Quick Start basics: Data quality checking

1. Aim is to avoid build script crashes when new data loaded

Refer to Quick start Data Transfer guide for checklist of data items that require regular checking

1.1 Establish a protocol and quality check log to minimise workload
1.2 Timing and frequency

- Undertake regular checks between data transfers
- Data will need to be quality controlled at minimum prior to each transfer of 'new' data (surveys and resources)

1.3 If errors discovered during a build

- log the feedback from the build team
- correct your original data immediately
- sign-off on the correction

2. Tools to help the process

Print folder & file lists to text for faster/easier checking (instructions at links below):

Windows: see
https://support.microsoft.com/en-au/help/196158/how-to-create-a-text-file-list-of-the-contents-of-a-folder

Mac: see free standard 'Print Windows' at
http://searchwaresolutions.com/default.html

You may have to print the file lists by folder.

3. Key areas of concern

3.1 Folder structure and folder/file naming conventions

These must follow the specified conventions for Minimap, Image processing and Resource Inventory (see guides for info/ illustration).

Be aware
Underscore and/or hyphen may be used in filenames (as in specified templates) with the file extension separated by a full stop from the filename as customary. Otherwise, filenames and especially folder names should not contain any spaces, symbols or punctuation. Pay close attention to case.

3.1.1 Check for 'typos' in folder and filenames, such as arbitrary spaces or change in case, especially file extensions:

- Open ‘list of content’ text file(s) in text editor e.g. Word or similar
- Set up find/replace dialogue to locate and highlight 'dis-allowed' characters
- Re-name actual files in directory/folder to fix issues identified in text file
- Re-print (to text) the ‘edited’ folder and file list(s)

3.1.2 Check names of folders/files exactly match those listed in:

- ResourceInventory.xls
- MinimapNodeSurveyData.xls
- surveys.csv

Suggestion:

- Merge all the re-printed lists from each main directory into a single document, create an equivalent document from appropriate excel file and ‘compare’ the two documents’ e.g for Resources
    - Merge the ‘content list’ text files following the ResourceInventory category & sub-category pattern (doc 1)
    - Copy file lists from Resource Inventory spreadsheet into text (e.g. Word) document (doc 2)
    - Open both docs 1 & 2, use ‘Compare’ from Review tab in Word (windows) or from Track Changes (from Tools menu) in Word for Mac

3.2. Nodes and Surveys:

3.2.1 Use entries in MinimapNodeSurveyData.xls to help check

- Coordinates of nodes make positional sense on minimap/ grid
- Floor levels appropriately mapped (a LevelXPlan.png exists for each level included in new surveys)
- Nodes documented and match survey node panorama jpg filenames
- Each survey Virtual Tour includes all nodes surveyed on that survey date and no nodes shot on a different day i.e. date of node photos comprising panoramas matches survey id date

3.2.2 Inspect quality of node panoramas for each survey

- Survey panoramas should make sense from perspective of site compared to minimap position (e.g. begin at north)
- Consider excluding from survey Virtual Tour any images of poor quality

3.2.3 If necessary re-render the virtual tour after dropping or re-stitching any dubious panorama images.

3.2.4 Check that ‘surveys.csv’ contains id and date entries for all previously loaded survey tours plus ONLY the additional tours to be loaded on this occasion (i.e. exclude the data for surveys completed but not yet processed or included in current data transfer.

3.2.5 Survey content should show construction progression (in appropriate sequence) in comparison to earlier surveys (if not, re-check survey/photos/ for image sequence errors)