**Data Types**

The main products of this proposal are a plan for structuring and storing historical data, and a Resource Description Framework (RDF) model that can be used to make those data available via Linked Data. Those products will be comprised of the following object types:

* Electronic documents containing minutes, procedings, and working materials generated during project related activities (e.g. minutes from consultation workshops);
* Electronic documents containing expository prose about the plan for structuring and storing historical data;
* Images and graphs depicting data models, infrastructural models, and workflows;
* Code and pseudocode related to the implementation of the plan and the RDF model;
* Samples of digitized historical data.

All of those objects will be stored in an open-access Github repository (see Preservation), along with descriptive metadata.

**Data Standards**

Metadata describing the objects above will comply with the Dublin Core metadata standard. Metadata records will be implemented in XML and stored along with the objects that they describe in the project’s GitHub repository (see Preservation).

Code and pseudocode will be documented using ReST-formatted docstrings and in-code comments, and fully comply with the PEP 8 Style Guide for Python Code in order to maximize clarity and reusability. Where appropriate, documentation will be transformed to HTML and/or PDF formats for dissemination.

**Preservation**

The data objects described in Data Types, as well as the metadata described in Data Standards, will be stored in a public Github repository (<https://github.com/erickpeirson/Materiality-of-Risk>). GitHub is an open source repository where project participants can store revisions of data objects, including source code, documents, and other electronic materials. We will take avantage of the tools that GitHub provides for version control and annotation. This will ensure fidelity and persistence of project data, appropriate attribution, and the ability to inspect the history of specific data objects over the course of the project. **Deposition of material into the GitHub repository will occur on an ongoing basis**, to ensure timely, rapid, and persistent data availability.

A complete backup of the project’s Git repository will be stored on ASU's campuswide research storage system, and the services hosted on the university's UNIX Virtual Server facility.   These two services are designed to bring the advantages and economies of scale of professional IT facilities to research projects across the campus in a cost effective way.

The research storage facility provides a high integrity space for storing data, using enterprise-class NAS (Network Attached Storage) equipment.   In addition to the primary copy of the data being stored on redundant disk behind a cluster of NAS filers with failover capability, a backup copy of the data is created daily on a separate disk storage system in a separate building to provide protection and rapid restore from any data loss.   The snapshots of both the primary and backup copies of the data are taken daily, and access to the snapshots on the primary volume is available to users to make the recovery of accidentally deleted files as seamless as possible.   Daily snapshots are retained for a month, and monthly snapshots are retained for four months, with additional storage of snapshots available on demand.

The Virtual Server facility consolidates individual servers into a large central farm of virtual machines.   While the user of the virtual machine still retains the ability to tailor the software environment as if they owned their own physical server, individual research projects will not have to bear the cost of maintaining a physical system.  If the underlying physical server fails, the virtual server can be restarted immediately on another physical server in the farm.   The staff for the research project need not worry about service contracts, hardware maintenance, or periodic replacement of physical systems.   Consolidation into a single farm also allows the university's professional IT staff to maintain proper backups of the virtual server, as well as providing regular security sweeps to look for vulnerabilities or unusual behaviour.  Both the research storage and virtual server facilities are part of ASU's coordinated strategy of providing outstanding stewardship for cyberinfrastructure investment made by federal grants and contracts.

**Dissemination & Access**

All data objects generate by this project will be made publicly available via the GitHub repository (see Preservation) under a [ what license do you want to use ? ] license, except where prohibited by copyright.

[ you probably want to add more about dissemination, depending on your plans ]

**The use of all data obtained in the course of this research will adhere to established standards of applicable licenses and permissions, and all sources will be fully attributed.**