



Natural Hazard and Disaster Reconnaissance

**DATA PROCESSING AND ARCHIVING GUIDELINES**

**V1.0**

**Rev Date: 12/1/2022**

## Overview

The RAPID Facility (RAPID) supports reconnaissance teams by providing hardware and software for both data collection and processing, as well as the management of these data. In conjunction with DesignSafe-CI, the RAPID implements a data management plan that addresses collection, metadata, security, privacy, backups, archiving, publication, sharing, and re-use. The RApp software aids data collection efforts and provides tools for gathering metadata during the data collection process. The RAPID also provides devices (iPads) and storage for data backup in the field. Users are able to upload data files to DesignSafe-CI from the field, the RAPID Facility, or any place with sufficient network connectivity. The RAPID data management responsibilities cover all the steps from data collection to transferring data to DesignSafe-CI. To ensure the smoothest workflow for moving data from RAPID equipment to DesignSafe-CI, the RAPID works closely with DesignSafe-CI to establish common data models and file formats.

## Purpose

- Assist users with understanding data formats and empower them to extract more information/products from the data
- Assist users with preparing data to share with the larger user community
- Ensure data products meet quality standards and are processed systematically
- Promote data re-use

## Roles and Expectations

- RAPID users are expected to publish their data in a timely manner on DesignSafe (ideally within 6 months of collection) to share with the broader community. Not only does the RAPID encourage data sharing as a value, NSF expects RAPID users to share data and may require evidence of that at the reporting stage. Users may continue to process and refine data in DesignSafe at a later date.

- RAPID users are expected to comply with disciplinary ethical standards related to data collection and publication. At minimum, RAPID users should receive approval from their institutional review board (IRB) for collection, sharing and reuse of data for human subjects research. All human subjects data should be collected and published in accordance with protocols approved by the IRB. Additional IRB approvals may be necessary for multi-institutional teams or for international research.
- Extreme Events Research/Reconnaissance (EER) networks (EERs) are expected to promote these guidelines and build up data processing capacity.
- CONVERGE promotes data sharing across a variety of disciplines and organizations. They also help inform organizations and researchers on interdisciplinary data needs to promote the collection of high-quality datasets that can be drawn upon from multiple disciplines.
- RAPID staff will assist with the data processing of NSF funded or EER-led datasets from natural hazard and disaster reconnaissance deployments.
- RAPID will provide data processing training opportunities.
- Additional data analysis will incur data processing charges and is dependent on RAPID staff availability.
- DesignSafe-CI provides infrastructure, assists with metadata, and develops tools that enable sharing and use of the data, such as HazMapper.
- The research community are expected to use the data, produce and share more advanced products, and notify authors of issues in data when discovered.

Non-NSF funded projects may not have the same requirements to archive data. However, these guidelines should still be followed as much as possible for efficiency and to ensure that systematic, quality products are produced to promote broader usage.

## Data Types and Formats

Data are collected via a wide variety of instruments such as lidar scanners, GNSS receivers, total stations, digital levels, aerial imaging drones, accelerometers, penetrometers, digital cameras, and the RAPID Recon App (RApp). General collected data includes images in JPEG, PNG, TIFF, and RAW formats; audio files recorded in MP3 and CAF formats; text files stored in both TXT and RTF formats; questionnaires stored in XML and/or JSON formats; and geographic items such as markers, areas, and routes stored in GeoJSON format. More specialized instruments also gather data in a variety of formats, many of which are proprietary and specific to the manufacturer, such as Leica DBX, m00, blk, and bin files for total stations, GNSS receivers, BLK scanners, and P50 scanners, respectively. However, these data can be converted into open or more commonly available formats via the manufacturer's software after initial processing. The RAPID preserves copies of the initial source data in the proprietary format to minimize data loss, but also works with users to publish their data and derived products in formats that are open, accessible, and widely adopted in the industry to support the reuse and longevity of the data. For example, SfM/MVS point clouds are made available in ASPRS LAS/LAZ, lidar point clouds in APSRS LAS/LAZ or ASTM E57, orthophotos from drone imagery

in GeoTiff, terrain or object models in OBJ or DXF format, and extracted features or data subsets in CSV format.

Metadata is attached to all data sets in conformance with DesignSafe-CI established standards. Metadata includes the event, location, instrument, time of day, and other relevant information. As much as possible, RAPID software prompts and/or fills out metadata for the user at the time of data collection. Users have the ability to modify or add metadata at a later time.

## Data Storage and Retention

Users are provided with data loggers, high-speed USB drives, SD cards, and portable hard drives to back up collected data in the field. The provided storage ensures users are able to make a sufficient number of backups in the field, with additional devices provided to account for drive or device failures. Upon returning from the field, data is not deleted from instruments or backup devices until the data has been moved to RAPID or DesignSafe-CI servers and data quality is verified.

Data products derived from other data files are treated in a similar fashion to raw data files. Facility personnel provide details on required metadata, however, users are responsible for adding appropriate metadata information. Data products created on RAPID computers are automatically backed up to ensure redundancy.

DesignSafe-CI provides data archiving and curation services for all RAPID deployments. Post-deployment, data are transferred to DesignSafe-CI via Globus.org by RAPID personnel or Users. Data are automatically copied from RApp to DesignSafe-CI when connected to a stable wifi network. Once data are uploaded to DesignSafe-CI, backing up, versioning, and archiving are performed automatically per DesignSafe-CI policies and procedures. The DesignSafe CI maintains all uploaded data on the storage resources of the Texas Advanced Computing Center. These resources are redundant and geographically replicated to ensure robust archival and preservation of data. The DesignSafe Data Depot is designed to meet NSF requirements for preservation.

## Data Publication, Sharing, and Re-use

All users should follow the [RAPID's Data Publishing Guidelines](#). The RAPID works closely with project PIs to ensure that raw instrument data is processed into a form that is usable for both their project and by the greater research community. RAPID staff provides a supporting role in helping PIs to develop products for traditional academic forums such as journals, conferences, and workshops.

DesignSafe-CI provides mechanisms for making data available for publication and assigning DOI numbers ([Refer to DesignSafe's website](#)). Metadata must meet a minimum requirement in order to be publishable. Users have complete control over the data publication process. RAPID staff can provide PIs guidance in the curation and publication process.

DesignSafe-CI also provides mechanisms, such as [Data Depot](#) and [HazMapper](#), that allow users to make data and derived data products available for use by other users and researchers. Users have complete control over which data and products are made available for sharing.

The RAPID places significant importance on preserving the raw data such that investigators have the ability to verify the quality of derivative products or create new products based on different processing procedures for their specific application.

A dataset is cited using the same format used to cite a journal article or a conference proceeding in the reference section of a paper. You can use the citation language and DOI conveniently provided by DesignSafe on the data publication's landing page (See [Data Reuse](#) on DesignSafe's website). In addition to citing a dataset, please cite the RAPID Facility with the following two publications:

Berman, J.W., Wartman, J., Olsen, M., Irish, J.L., Miles, S.B., Tanner, T., Gurley, K., Lowes, L., Bostrom, A., Dafni, J., Grilliot, M., Lyda, A., Peltier, J., 2020. Natural Hazards Reconnaissance With the NHERI RAPID Facility. *Front. Built Environ.* 6, 1–16. <https://doi.org/10.3389/fbuil.2020.573067>

Wartman, J., Berman, J.W., Bostrom, A., Miles, S., Olsen, M., Gurley, K., Irish, J., Lowes, L., Tanner, T., Dafni, J., Grilliot, M., Lyda, A., Peltier, J., 2020. Research Needs, Challenges, and Strategic Approaches for Natural Hazards and Disaster Reconnaissance. *Front. Built Environ.* 6, 1–17. <https://doi.org/10.3389/fbuil.2020.573068>

Human subjects data should be published according to protocols approved by relevant IRBs. RAPID users may wish to take additional steps to mitigate the potential for unintended consequences of data publication for disaster-affected individuals and communities, as well as to reduce the potential for retraumatization and protect privacy, including facial or license plate blurring, anonymization or aggregation of data to prevent creation of individually identifiable data through triangulation with other data sources, and/or removal of aggregation of geolocation data from RApp datasets.

## Data Quality (DQ) Indicators

Data are produced both by researchers and RAPID staff. To build confidence in data products as well as communicate the intent of the original data, the following data processing indicator levels are provided.

**DQ0:** Raw data from the sensor; data has not been processed. Spot checks of the data have been conducted in the field or shortly thereafter to ensure its basic integrity.

**DQ1:** RAPID staff or user has performed basic processing of the data with minimal Quality Assurance/Quality Control (QA/QC). Includes data that was not collected with rigorous workflows in the field, but was sufficient for the user's needs and time constraints.

**DQ2:** RAPID staff or appropriately trained and experienced user has reviewed the dataset and performed QA/QC. Data processing has been completed with basic workflows and primarily automated processes.

**DQ3:** RAPID staff or appropriately trained and experienced user(s) have thoroughly reviewed the dataset, with enhanced processing strategy and rigorous QA/QC. Data have been processed using advanced workflows (e.g., rigorous least squares adjustments, manual cleanup for ground filtering, etc.) for high-quality results.

While the RAPID certainly encourages the highest quality data possible, it also recognizes the unique constraints associated with collecting data in reconnaissance. In some cases, DQ1 may be sufficient to meet the objectives of the reconnaissance deployment, and the level of effort to produce a DQ3 dataset is not needed or may not be possible in a timely manner.

These DQ Indicators should be recorded in the Metadata of the project.

There are many factors that dictate what resources the RAPID can provide. If there are sufficient staff resources, the RAPID can support all DQ levels for a fee depending on the project and the time commitment required. Users should discuss their data processing needs with the RAPID Facility.

*Examples of data quality indicator levels for different data sources.*

	<b>DQ 0</b>	<b>DQ 1</b>	<b>DQ 2</b>	<b>DQ 3</b>
<b>Typical RAPID facility support provided without charge</b>	All projects	NSF Projects	NSF-funded EERs.	Paid services only
<b>Lidar (TLS / ULS)</b>	Raw scans in manufacturer proprietary format.	Basic or preliminary local registration and trajectory processing. Basic QA/QC (Visual, Slices).  <u>Primary Products:</u> E57/LAZ files and reports.	Register data with default workflow and settings. More QA/QC (Random check, cut planes, visual), georeferencing if possible.  <u>Primary Product:</u> E57/LAZ files, QA/QC screenshots, and registration diagnostics.	Custom workflow and settings to push the registration accuracy to meet the requirements of the users' applications. georeferencing if possible.  <u>Primary Product:</u> E57/LAZ files, QA/QC screenshots, and registration diagnostics.
<b>Drone/UAS SfM MVS Photogrammetry</b>	Raw images in manufacturer proprietary format.  RTK GNSS logs from flights  Flight trajectory	Basic dense point cloud or orthophoto reconstruction in Pix4D with minimal targets. Basic QA/QC  <u>Primary Products:</u> LAZ	Perform more detailed point cloud reconstruction and thorough image quality assessment. Dense Target network. Thorough	Custom workflow and settings to push the georeferencing accuracy (e.g., strip adjustments) to meet the requirements of

		files, moderate resolution orthophoto (dm resolution), and control point residuals.	QA/QC (Random check, cut planes, visual).  <u>Primary Product:</u> LAZ files, orthophoto (cm resolution), and control point residuals.	the users' applications.  <u>Primary Product:</u> LAZ files, orthophoto (cm resolution), and control point residuals.
<b>Street View</b>	Raw data from device  Processed panorama images	Same as DQ 2	Data upload to Mapillary	
<b>Site Characterization</b>	<i>Standards in progress</i>			
<b>GNSS</b>	Raw .m00 files and project files	Same as DQ 2	Raw .m00 files and project files  Processed GNSS to the best coordinate quality possible given data collection limitations  <u>Primary Products:</u> CSV, Point Quality Report	
<b>Z-Boat</b>	Raw trajectory and sounding data.	Processed inertial trajectory  Elevation-based bathymetry exported in preferred coordinate system	Geodatabase - Polyline trajectories - Interpolation rasters - Projection files	Echosounder amplitude values  More rigorous QA/QC

# Coordinate Systems

**UNLESS OTHERWISE SPECIFIED BY THE USER, FOR ALL PROJECTS IN THE UNITED STATES, THE RAPID FACILITY WILL DEFAULT TO COLLECT, PRODUCE, AND PROVIDE DATA IN THE 1983 STATE PLANE COORDINATE SYSTEM IN SI UNITS (e.g., METRIC SYSTEM) REFERENCED TO NAD83(2011) EPOCH 2010.00 WITH GEOID18 FOR THE VERTICAL. (NOTE THAT AT A FUTURE DATE (EST. 2025) THIS WILL BE REPLACED WITH A NEW VERSION OF THE STATE PLANE COORDINATE SYSTEM). FOR INTERNATIONAL LOCATIONS, THE MOST RELEVANT NATIONAL GRID OR UTM COORDINATES WILL BE USED.**

The project team should discuss and determine the best coordinate system for the project as soon as possible, ideally before data collection. Coordinate system selection must consider many factors. For example, coordinate systems that cover large areas tend to have more distortion. Selecting a highly specialized coordinate system may create issues when working in some software and sharing data with others. It is also essential that everyone working on the project is consistently using the same system, including the projected coordinate system, reference frame, and vertical datum with the specified units and epochs. The RAPID facility highly recommends using the EPSG code (<https://epsg.io/>) in the metadata and all communications to avoid confusion between similar coordinate systems.

All project personnel must clearly specify the following in the project metadata:

- Projected Coordinate System (e.g., WA State Plane N)
- Reference Frame (e.g., NAD 83 (2011))
- EPSG code of the coordinate system (e.g., EPSG:6596)
- Epoch (e.g., 2010.00)
- Geoid (e.g., Geoid18)

## Datums/Epochs

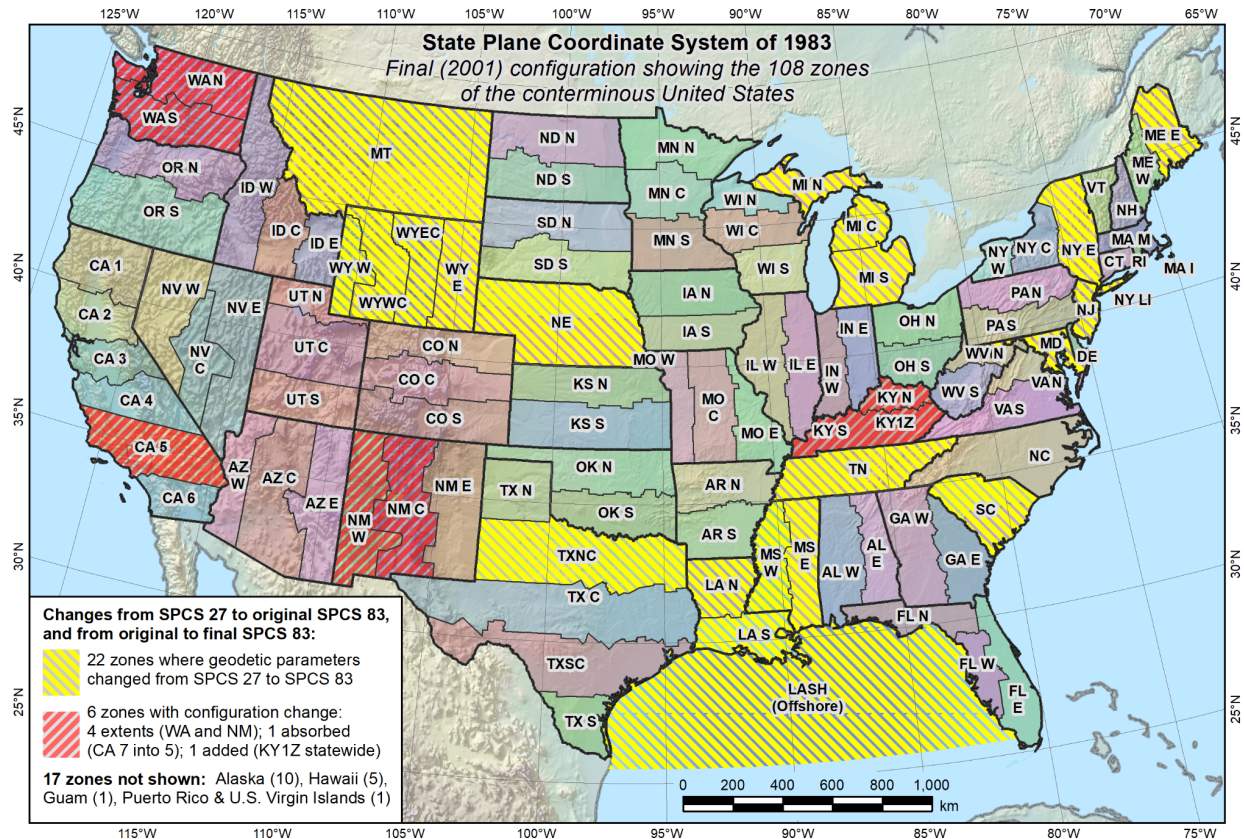
Datums or reference frames are set up so that measurements can be taken with a clear reference. There are many models of the shape of the earth that are used in these reference frames. Periodically reference frames are updated to account for tectonic changes. The center of mass of the earth changes with time and methods to observe and model the earth also improve. Geodesists track these with GPS\GNSS, gravimetric data, and other survey observations and make periodic adjustments to provide better models for positioning with minimal error. The coordinates of an object will change with time as the earth is dynamic. The epoch of observation can help relate data collected over time in a dynamic location. Epochs are always important to record but are critical given seismic movement in earthquake-prone regions (e.g., California) and in areas with substantial active subsidence (e.g., Houston).

## Geoids

A 'geoid' is the shape that the ocean surface would take under the influence of Earth's gravity and rotation if other influences such as winds and tides were absent. This smooth, but irregular surface is extended through the continents to create a mathematical "sea level" from which to measure elevations. Technically it models an equipotential gravimetric surface as sea level changes with time.

Once the NGS releases the new Reference Planes and Associated State Plane Coordinate System of 2022, the metric version of this system will be the preferred system for RAPID

surveys in the US. Different zones exist in the state plane coordinate system so the most relevant zone should be selected.



**State Plane coordinate system zones (developed by Michael Dennis (NGS)).**

Here is a good tutorial explaining datums, geoid, and more and why they are important:  
<https://vdatum.noaa.gov/docs/datums.html>

Component	Link	Comments
SPCS2022	<a href="https://geodesy.noaa.gov/datums/newdatums/index.shtml">https://geodesy.noaa.gov/datums/newdatums/index.shtml</a>	Forthcoming
NATRF2022		Forthcoming
GEOID2022		Forthcoming
SPCS83	<a href="https://geodesy.noaa.gov/SPCS/">https://geodesy.noaa.gov/SPCS/</a> <a href="https://pubs.er.usgs.gov/publication/pp1395">https://pubs.er.usgs.gov/publication/pp1395</a>	Current system, will be replaced by SPCS2022.
NAD83(2011) Epoch 2010.00	<a href="https://beta.ngs.noaa.gov/web/surveys/NA2011.20111122/">https://beta.ngs.noaa.gov/web/surveys/NA2011.20111122/</a>	Current



Geoid18	<a href="https://geodesy.noaa.gov/GEOID/GEOID18/">https://geodesy.noaa.gov/GEOID/GEOID18/</a>	Current
---------	---	---------

Outside the continental US, the local national grids are generally preferred if they are mature systems. If no coordinate system is decided, the RAPID will generally use UTM as it is widely used. While UTM is a common system used across the globe, caution should be exercised when using this system:

- UTM suffers from significant distortion due to the projection (1:1000 nominally. That means that if you measure 1000 m on the ground, that same distance based on the projected grid could differ by ~1 m. However, the actual amount varies depending on the topographic relief and variability.
- Selecting the appropriate UTM Zone for your study area is critical. When your study area spans multiple zones, select the one that best matches the majority of your study area.

## Data Products

The following are suggested data products to be created and provided to the community. **ALL DATA SHOULD BE COLLECTED, PRODUCED, AND PROVIDED IN SI UNITS (e.g., Metric system) to minimize downstream confusion, issues, and errors.**

If partners (e.g., government agencies) want data in US Customary units, clearly marked alternate versions can be produced, but should simply be produced by scaling the data and exporting separate versions on the final products only. Once the SPCS is released, there will only be one definition of the foot: the international foot that equals 0.3048 m EXACTLY. [Read more here.](#)

Data Source	Suggested shared product	Data Format	Notes
All Sensors	<ul style="list-style-type: none"> <li>• Metadata</li> </ul>	<ul style="list-style-type: none"> <li>• Docx or PDF</li> </ul>	<ul style="list-style-type: none"> <li>• Write up of data processing and QA/QC methodologies</li> <li>• Error report</li> <li>• Notes on issues in data</li> </ul>
Terrestrial Laser Scanning	<ul style="list-style-type: none"> <li>• Registered/georeferenced point cloud</li> </ul>	<ul style="list-style-type: none"> <li>• ASTM E57 (preferred) or ASPRS LAZ (v1.4)</li> </ul>	<ul style="list-style-type: none"> <li>• E57 is preferred to provide photographs with data</li> <li>• Each station should be in a separate E57 file</li> </ul>
	<ul style="list-style-type: none"> <li>• Inventory file of scan</li> </ul>	<ul style="list-style-type: none"> <li>• CSV and KMZ</li> </ul>	<ul style="list-style-type: none"> <li>• XYZ, ScanName</li> </ul>

	locations		<ul style="list-style-type: none"> <li>Roll Pitch Yaw Preferred. (Euler angles in a right-handed coordinate system)</li> </ul>
	<ul style="list-style-type: none"> <li>Scan position images</li> </ul>	<ul style="list-style-type: none"> <li>JPEGs</li> </ul>	
<b>UAS lidar or UAS SfM/MVS photogrammetry</b>	<ul style="list-style-type: none"> <li>Georeferenced, unclassified point cloud, black points removed</li> </ul>	<ul style="list-style-type: none"> <li>ASPRS LAZ (v1.4 or higher)</li> </ul>	<ul style="list-style-type: none"> <li>Each flight line should be in a separate .laz file for lidar, single file for SfM/MVS</li> <li>Classified for topographic datasets only require additional data processing</li> </ul>
	<ul style="list-style-type: none"> <li>Bare Earth Raster DEM</li> </ul>	<ul style="list-style-type: none"> <li>ESRI BIL or BSQ format</li> </ul>	<ul style="list-style-type: none"> <li>Single raster per site (topographic project only)</li> </ul>
	<ul style="list-style-type: none"> <li>Photographs or video logs</li> </ul>	<ul style="list-style-type: none"> <li>JPGs or MP4s</li> </ul>	<ul style="list-style-type: none"> <li>Store in subfolder(s)</li> </ul>
	<ul style="list-style-type: none"> <li>Trajectories</li> </ul>	<ul style="list-style-type: none"> <li>KMZ, OUT, CSV</li> </ul>	
	<ul style="list-style-type: none"> <li>Photograph locations</li> </ul>	<ul style="list-style-type: none"> <li>KMZ, CSV</li> </ul>	
<b>Street View</b>	<ul style="list-style-type: none"> <li>Photographs or video logs</li> </ul>	<ul style="list-style-type: none"> <li>JPGs or MP4s</li> </ul>	<ul style="list-style-type: none"> <li>Store in subfolder(s)</li> </ul>
	<ul style="list-style-type: none"> <li>Drivepath trajectory</li> </ul>	<ul style="list-style-type: none"> <li>KMZ</li> </ul>	
	<ul style="list-style-type: none"> <li>Photograph locations</li> </ul>	<ul style="list-style-type: none"> <li>KMZ, CSV</li> </ul>	
<b>Z Boat</b>	<ul style="list-style-type: none"> <li>Georeferenced point cloud of depths and/or elevations</li> </ul>	<ul style="list-style-type: none"> <li>XYZ</li> <li>ASPRS LAZ (v1.4)</li> </ul>	<ul style="list-style-type: none"> <li>Each survey session should be a separate laz file</li> </ul>
	<ul style="list-style-type: none"> <li>Trajectories</li> </ul>	<ul style="list-style-type: none"> <li>OUT</li> </ul>	
	<ul style="list-style-type: none"> <li>Interpolated surfaces</li> </ul>	<ul style="list-style-type: none"> <li>TIF</li> </ul>	
<b>GNSS</b>	<ul style="list-style-type: none"> <li>Raw static data</li> </ul>	<ul style="list-style-type: none"> <li>Provide both m00 Leica Proprietary format and Rinex 3.11</li> </ul>	

	<ul style="list-style-type: none"> <li>Final point coordinates</li> </ul>	<ul style="list-style-type: none"> <li>CSV file</li> </ul>	<ul style="list-style-type: none"> <li>CSV file should contain point ID, XYZ coordinates, standard deviations/precision estimates per point, and solution type per point (e.g., float/fixe).</li> </ul>
<b>Total Station Survey Data</b>	<ul style="list-style-type: none"> <li>Final point coordinates</li> </ul>	<ul style="list-style-type: none"> <li>CSV file</li> </ul>	<ul style="list-style-type: none"> <li>CSV file should contain point ID, XYZ coordinates, standard deviations/precision estimates per point</li> </ul>
<b>RApp</b>	<ul style="list-style-type: none"> <li>Via Hazmapper &amp; DesignSafe DOI</li> </ul>		<ul style="list-style-type: none"> <li>Import directly into DesignSafe</li> </ul>
<b>Seismometers</b>	TBD	TBD	TBD
<b>ATOM System</b>	TBD	TBD	TBD
<b>GIS extracted features</b>	<ul style="list-style-type: none"> <li>Points, polylines, polygons of extracted features</li> </ul>	<ul style="list-style-type: none"> <li>GDB of .shp files</li> </ul>	
<b>2D Modeled objects</b>	<ul style="list-style-type: none"> <li>2D linework</li> </ul>	<ul style="list-style-type: none"> <li>DXF</li> </ul>	
<b>3D Modeled objects</b>	<ul style="list-style-type: none"> <li>3D models</li> </ul>	<ul style="list-style-type: none"> <li><a href="#">IFC</a></li> </ul>	<ul style="list-style-type: none"> <li>DXF may be used as a substitute if IFC is not a possibility in the software.</li> </ul>
<b>Questionnaires</b>		<ul style="list-style-type: none"> <li>JSON, XML</li> </ul>	

RAPID users and staff may produce and publish files in other formats, including available proprietary formats. However, adherence to these standards provides data in open formats that are more broadly accessible.

## Metadata Specifications

Ideally, RAPID users will utilize RApp to populate the metadata in DesignSafe. A short writeup of the project should be provided in DesignSafe with the data so that other users understand how the data were collected and processed, which is not captured in the metadata. This writeup should include:

- Data collection objectives (i.e., why the data were collected)
- List of Equipment used (including model numbers)
- Field methods employed, including a : description of the field workflow and general instrument settings (e.g., panoramic scan at 0.02 m at 50 m resolution)
- Short description of data processing (e.g., registration methods, any cleaning of the data, software tools used)
- Details of the coordinate system used for the project
- Statement of accuracy (e.g., a summary of outputs), the associated DQ level, and any issues noted in the processing or QA/QC

# Appendix A. File Structure

## RAPID Data Structure within the DesignSafe Project

/Parent DesignSafe Project (Internal RAPID use only: RPD####\_PI-Name\_StudySite)

- RApp
  - RApp data collected during deployment in this folder.
- RAPID\_EF
  - 1\_Raw
    - Place all raw data and data collection field notes gathered during the deployment in this folder. It should be a direct copy from a primary field backup device.
  - 2\_Processing
    - Important processing folders, notes, and files should be placed here. These are resources that are pivotal to the processing work and should be kept in case items need to be revisited, but would not be delivered to the public or published.
  - 3\_Deliverables
    - Final editions of the processed data. What could be published or delivered to the public.
  - 4\_Metadata
    - Keep all supporting documents that would be used in any future metadata report. The final metadata report will be kept here as well, including the deployment report.

## Standard Folder and File Naming Conventions

Avoid the use of special characters and spaces when naming folders and files. Some programs and scripts are unable to access file paths that contain them, causing processing to fail.

## RAPID Recon App (RApp) Folder Structure

RApp data is organized by the data collector. A folder is created for each DesignSafe user in the project who collects data with RApp. Within each user folder, data (e.g., photos, voice notes, and videos) are either listed as single files or grouped by collection folder.

## Raw Data Folder Structure

The raw data folder is divided into instrument subfolders. Within those folders, data is further divided by collection location and date. The raw data files are the files directly pulled from the instrument. Ideally, these raw files should be named with a location and date.

**1\_Raw >> *Instrument* >> *Location/Collection Date Folder* >> raw data files**

## Processing Folder Structure

The processing data folder is divided into software program folders. Within these folders, data is organized by 'project folders' that are automatically generated upon project setup in the

software. The name given to a project in the software should be intuitive, including the data collection location and date, so that the resulting ‘project folders’ and files within automatically have descriptive names.

**2\_Processing >> Software Program >> Processing Project Folder >> <description\_YYYYMMDD>**

Additionally, processing note files should be included to describe the settings and steps used to process the data.

Deliverables Folder Structure

The deliverables folder structure will vary slightly depending on deployment magnitude.

For larger deployments that span multiple trips, locations, or epochs, the deliverables structure is:

**3\_Deliverables >> location/epoch >> data type >> data type structure >> file <uniqueid\_description\_YYYYMMDD>**

Note: Not all deployments will have multiple trips or epochs; they may just have multiple locations. In this instance, exclude the trip and epoch folders.

For smaller deployments with reduced duration and spatial extent, exclude the location/epoch folders, so that the data structure is:

**3\_Deliverables >> data type >> data type structure >> file <uniqueid\_description\_YYYYMMDD>**

In both cases, the deliverables data is primarily classified by *data type*. The available *data type* classifications are DEM, Geotech, Imagery, Mesh, Pointclouds, Structural, Survey, Mapping, and Other. Deployments will often have a few, not all, of these data types.

Within each *data type* folder, there is a *data type-specific file structure*, which is outlined in the following table. The underlined subclassifications denote a folder, with the bulleted data indicating what file extension types may be present in each folder. Please contact the RAPID Facility to discuss the deliverables you are hoping to obtain.

**Data Deliverables and their associated file structure**

<b>DTM</b>	<b>Geotech</b>	<b>Imagery</b>	<b>Mesh</b>
<u>DSM</u> - .prj - .tfw - .tif	TBD	<u>Ortho</u> - .prj - .tfw - .tif	<u>Mesh</u> - .ply/.fbx - .obj - .mtl - .jpg/.png
<u>Bare Earth DEM (upon request)</u> - .prj		<u>Pano (Streetview)</u> - .jpg - .tif	<u>Spectral</u> - .ply

<ul style="list-style-type: none"> <li>- .tfw</li> <li>- .tif</li> </ul>		<u>Spectral</u> <ul style="list-style-type: none"> <li>- .tif</li> </ul> <u>Google Tiles (upon request)</u> <ul style="list-style-type: none"> <li>- Zoom-level folders</li> <li>- .html</li> <li>- .kml</li> </ul>	<ul style="list-style-type: none"> <li>- .jpg</li> </ul>
<b>Pointclouds</b>	<b>Structural</b>	<b>Survey</b>	<b>Mapping</b>
<u>SfM</u> <ul style="list-style-type: none"> <li>- combined .laz</li> </ul> <u>TLS</u> <ul style="list-style-type: none"> <li>- individual .e57</li> <li>- individual .laz</li> <li>- combined .laz</li> </ul> <u>ULS</u> <ul style="list-style-type: none"> <li>- combined .laz</li> </ul> <u>Bathy</u> <ul style="list-style-type: none"> <li>- depth 1 XYZ</li> <li>- depth 2 XYZ</li> <li>- elevation XYZ/.laz</li> </ul> <u>Z Metadata</u> <ul style="list-style-type: none"> <li>- reports</li> </ul>	TBD	CSV file (XYZ PointID)	<u>GIS</u> <ul style="list-style-type: none"> <li>- GDB of vector and/or raster data</li> </ul>

\***Other** accounts for sensor data and for future instruments with other standard deliverables.

Each of the files contained in these folders follows the naming convention:

**<uniqueid\_description\_YYYYMMDD>**

where

- *Uniqueid* is an identifier to help users distinguish what the data is representing (ex. Location or epoch).
- *Description* is a short statement to provide context should a user download data individually (ex. Scan setup #, TLS, SfM, equipment used (RTC360 or P50), ortho, pano, spectral, report).
- YYYYMMDD is the data collection date.

# Appendix B. Common Data Formats

## A.1 APSRS LAZ/LAS

ASPRS LAS has multiple versions that enable different information to be stored. The current standard is LAS 1.4 Rev 15.

[http://www.asprs.org/wp-content/uploads/2019/07/LAS\\_1\\_4\\_r15.pdf](http://www.asprs.org/wp-content/uploads/2019/07/LAS_1_4_r15.pdf)

LAS also has different point record formats to store data. XYZ data is necessary for all point clouds. RGB should be included for colorized point clouds. Intensity (strength of the return signal) of the point clouds can reflect the characteristics of the object surface to some degree. Including timestamps can also be helpful for kinematic data or test data that may need to be time synchronized. The return number and number of returns are common fields in airborne lidar and some of the UAS lidar data.

LAZ is a lossless compressed LAS file. Whenever possible LAZ should be used to minimize file size (it compresses by a factor of approximately 8). This saves space on the servers and makes upload/download times much faster. When necessary, the EZPC Toolkit can be used to convert between LAS and LAZ.

Georeferencing information can be stored in the LAS header; however, not all software handles this properly.

For classification, the ASPRS codes should be used where 1= unclassified and 2= classified when ground filtering has been done on the dataset. (Certainly, additional classifications can be done in the files beyond codes 1 and 2).

Limitations: ASPRS LAS does not directly store the data structure of how the data were collected like E57. Also, images and associated calibration information are stored as separate files and not linked to the point cloud.

## A.2 ASTM E57

ASTM E57 format  
(<https://www.ri.cmu.edu/publications/the-astm-e57-file-format-for-3d-imaging-data-exchange/>) stores point clouds and image data with their acquisition structures and camera calibration information, respectively. It enables multiple scans and images to be stored in the same file as the scan if desired. The standard and implementation are available at <https://www.astm.org/Standards/E2807.htm> and <http://www.libe57.org/>.

Limitations: Unlike ASPRS LAS, E57 format does not have a compressed version (although the committee is working on that), which can result in large files. Practically speaking, it is preferable to have each scan in a separate e57 file even though you can store them all in one. This will make it easier for people to download files and minimize the chance of corruption\copy errors that are more likely for very large files.



### A.3 OBJ format

The Wavefront OBJ format is an ASCII data format that stores the surface models. The OBJ file itself includes all the geometric information of the 3D model (e.g., vertices, facets) while the color and texture information can be stored in a PNG file which is associated to the OBJ file via a Wavefront Material Template Library (MTL) file. More information about the OBJ format can be found here: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000507.shtml>

### A.4 ESRI BIL/BIP/BSQ

ESRI BIL, BIP, and BSQ formats can be used to store a raster in a binary file containing one or multiple channels. An ASCII header file in HDR format is required to preserve the basic information of the raster. More information can be stored in other associated files such as world files (world files have the extension of the first and last letters of the image extension followed by a W). and statistics data. It is worth noting that the world files can also be used in mapping other raster formats such as JPG, TIFF, and PNG. More information can be found here: <https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/bil-bip-and-bsq-raster-files.htm>

### A.5 IMAGES

Type	Uses	Comments
JPG, JPEG	photographs	JPGs are very mature for photographs; however, they are lossy. As a result, they should be avoided for most other data production (e.g., image renders, screenshots)
PNG	Image renders	Lossless
TIFF, GeoTIFF	GIS raster layers and publication figures	Can implement compression, some lossy and some lossless

### A.6 Geodatabase

Geodatabase is a collection of geographic datasets of various types held in a file and folder structure supported by ArcGIS. Note that it is not recommended to store point cloud data directly into a geodatabase. More information can be found here: <https://pro.arcgis.com/en/pro-app/latest/help/data/geodatabases/overview/what-is-a-geodatabase.htm>

## A.7 RINEX Format

The Receiver Independent Exchange (RINEX) format that was developed for the easy exchange of GNSS data. Most of the GNSS units and software support this ASCII-based format to store the observation information. The latest version is RINEX 4.00 ([https://files.igs.org/pub/data/format/rinex\\_4.00.pdf](https://files.igs.org/pub/data/format/rinex_4.00.pdf)), released in December 2021. It is worth noting that there is a file naming convention for the RINEX format, and once the files are exported from the equipment or software, the file names should not be modified. All the associated metadata should be stored and archived properly as well.

## Appendix C: Coordinate Systems

This table provides information on recommended, robust local coordinate systems at locations across the globe where RAPID has worked.

Location	Coordinate Systems	Entity	Link
New Zealand	NZGD2000	LINZ	<a href="https://www.linz.govt.nz/data/geodetic-system/datums-projections-and-heights/geodetic-datums/new-zealand-geodetic-datum-2000-nzgd2000">https://www.linz.govt.nz/data/geodetic-system/datums-projections-and-heights/geodetic-datums/new-zealand-geodetic-datum-2000-nzgd2000</a>
United States	State Plane NAD83 (2011) NAVD88 height	NOAA/NGS	<a href="https://geodesy.noaa.gov/SPCS/">https://geodesy.noaa.gov/SPCS/</a>
Germany	ETRS89 / UTM zone 32N DHHN2016 height using GCG 2016 Geoid	BKG	<a href="http://gdz.bkg.bund.de/index.php/default/quasigeoid-der-bundesrepublik-deutschland-quasigeoid.html">http://gdz.bkg.bund.de/index.php/default/quasigeoid-der-bundesrepublik-deutschland-quasigeoid.html</a>