA.
TODO: IP.
code: one of
   blue    (Non-confidential information that can be stored and shared freely.),
   green   (Potentially identifiable but not…),
   yellow  (Potentially harmful personal information…),
   orange  (May include sensitive, identifiable personal information…),
   red     (Very sensitive identifiable personal information…),
   crimson (Requires explicit permission for each transaction…)
.
Handling: storage, transit, authentication, auth.
storage: one of clear, encrypt, doubleEncrypt.
standards: some of HIPAA, FERPA, ElectronicWiretapping, CommonRule, CIPSEA.
```

Value Types: Atomic, Aggregate, Compound

Supports partial definitions

components

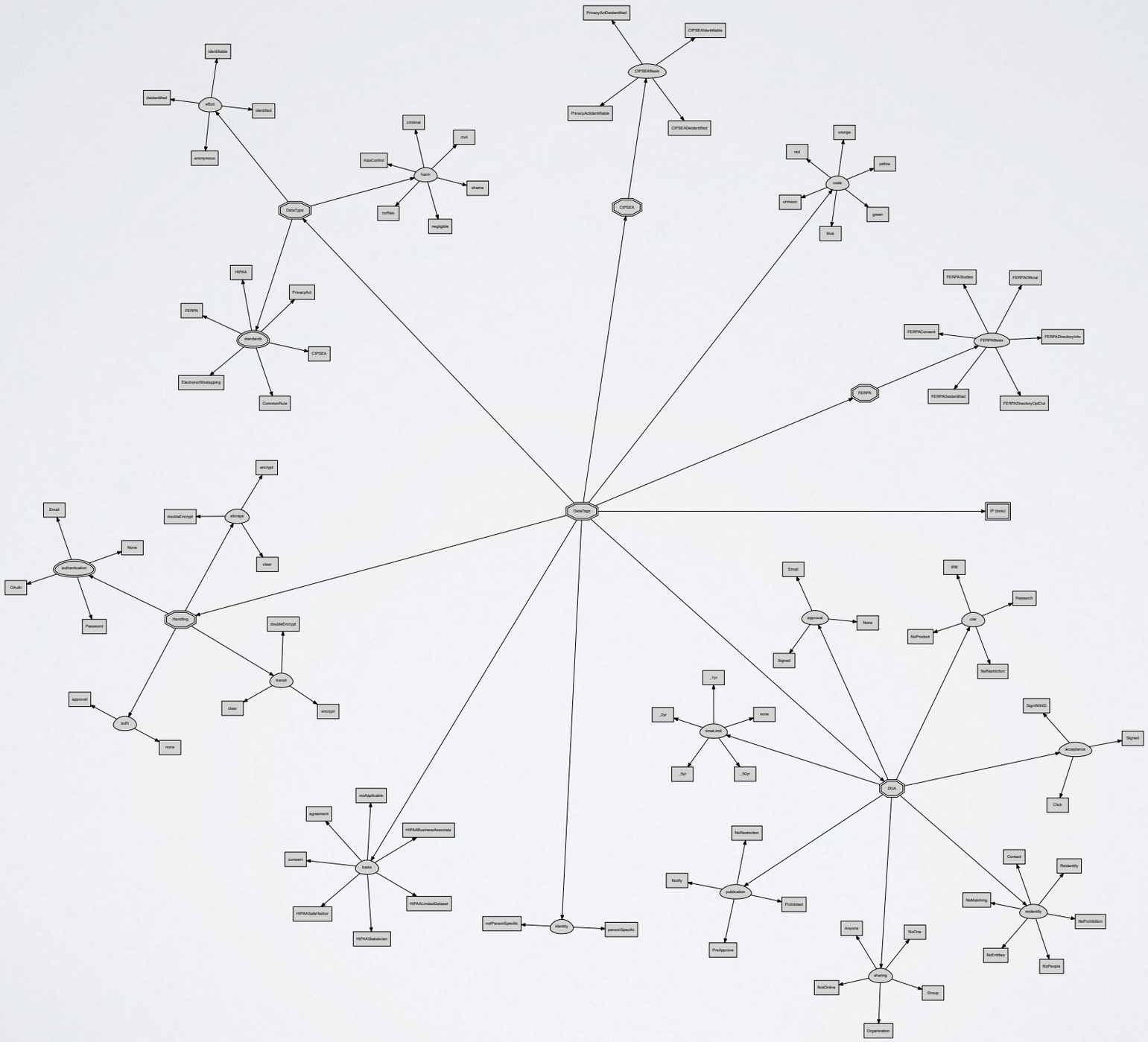
Tools



- Editing: Any text editor
- Compiler
- Visualizers
  Using Graphviz
- Runtime Engine
  Java library
- Command-line Runner

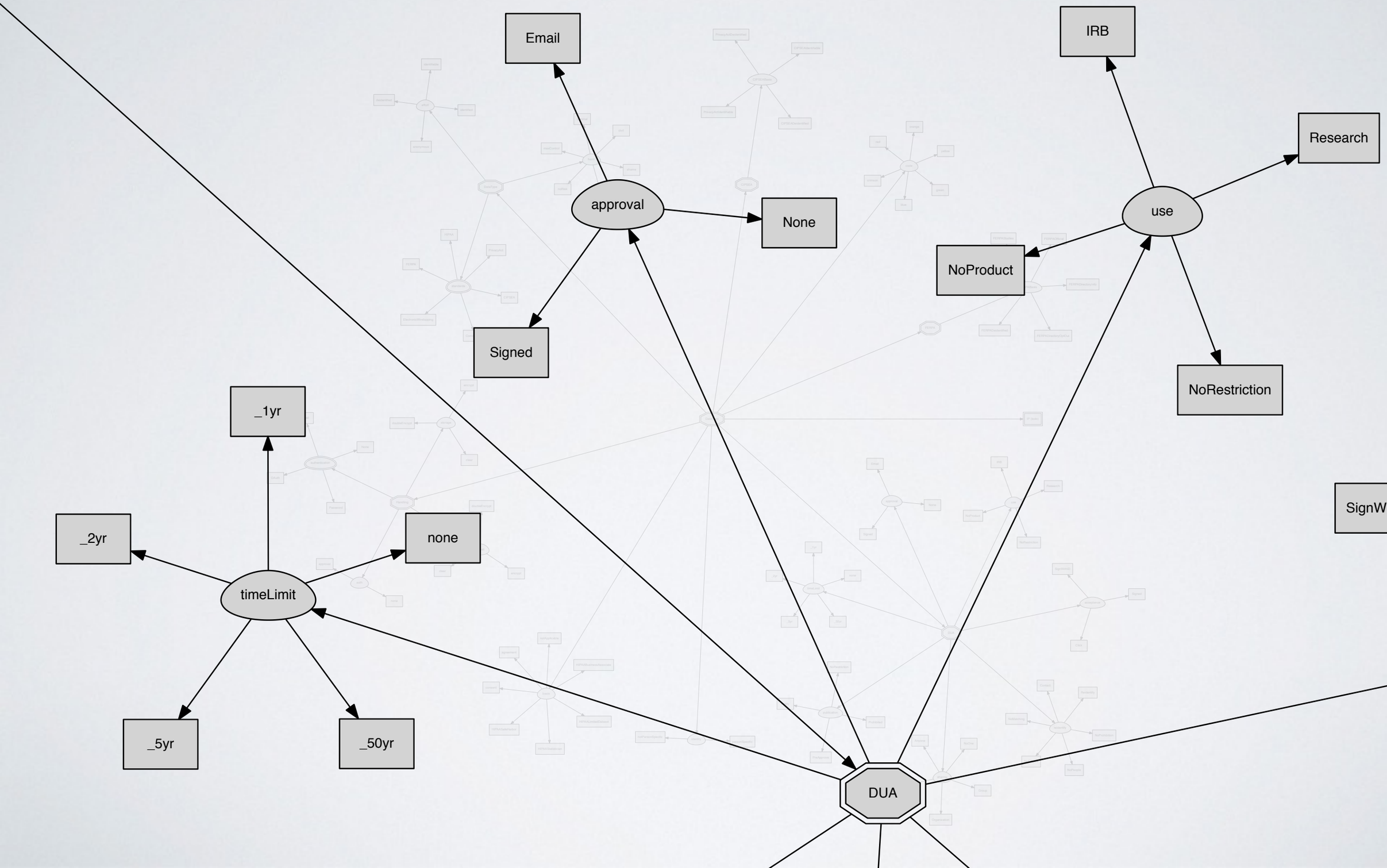
Tools

Visualizers: dtDef



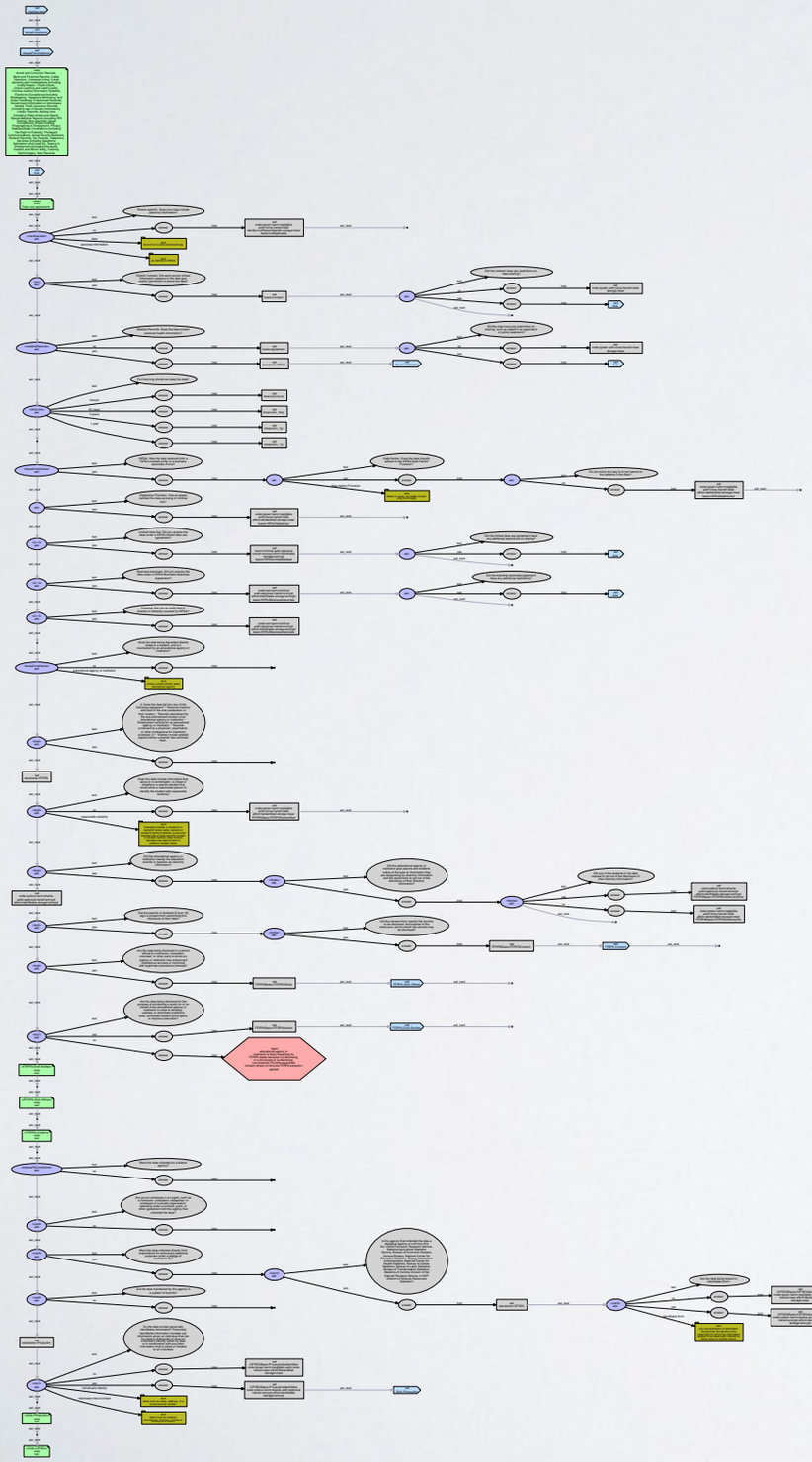
Tools

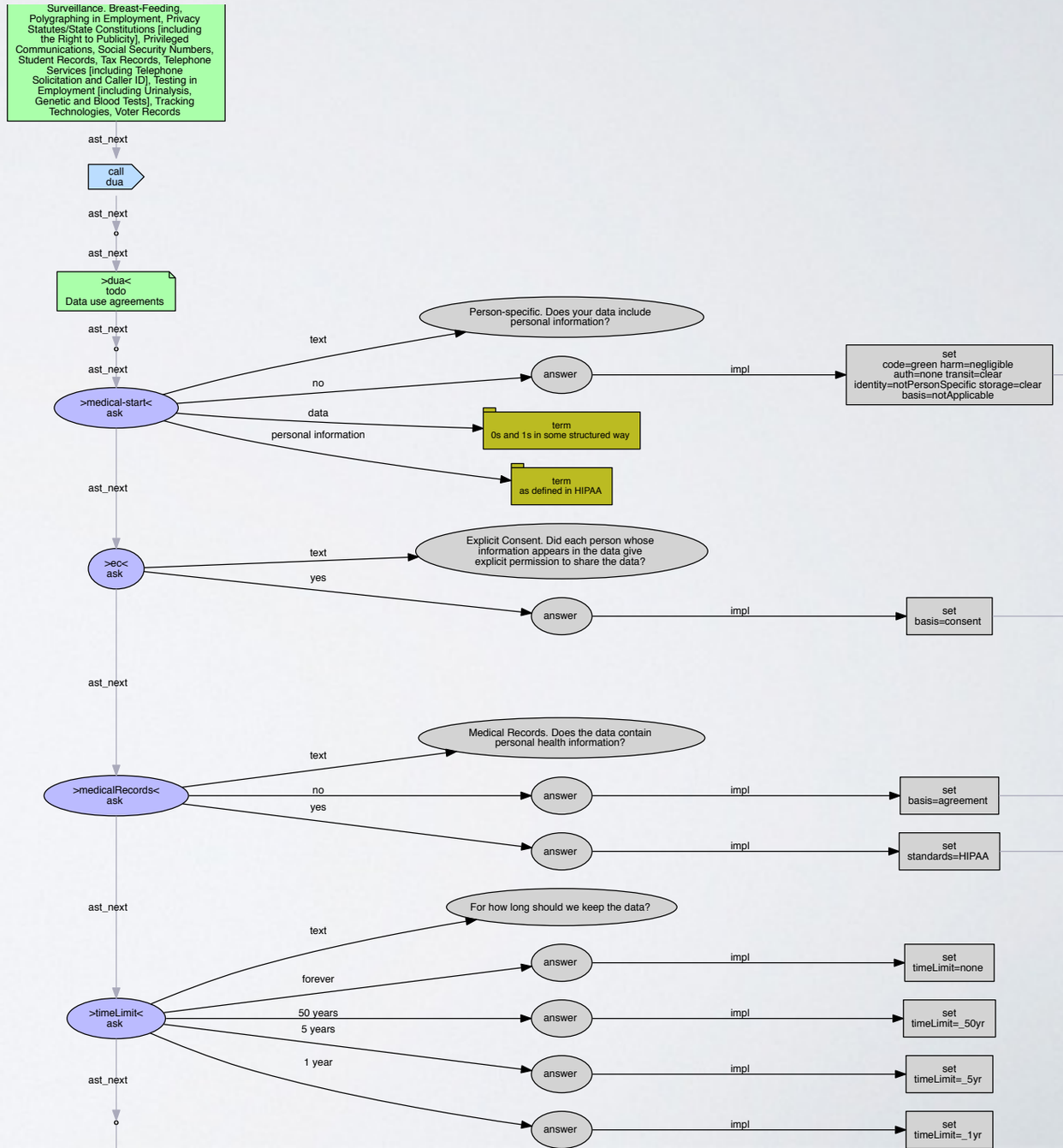
Visualizers: dtDef



Tools

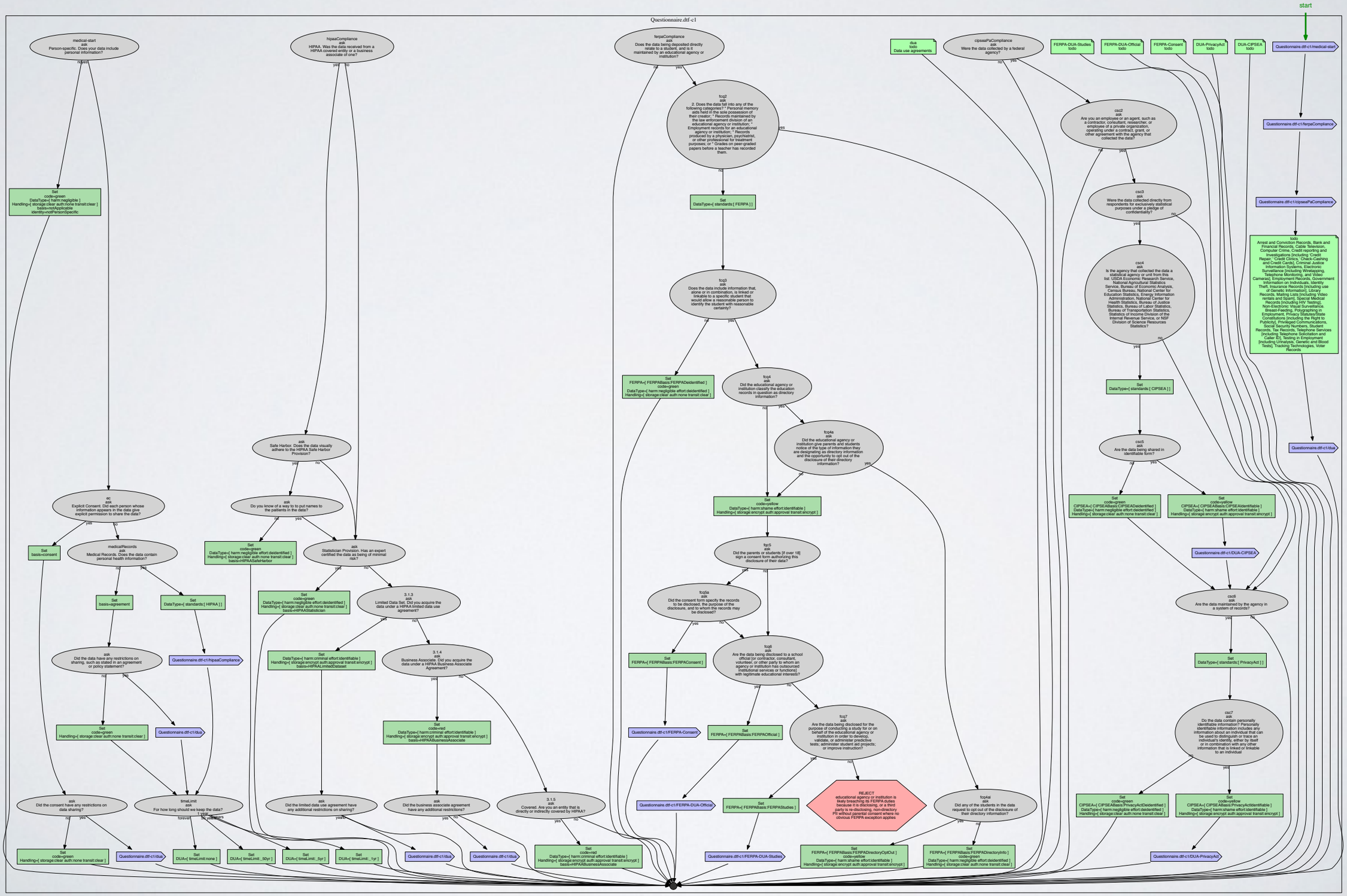
Visualizers: dtf/ast





Tools

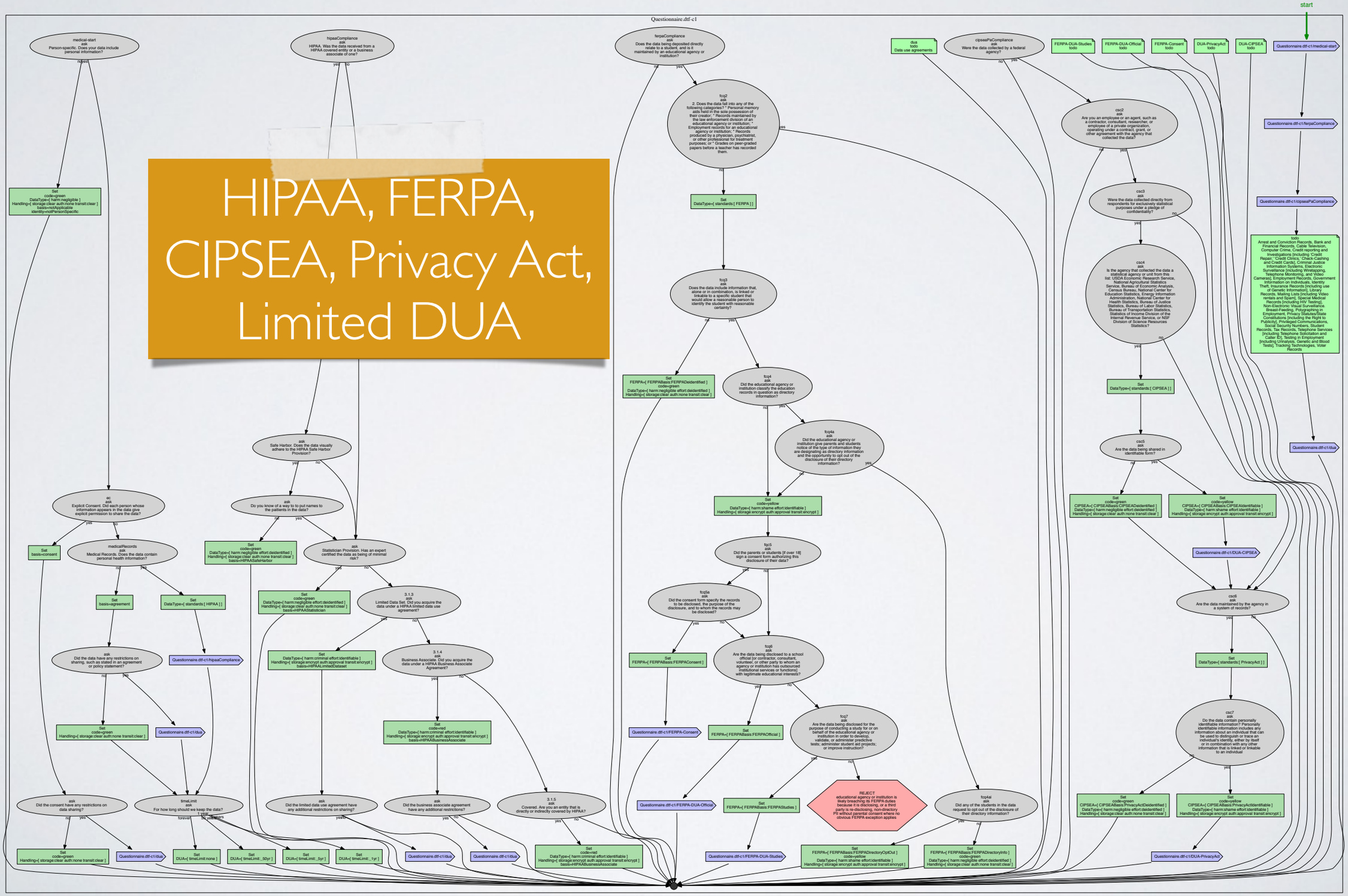
Visualizers: dtf/Chart



Tools

Visualizers: dtf/Chart

HIPAA, FERPA, CIPSEA, Privacy Act, Limited DUA



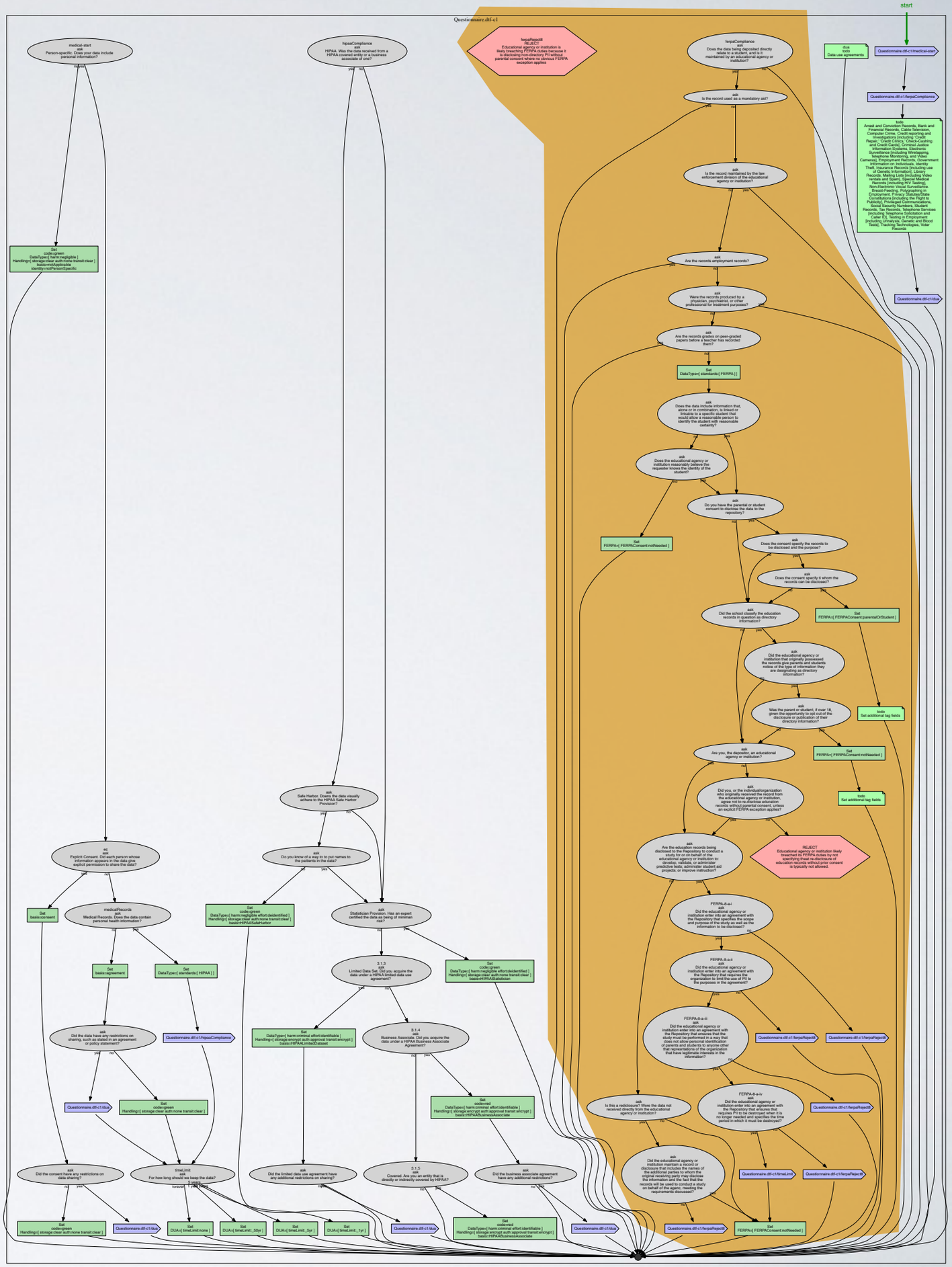
Tools

Using the language and
tool to collaborate

#TrueStory

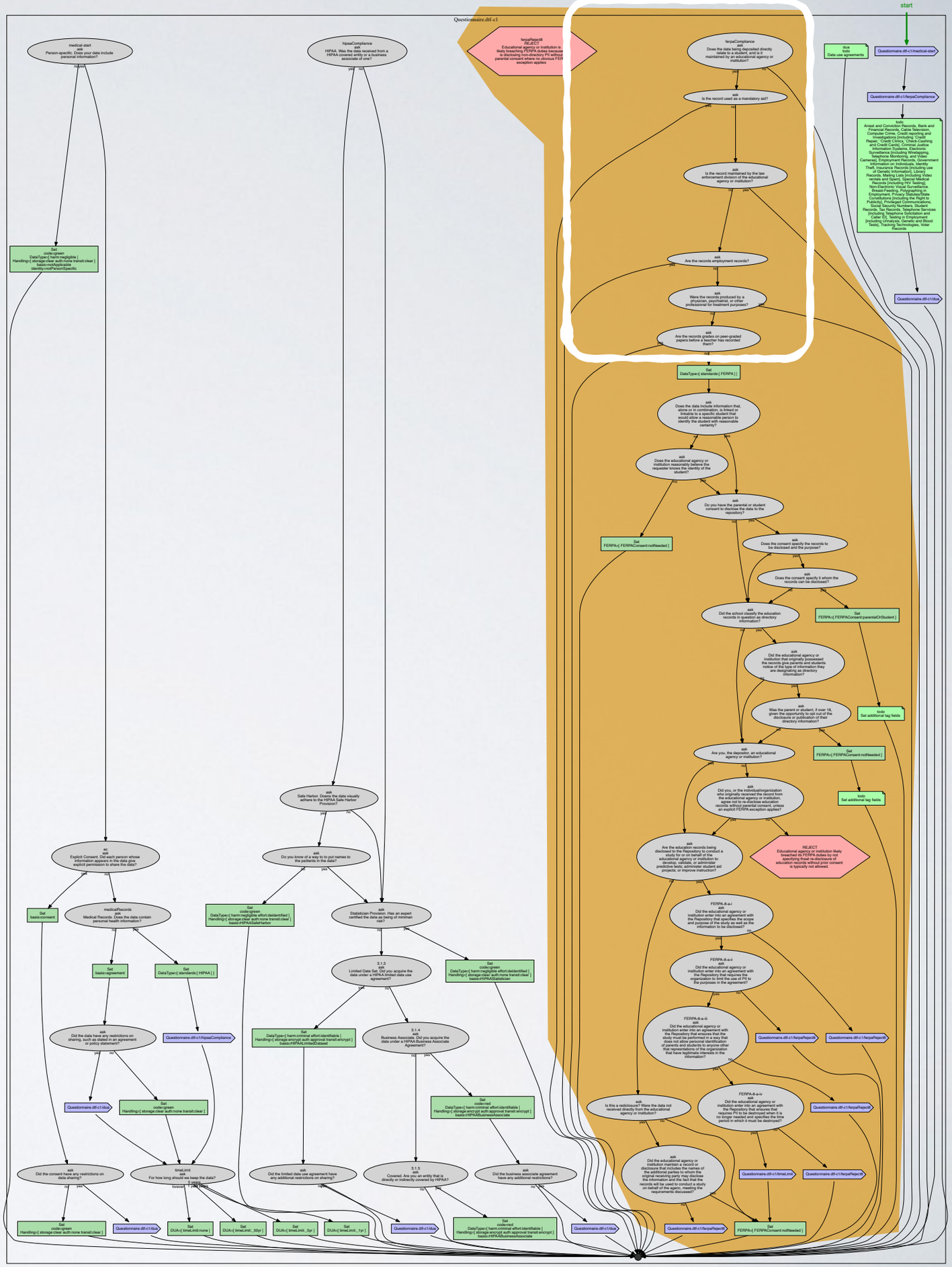
FERPA Interview

Initial FERPA interview, based on Berkman's summer interns work



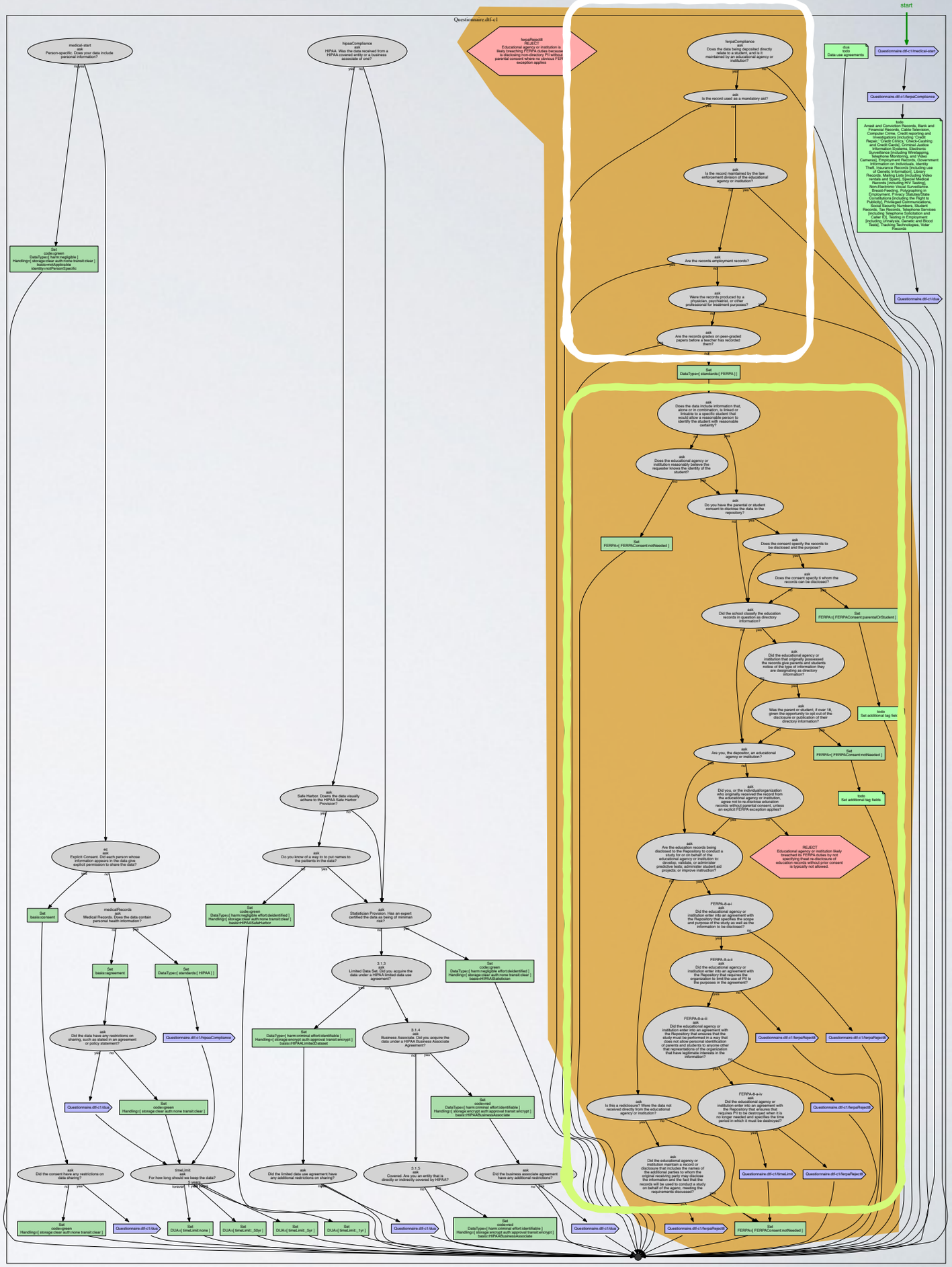
FERPA Interview

Initial FERPA interview, based on Berkman's summer interns work

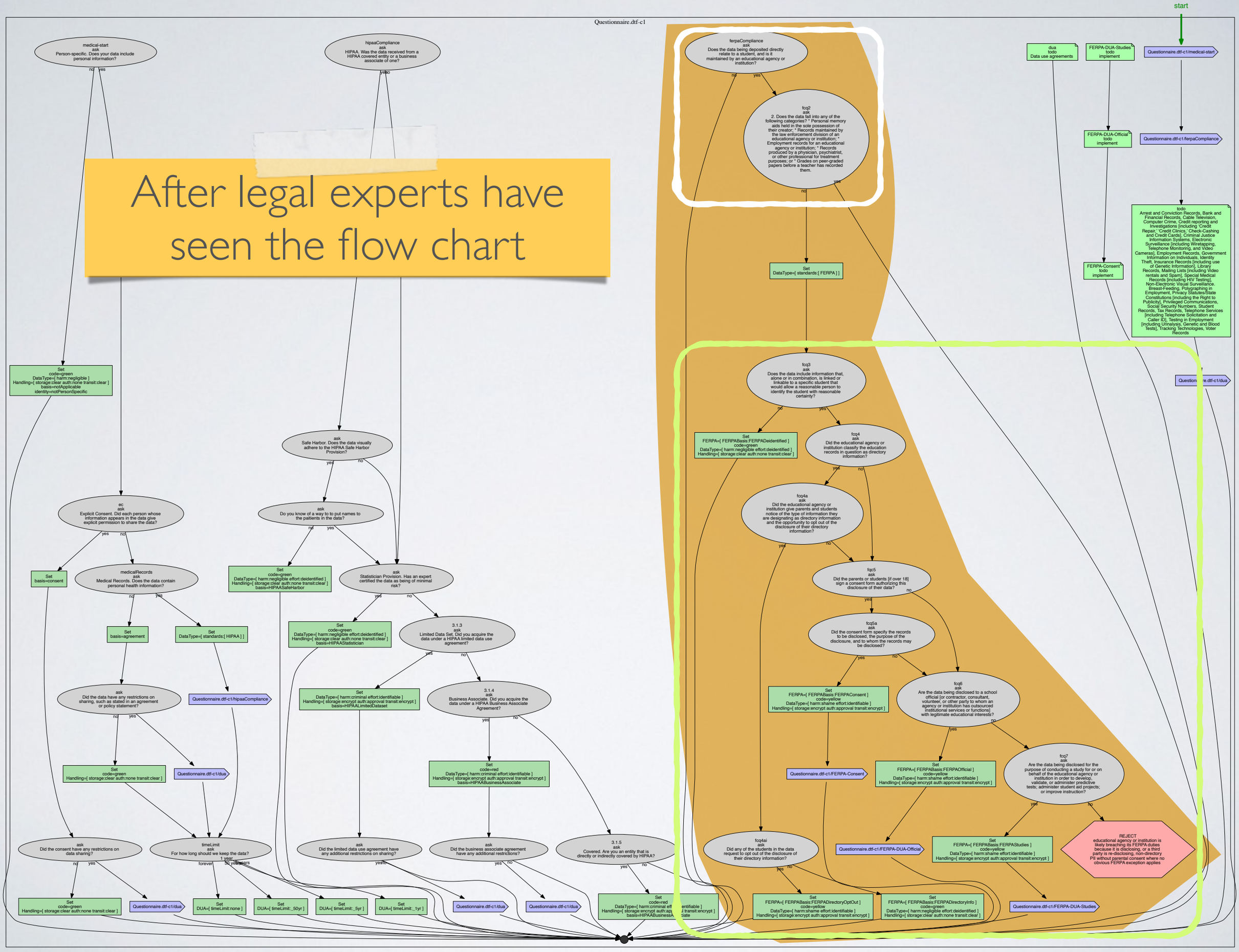


FERPA Interview

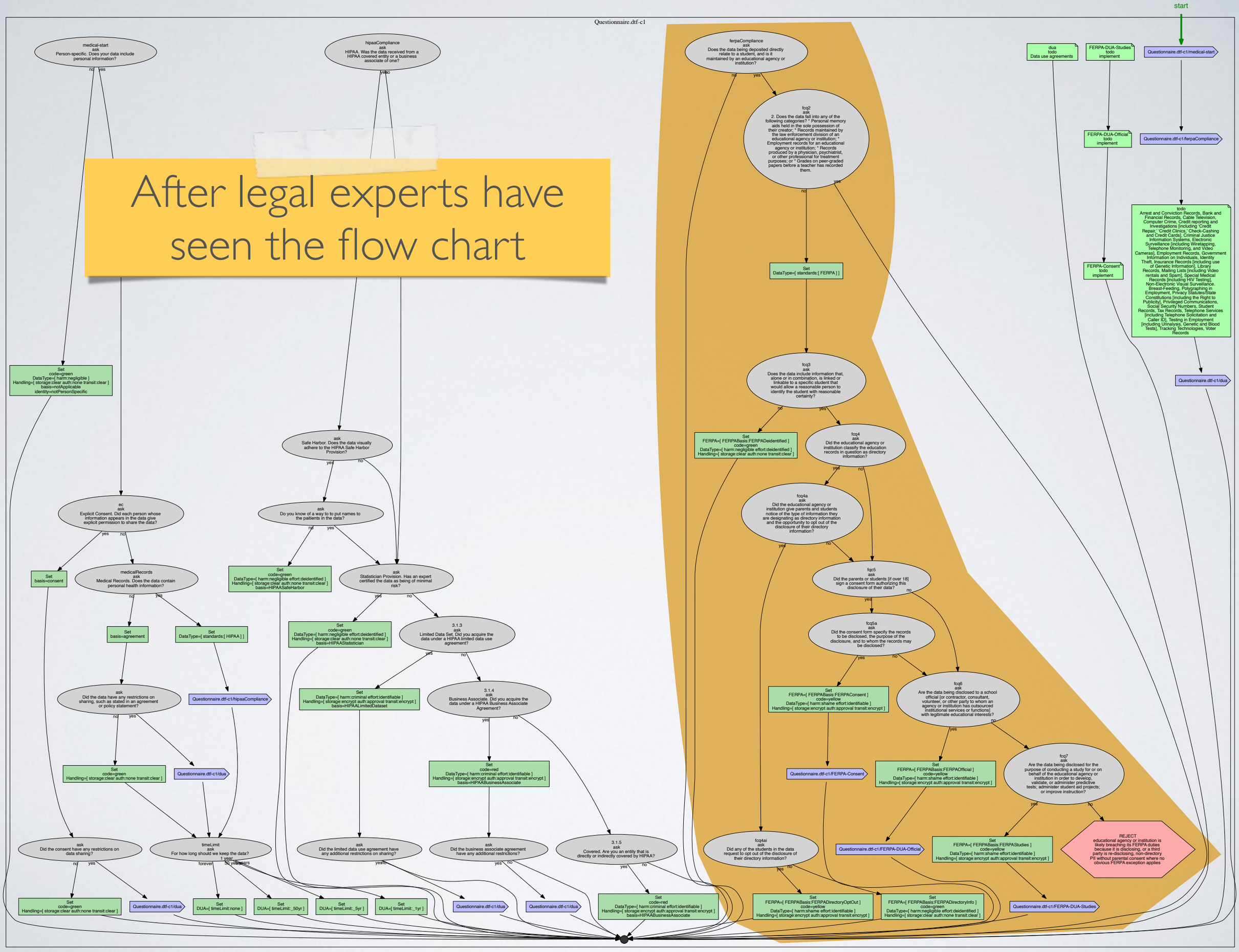
Initial FERPA interview, based on Berkman's summer interns work

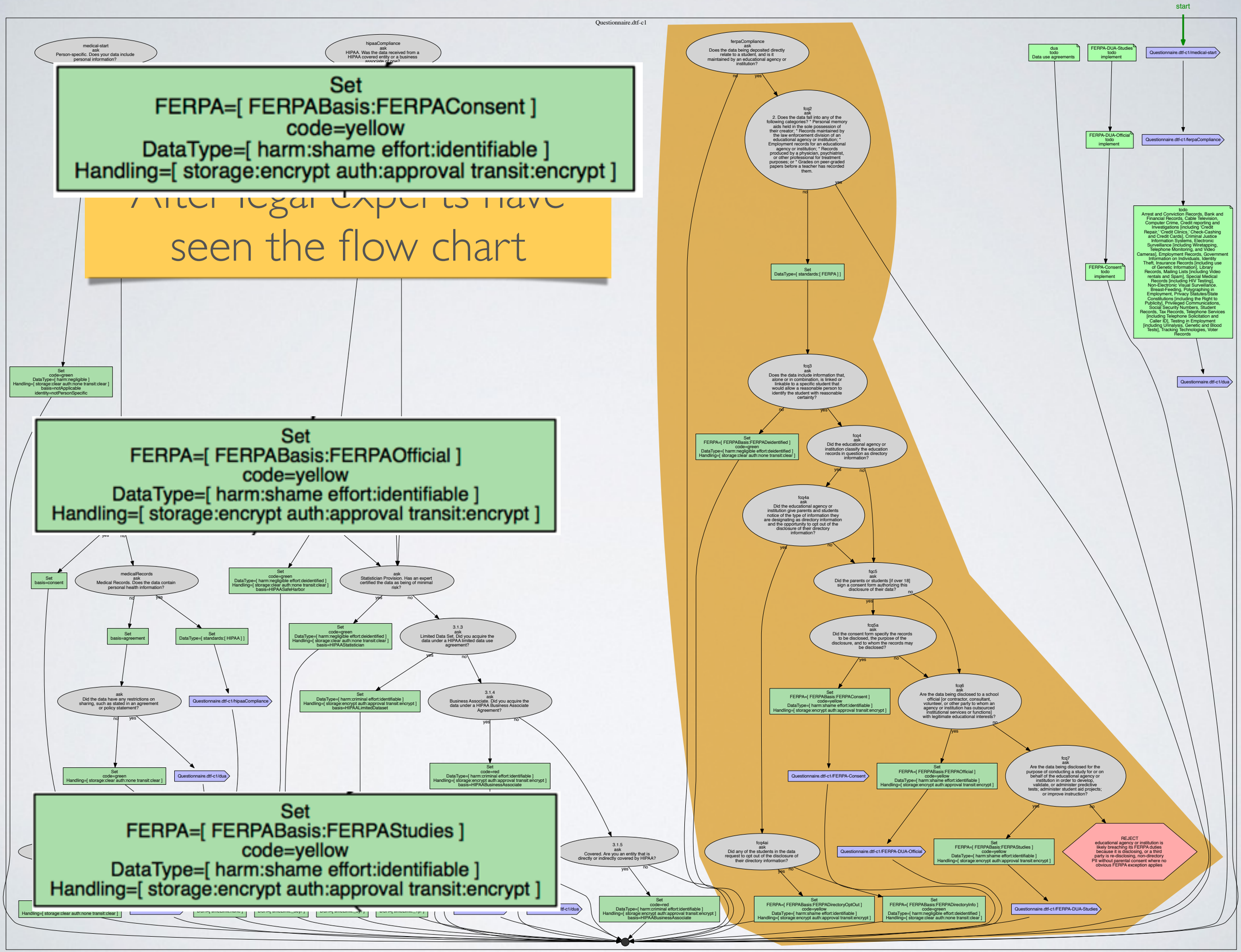


After legal experts have seen the flow chart

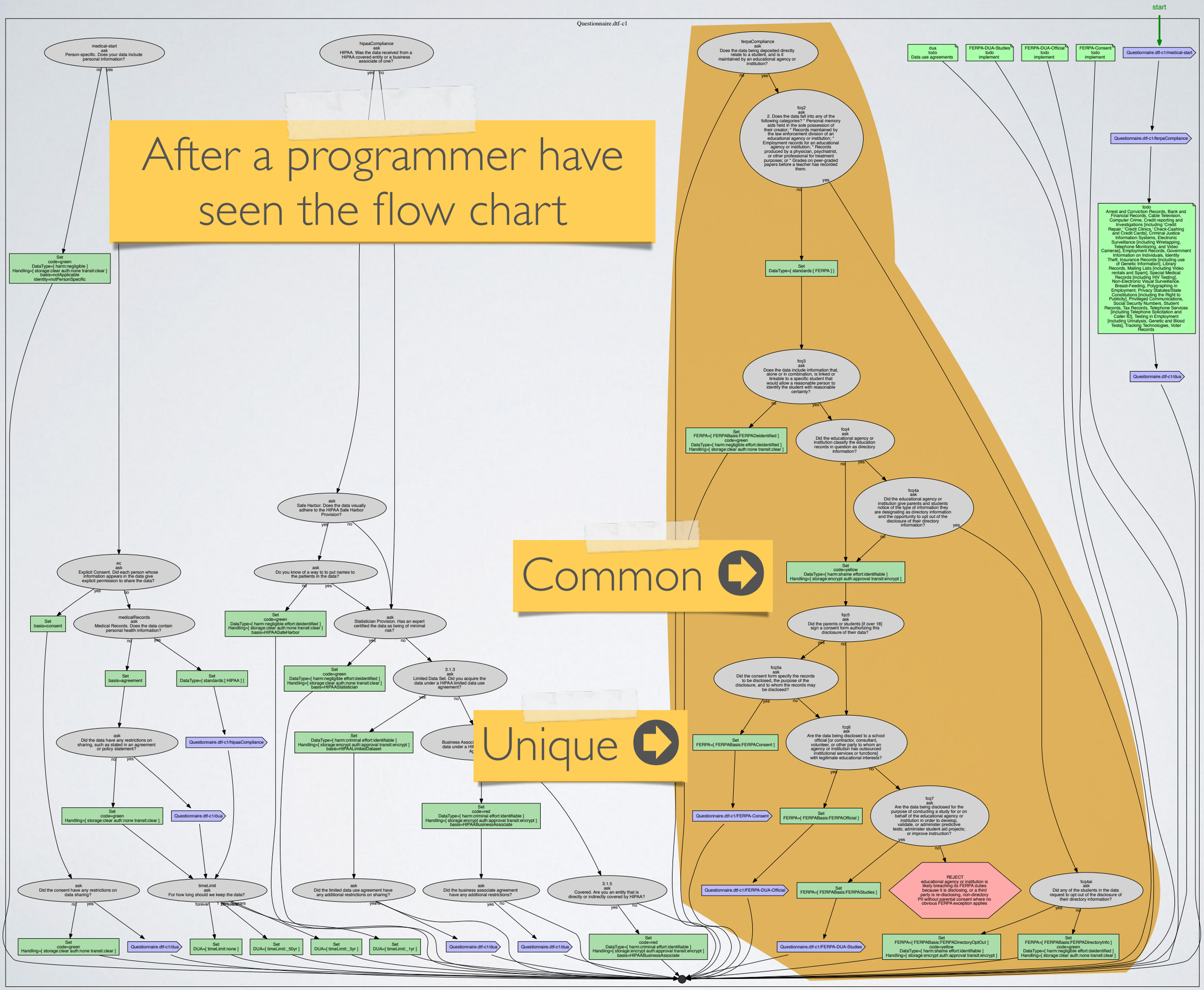


After legal experts have seen the flow chart

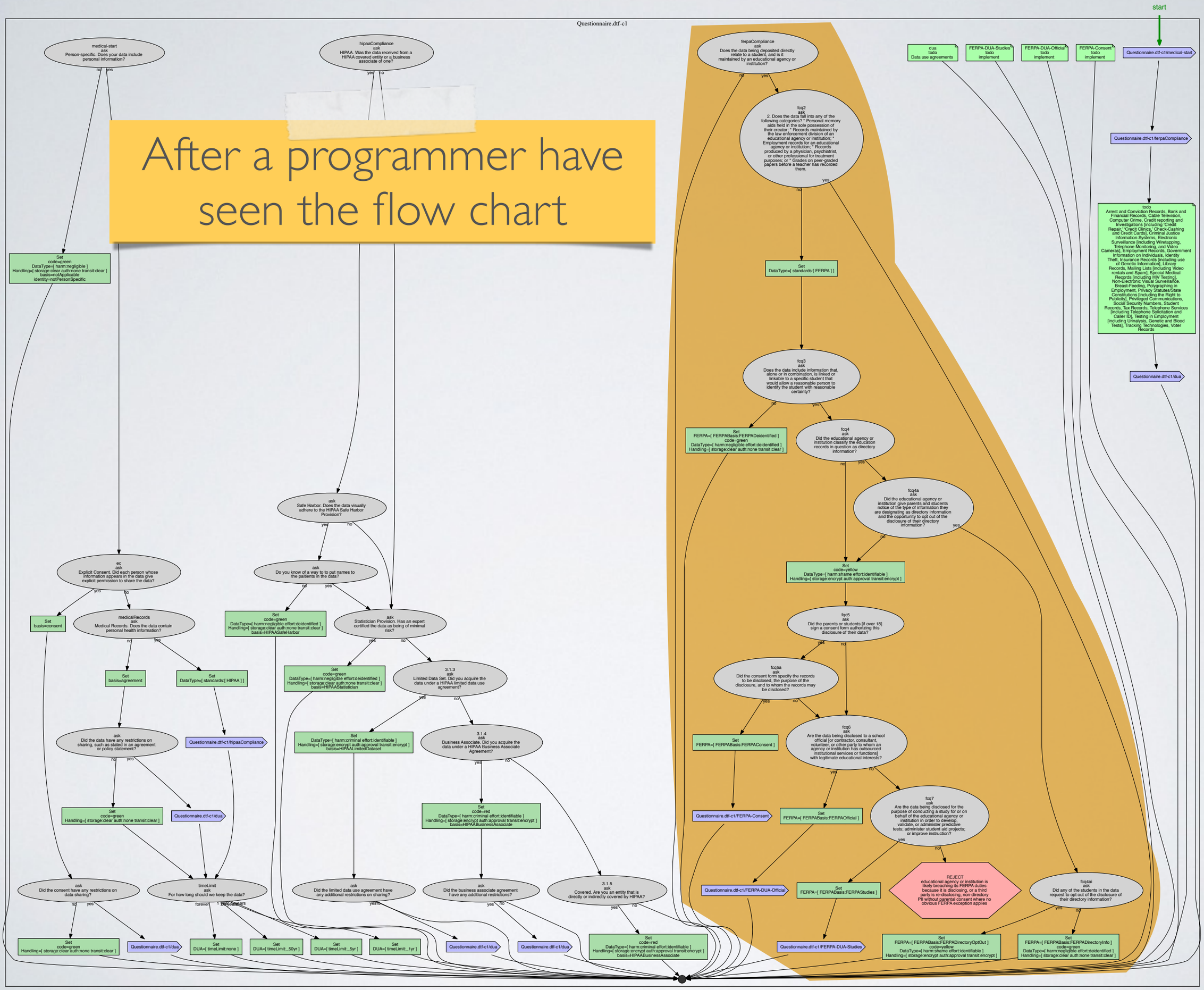



Questionnaire.dtf-c1
start
medical-start
ask
Person-specific. Does your data include personal information?
hipaaCompliance
ask
HIPAA. Was the data received from a HIPAA covered entity or a business associate of one?
Set
FERPA=[FERPABasis:FERPAConsent]
code=yellow
DataType=[harm:shame effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
After legal experts have seen the flow chart
Set
code=green
DataType=[harm:negligible]
Handling=[storage:clear auth:none transit:clear]
basis=notApplicable
identity=notPersonSpecific
Set
FERPA=[FERPABasis:FERPAOfficial]
code=yellow
DataType=[harm:shame effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
Set
basis=consent
medicalRecords
ask
Medical Records. Does the data contain personal health information?
Set
code=green
DataType=[harm:negligible effort:deidentified]
Handling=[storage:clear auth:none transit:clear]
basis=HIPAASafeHarbor
ask
Statistician Provision. Has an expert certified the data as being of minimal risk?
Set
basis=agreement
Set
DataType=[standards:[HIPAA]]
Set
code=green
DataType=[harm:negligible effort:deidentified]
Handling=[storage:clear auth:none transit:clear]
basis=HIPAAStatistician
3.1.3
ask
Limited Data Set. Did you acquire the data under a HIPAA limited data use agreement?
ask
Did the data have any restrictions on sharing, such as stated in an agreement or policy statement?
Questionnaire.dtf-c1/hipaaCompliance
Set
DataType=[harm:criminal effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
basis=HIPAALimitedDataset
3.1.4
ask
Business Associate. Did you acquire the data under a HIPAA Business Associate Agreement?
Set
code=green
Handling=[storage:clear auth:none transit:clear]
Questionnaire.dtf-c1/dua
Set
code=red
DataType=[harm:criminal effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
basis=HIPAABusinessAssociate
Set
FERPA=[FERPABasis:FERPAStudies]
code=yellow
DataType=[harm:shame effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
3.1.5
ask
Covered. Are you an entity that is directly or indirectly covered by HIPAA?
Handling=[storage:clear auth:none transit:clear]
Set
code=red
DataType=[harm:criminal effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
basis=HIPAABusinessAssociate
ferpaCompliance
ask
Does the data being deposited directly relate to a student, and is it maintained by an educational agency or institution?
fcq2
ask
2. Does the data fall into any of the following categories? * Personal memory aids held in the sole possession of their creator; * Records maintained by the law enforcement division of an educational agency or institution; * Employment records for an educational agency or institution; * Records produced by a physician, psychiatrist, or other professional for treatment purposes; or * Grades on peer-graded papers before a teacher has recorded them.
Set
DataType=[standards:[FERPA]]
fcq3
ask
Does the data include information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person to identify the student with reasonable certainty?
Set
FERPA=[FERPABasis:FERPADeidentified]
code=green
DataType=[harm:negligible effort:deidentified]
Handling=[storage:clear auth:none transit:clear]
fcq4
ask
Did the educational agency or institution classify the education records in question as directory information?
fcq4a
ask
Did the educational agency or institution give parents and students notice of the type of information they are designating as directory information and the opportunity to opt out of the disclosure of their directory information?
fcq5
ask
Did the parents or students [if over 18] sign a consent form authorizing this disclosure of their data?
fcq5a
ask
Did the consent form specify the records to be disclosed, the purpose of the disclosure, and to whom the records may be disclosed?
Set
FERPA=[FERPABasis:FERPAConsent]
code=yellow
DataType=[harm:shame effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
fcq6
ask
Are the data being disclosed to a school official [or contractor, consultant, volunteer, or other party to whom an agency or institution has outsourced institutional services or functions] with legitimate educational interests?
Questionnaire.dtf-c1/FERPA-Consent
Set
FERPA=[FERPABasis:FERPAOfficial]
code=yellow
DataType=[harm:shame effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
fcq7
ask
Are the data being disclosed for the purpose of conducting a study for or on behalf of the educational agency or institution in order to develop, validate, or administer predictive tests; administer student aid projects; or improve instruction?
fcq4ai
ask
Did any of the students in the data request to opt out of the disclosure of their directory information?
Questionnaire.dtf-c1/FERPA-DUA-Official
Set
FERPA=[FERPABasis:FERPAStudies]
code=yellow
DataType=[harm:shame effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
REJECT
educational agency or institution is likely breaching its FERPA duties because it is disclosing, or a third party is re-disclosing, non-directory PII without parental consent where no obvious FERPA exception applies
Set
FERPA=[FERPABasis:FERPADirectoryOptOut]
code=yellow
DataType=[harm:shame effort:identifiable]
Handling=[storage:encrypt auth:approval transit:encrypt]
Set
FERPA=[FERPABasis:FERPADirectoryInfo]
code=green
DataType=[harm:negligible effort:deidentified]
Handling=[storage:clear auth:none transit:clear]
Questionnaire.dtf-c1/FERPA-DUA-Studies
dua
todo
Data use agreements
FERPA-DUA-Studies
todo
implement
FERPA-DUA-Official
todo
implement
FERPA-Consent
todo
implement
Questionnaire.dtf-c1/medical-start
Questionnaire.dtf-c1/ferpaCompliance
todo
Arrest and Conviction Records, Bank and Financial Records, Cable Television, Computer Crime, Credit reporting and Investigations [including 'Credit Repair,' 'Credit Clinics,' Check-Cashing and Credit Cards], Criminal Justice Information Systems, Electronic Surveillance [including Wiretapping, Telephone Monitoring, and Video Cameras], Employment Records, Government Information on Individuals, Identity Theft, Insurance Records [including use of Genetic Information], Library Records, Mailing Lists [including Video rentals and Spam], Special Medical Records [including HIV Testing], Non-Electronic Visual Surveillance. Breast-Feeding, Polygraphing in Employment, Privacy Statutes/State Constitutions [including the Right to Publicity], Privileged Communications, Social Security Numbers, Student Records, Tax Records, Telephone Services [including Telephone Solicitation and Caller ID], Testing in Employment [including Urinalysis, Genetic and Blood Tests], Tracking Technologies, Voter Records
Questionnaire.dtf-c1/dua

After a programmer have seen the flow chart

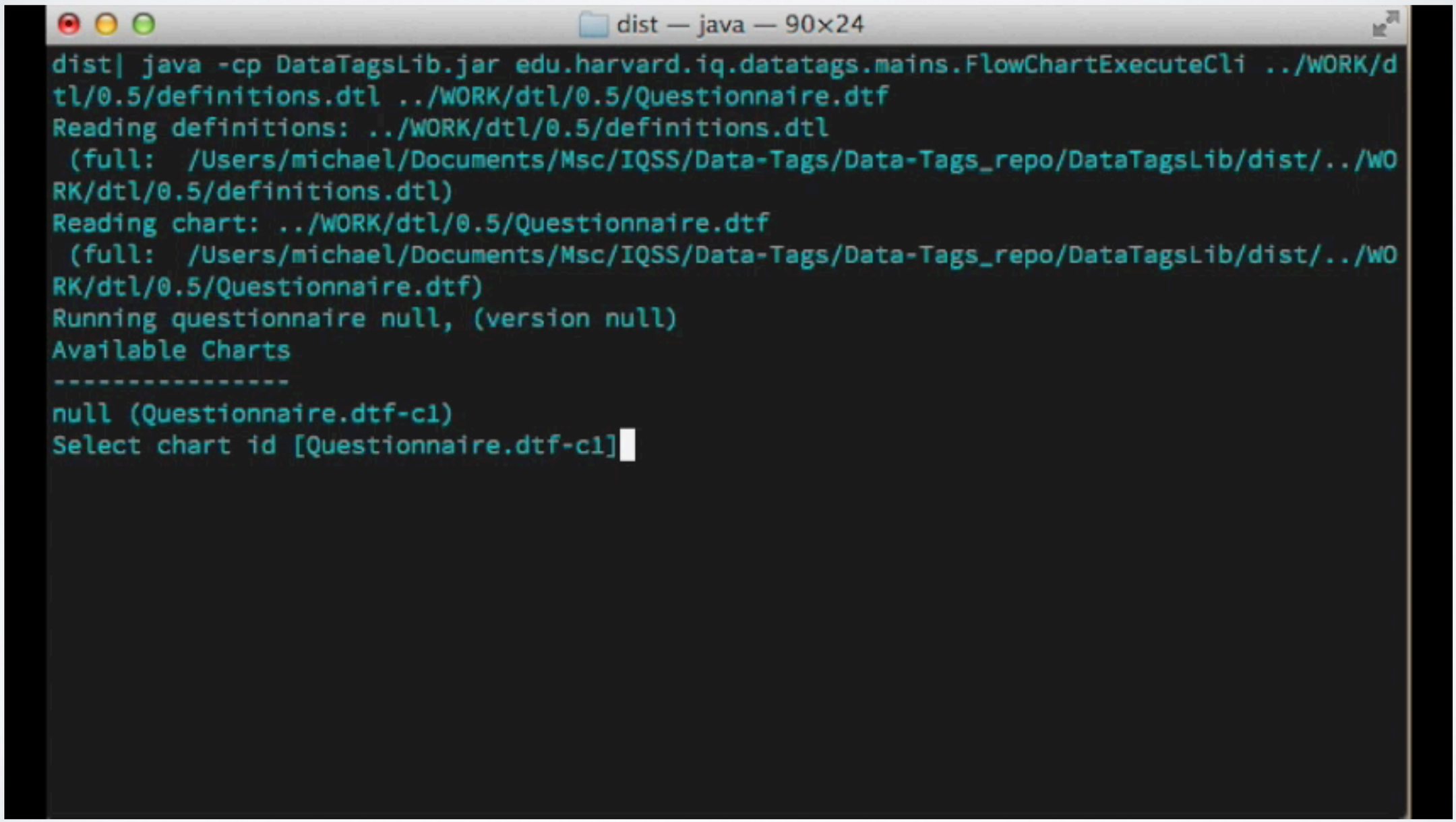


After a programmer have seen the flow chart



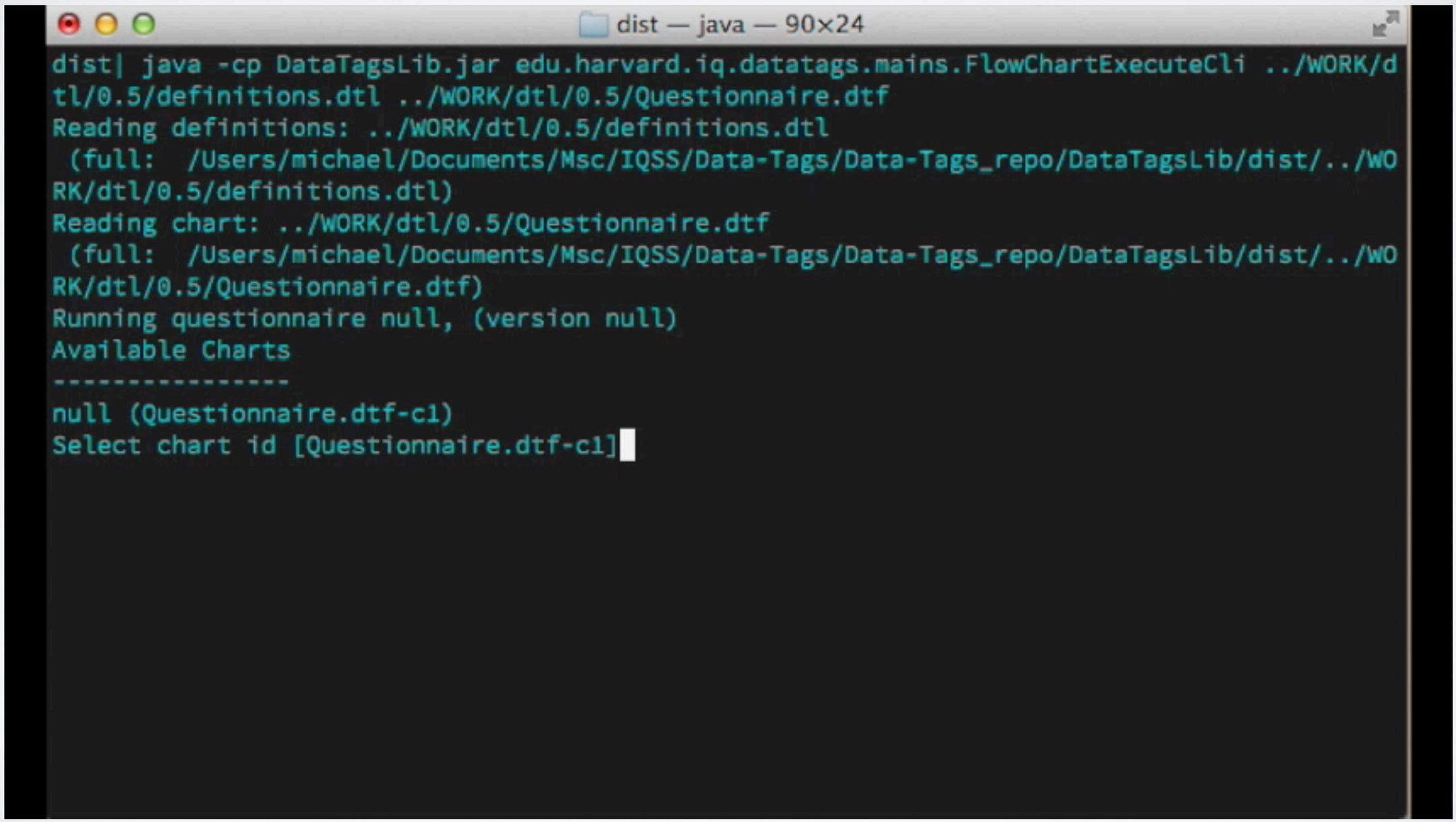
Tools

CLI Runner



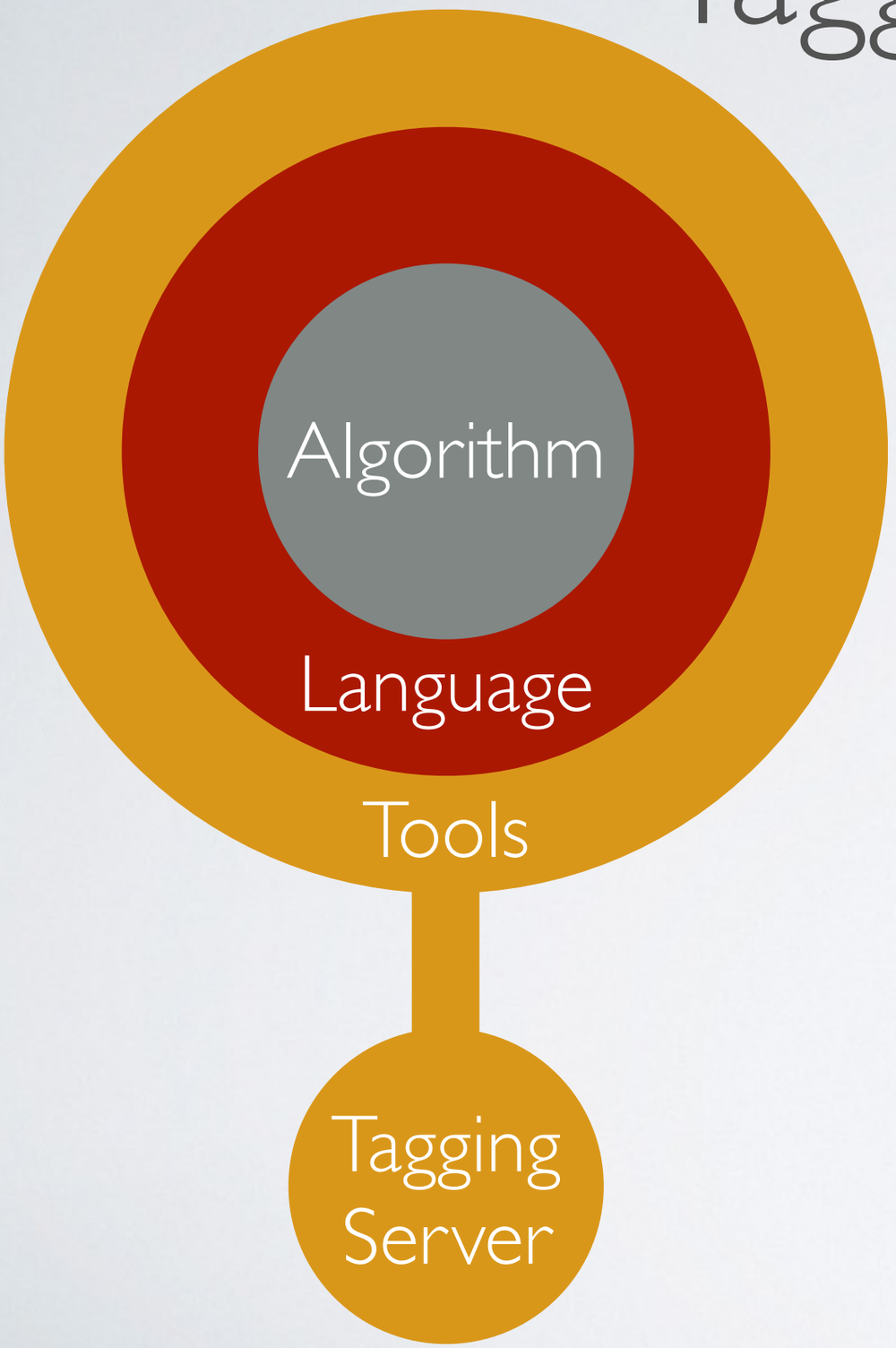
Tools

CLI Runner



components

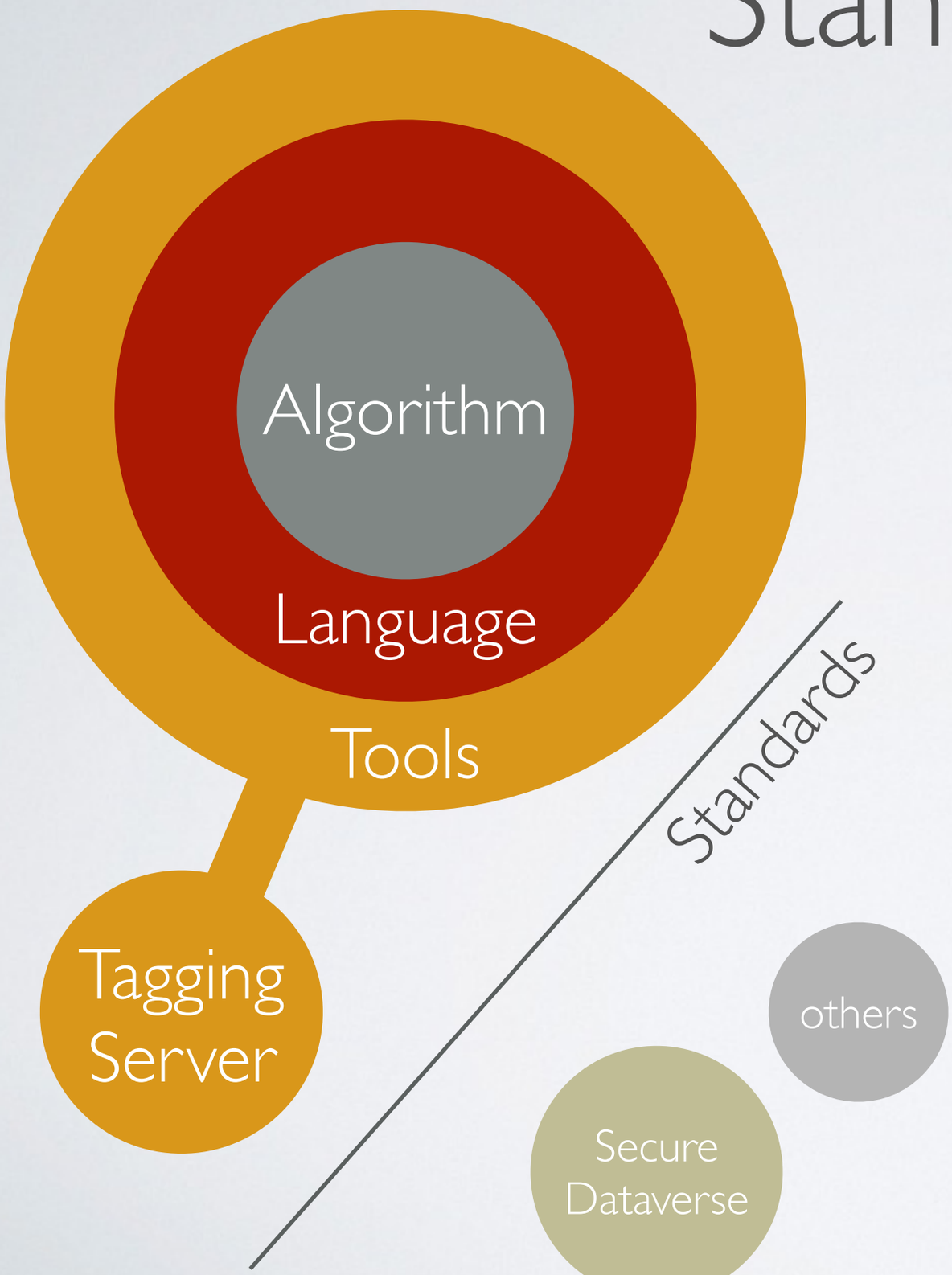
Tagging Server



- Current IQSS focus
- Web UI
  - Focus on usability
- API, Integration with other systems
  NSF requirement
- Will allow other teams to develop tagging interviews

components

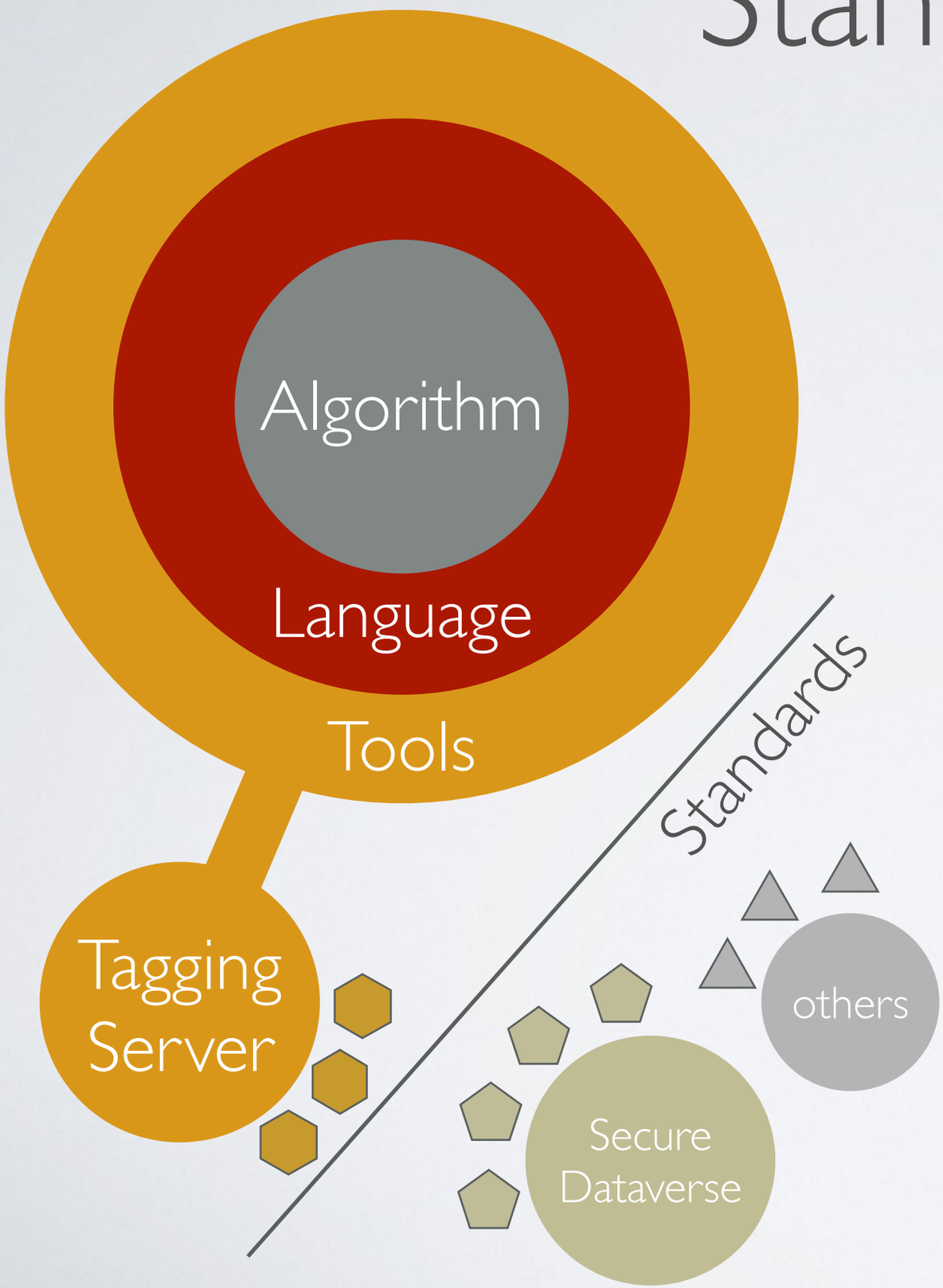
Standards



- Repositories need to implement behaviors according to the tags
- Software, Server and site security needed for the DataTags and the Repository
- Certification and auditing for software systems and instances

components

Standards



- Repositories need to implement behaviors according to the tags
- Software, Server and site security needed for the DataTags and the Repository
- Certification and auditing for software systems and instances

Questions?

What we talk about when we talk about

DataTags

Thanks

IQSS

Data Privacy Lab

Dataverse Network PROJECT

Privacy Tools
for Sharing Research Data