Project Overview:

Prototype

System for working through DCF process and producing reusable artifacts

- **Templates**: Reusable and configurable information about policy, data, and controls used across many data sharing efforts

- **Web App**: Guides users through DCF process using library of available templates

- **Environments**: Virtual machines configured based on user responses and template options
Identify Utility & Risks

**Utility**
Specify how the data is intended to be used

**Risk**
Determine risks associated with the stated utility goals and their required data

Apply Disclosure Controls

**Operational Controls**
Changes to the data environment to limit the scope of available data

**Data Controls**
Alter data to remove or obfuscate the presence of sensitive information

**Policy Controls**
Legal restrictions on the use of the data to discourage misbehavior by recipient

Assess Impact

**Utility Impact**
Revisit goals and determine how disclosure controls may have changed them

**Risk Impact**
Revisit risks based on changes to available data and utility goals due to controls
Policy Template

“name”: “Privacy Notice”,
“description”: “A privacy notice and terms of use.”,
“categories”: [“upstream”, “terms”, “covenants”],
“text”: “We collect [#(COLLECTED_DATA)] kinds of information to measure the performance of your mobile broadband service.

[COLLECTED_DATA]
This data is protected using [PROTECTIONS]. You can find more detail in the FCC’s technical summary of this program.”,
“questions”: [
    {“question”: “Enumerate data items collected.”, “answer”: “COLLECTED_DATA”},
    {“question”: “Enumerate protections for raw data after collection.”, “answer”: “PROTECTIONS”}
]

Technology Template

“name”: “Quantize_Location”,
“description”: “Aggregate location data into blocks.”,
“categories”: [“aggregation”, “location_data”],
“pointer”: “http://example.com/quantize_loc”
“parameters”: {
    “k”: “10”,
    “granularity”: “0.5”
}

Data Template

“name”: “GPS Data”,
“description”: “Lat. and long. data”,
“categories”: [“location_data”],
“parser”: “http://example.com/gps”
“schema”: “
{
    “accuracy”: “float”,
    “latitude”: “float”,
    “longitude”: “float”,
    “timestamp”: “datetime”
}”
Policy Template: Acceptable Use Agreement

“name”: “Acceptable Use Agreement (CAIDA)"
“description”: “A data use agreement that contains the data license, ...”
“categories”: [“use restrictions”, “data license”, ...]
“text”: “The following terms comprise the acceptable use agreement for all datasets made available to [RECIPIENT_NAME] by the [PROVIDER_NAME]. Certain datasets may have additional Supplemental provisions.

[DATAPROVIDER_NAME] has the authority and reserves the right, in its sole discretion, to refuse requests for dataset(s) or discontinue further access and use to anyone. If [RECIPIENT_NAME] feel [RECIPIENT_NAME]'s request is inappropriately denied please contact [PROVIDER_NAME] by sending a message to [PROVIDER_EMAIL].”

...”
“questions”:

[“provider_name”: “State data provider’s name.”,
 “provider_email”: “Enter data provider’s email.”,
 “recipient_name”: “State data recipient’s name.”,
 “recipient_email”: “Enter data recipient’s email.”,
 “data_type”: “Enumerate data types in this data release.”]
Data Template: DNS Message

"name": "dns_message"
"description": "Combined format for DNS query and reply messages"
"categories": ["enterprise", "domain"]
"parser": "http://redjack.com/dns_parser"
"schema":
{
    "name": "header",
    "type": {
        "name": "dns_header",
        "type": "record",
        "fields": [
            { "name": "id", "type": "int" },
            { "name": "flags1", "type": "int" },
            { "name": "flags2", "type": "int" },
            { "name": "qdcount", "type": "int" },
            { "name": "ancount", "type": "int" },
            { "name": "nscount", "type": "int" },
            { "name": "arcount", "type": "int" }
        ]
    }
}, ...

Current Progress: Prototype

- Templates: 50%
  - Data (DNS, Flow)
  - Policy (AUA)
  - Tech
  - DCF Questionnaire
  - VM Provisioning

- Web App: 0%
  - Collection
  - Sharing

- Environments: 0%
Promoting DCF use by the larger data sharing community and soliciting feedback

- **FCC Mobile Broadband**: Provide privacy guidance and case study for use of DCF in large-scale project; possible PREDICT dataset

- **Information Security Operations & Intelligence (ISOI13)**: "Less is More – A Disclosure Control Approach to Data Sharing"

- **International Association of Privacy Professionals (IAPP) Europe Data Protection Congress**, November, Brussels.
FCC Mapping Data

Collection of mobile broadband **speed** and **signal strength** indicators via phone apps

Recommendations after initial analysis:

- Quantize location information to 25km hexagon
- Min. measurements per carrier in each block *(k-anonymity)*
- Min. “landmarks” per block *(l-diversity)*
- Separate handset stats from location measurements
FCC Mapping Data
Opportunities and Challenges

Data ready to be released to PREDICT pending **final privacy analysis** and approval by **mobile carriers**

Strong desire for “centralized” FCC data repository and **DCF-like web app**

Interest in research-oriented data release via **secure virtual enclave**
Current Progress: Outreach

- **Privacy Analysis**: 66%
- **Case Study**: 50%
- **Data Import**: 25%

- Initial analysis
- Recommendations
- Final analysis
- Submission to WISCS (colocated w/ CCS)
- Mapping Data
- Research Data