

## Data management plan

Our data management system will effectively manage the broad suite of data collected and produced for the project, including curation of high-quality research data and metadata for the project team during the project, and for dissemination following research completion. Our data will include (1) quantitative social survey data, (2) qualitative interview and focus group audio files and transcripts, (3) *Envision* and other submodel input and output data. Table 1 provides a detailed description of each dataset we expect to generate and their file formats, storage, and expected dissemination mode. We will archive all data in commonly available, nonproprietary formats. To ensure availability of nonconfidential, copyrightable data, we will release our data, following publications, using the open access institutional repository (e.g., DXScholar, UO Scholar's Bank) of the senior personnel responsible for that data.

**Roles and responsibilities** (back-up personnel, should listed if I leave project, shown in parentheses)

Data will be managed and retained by the relevant PI or co-PI leading each project component. Max Neilsen-Incuse (Brent Steel) will be responsible for audiofiles, transcripts and survey data generated through survey activities, the SES Stakeholder Council, and other work in our four study areas. Lisa Gaines (Dwayne Irlaza) will be responsible for all data related to the CHANS Fellows program. John Bolte (Dave Hulse) will be responsible for *Envision*-related model data, code and output files. Final data analysis codes will be managed and retained by the project PI or co-PI leading the publication for which it is used. I Johnson will assume that role for any co-PI who leaves the project. All co-PIs and senior personnel shall have right to access all project data by requesting it from the responsible party.

### Metadata

Generally, metadata for social science data collected during this project will follow the standards of the Inter-University Consortium for Political and Social Science using the Data Documentation Initiative-Codebook (DDI-C) specification. Our metadata will be a codebook with information about the overall data including the purpose and history of the data, collection methods, sources, scale and temporal coverage, as well as data confidentiality. For spatial data used in the *Envision* model, we will adopt the data quality standards for spatial data as specified in the Spatial Data Transfer Standard (Federal Information Processing Standard [FIPS] Pub 73-1). These include the following criteria: lineage (source), positional accuracy, attribute accuracy, logical consistency and completeness. Metadata for spatial coverage used in this effort will be maintained in standard ArcGIS format compliant with the Federal Geographic Data Commission (FGDC) Content Standards for Digital Geospatial Metadata (CSDGM), v. 2.1. Data quality will be assessed and documented in the codebook using standard techniques for geospatial data and reported in metadata for each data theme. Our metadata codebook will contain the parameters used to initialize *Envision* and generate outputs. In addition, our codebook will provide information about names, labels, and descriptions of variables, schemes for coding data, and similar features. Institutional libraries will provide descriptive metadata for data sets in Scholars' Bank using the Qualified Dublin Core international metadata standard. This allows them to be searchable through Web search engines, such as Google, and routinely found by international search engines and harvesters, such as OAIster.

### Storage, backup, and security

During the project period, data will be stored in three locations: password-protected campus-based computers, Basecamp (a cloud file storage and project management service) or another similar system, and each university's institutional file servers. Basecamp will be used primarily for uploading and back up of audio files and other documents collected during fieldwork and for sharing files, messages, and planning among project staff. Institutional file servers will be used as our regular file storage and backup system, and for long-term storage of data generated at each University.

All data and software codes developed will be stored in an existing Subversion repository hosted at OSU providing software versioning and revision control. This will ensure the preservation of current and historical versions of source code, documentation and utilized datasets. All project data used in modeling will reside on secure and backed up servers at OSU. All unrestricted geospatial datasets will be made publicly available through the Subversion repository; additionally, many of the most relevant datasets generated will be made available through a project website maintained at OSU.

The cloud file storage, Subversion repository, and project management service will be accessible to

---

<sup>1</sup> <http://basecampHQ.com>. See also, <http://37signals.com/security> for more detailed information on security

all project participants via a password-controlled log in, with access controlled by the PI. Data stored on Basecamp is written to multiple disks instantly, backed up daily and stored in multiple locations with full redundancy. The institutional files server is backed up nightly to multiple locations on each University campus. Each university also conducts a weekly backup that is stored off-campus for disaster recovery purposes. Data on the institutional files server is restricted with access control lists; user security, passwords, and access regulated by each University's central Information Services division.

#### Protection of privacy

The PSU and OSU Institutional Review Boards will protect interview and survey subjects from undue risk. During the project period a key code will link informant names and titles with interviews, and respondent names with survey data. The survey key code will be destroyed immediately after the survey dataset is finalized. The informant key code will be destroyed one year after the project's end. Because interview and focus group data cannot be made anonymous at a reasonable cost, transcript data will not be directly released. Publications will use quotations and themes developed from the interviews. Any geospatial datasets that have confidential information or information identifiable to a single private party will not be made available without express permission of the data source. Non-confidential datasets acquired by the project will be treated in accordance with access standards specified by the data provider.

#### Archiving and preservation.

Most project data will be archived at each university's open access institutional repository (e.g., DXScholars, UO's Scholar's Bank). The complete metadata will be stored in XML, and simple Dublin Core records created to describe the package, and also be stored in institutional repositories, which have backup systems to retain data in perpetuity. Material stored in university open access repositories is incorporated into the university library digital preservation plans. Interview and focus group audio files and transcripts will be kept for at least five years on institutional file servers. *Envision* initialization files and model output data are too large to be maintained in the university open access institutional repositories. Oregon State University maintains a dedicated *Envision* website (<http://envision.bee.oregonstate.edu/>) where final model setup files and output data will be maintained for a minimum of five years after the project end. Shareable data will be made available following acceptance for publication of the relevant data sources. *Envision* model setup files and output data will be made publicly available after finalization and prior to publication. Project datasets will be made available to other parties upon request following the first published use of that data by project personnel or after two years.

**Table 1. Data type and plan**

Data type	File type - generation	File type - sharing	File Size	Storage during collection & analysis	Vehicle for public release
Metadata codebook	Excel files (.xlsx), Word (.docx)	.xml, .csv, .pdf	< MB	Basecamp & Institutional files server	PDX Scholar Scholars' Bank
Interview & focus group audiofiles	.wav	.wav	< GB	Basecamp & institutional files server	Will not be released-cannot be de-identified
Transcripts	Word (.docx)	.pdf	< 100MB	Basecamp & institutional files server	Will not be released-cannot be de-identified
Social survey data	Excel files (.csv)	.csv	< MB	Basecamp & institutional files server	PDX Scholar Scholars' Bank
<i>Envision</i> source code, setup files & output data	<i>Envision</i> files (.envx & .xml) Excel files (.csv)	.envx, .xml, .csv, or .Rdata	< TB	Institutional files server, Subversion repository	OSU <i>Envision</i> Website
Final data analysis codes	SAS, Word (.docx)	.txt or .pdf	< MB	Basecamp & institutional files server	PDX Scholar Scholars' Bank
White papers	Word (.docx); InDesign (.indd)	.pdf	< MB	Basecamp & institutional files server	PDX Scholar Scholars' Bank
Peer-reviewed journal articles	Word (.docx)	.pdf & paper	< MB	Basecamp & institutional files server	Print & electronic journals