

## GIS 5571 Quiz 1 – Chris Carter

### 1. Define:

- Representation - the use of a sign or likeness as a stand in for something else
- Abstraction – a simplified view of an entity, with omits unimportant details to draw attention to details of greater importance
- Interface – the method, process, or system through which a user tells an abstraction to perform a task or function
- Data pipeline – a series of processes, tools or function that function in order to produce a predetermined output from an input dataset or datasets
- Interoperability – Allowing data to be integrated into a single dataset even if created by different organisations and formatted differently
- Quality assurance – processes or methods to help prevent errors from being introduced into the data
- Quality control – processes or tools to identify errors that are already in the data

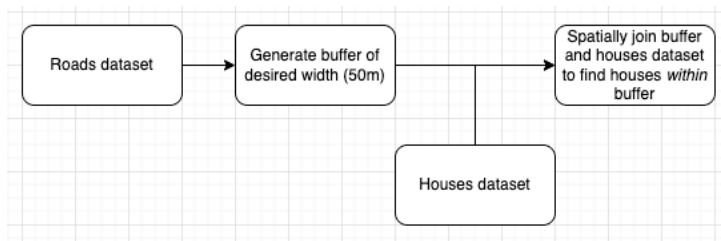
### 2. Computer systems rely heavily on abstractions and interfaces. Interfaces allow users to interact with computing systems without having to parse or understand the base mechanisms through which they work, while abstractions facilitate those interfaces by removing extraneous detail from the user's view

### 3. GIS abstractions:

- Raster – grid of values
- Vector – (series of) points with spatial relationships
- Network – topological relationships
- TINs
- Cubes – 3D raster data

### 4. ESRI implementations:

- Buffer vector data
- Spatial join two datasets



5.

### 6. The web inspector is a tool to see “inside” of a web page. It gives a user access to the source data that facilitates the display of a webpage (e.g. image sources, linked material). It allows users to inspect, debug and understand the inner workings of a webpage

### 7. import zipfile, io

```
quiz_data = requests.post(r'https://www.get-data.com/file.zip')
quiz_data_zip = zipfile.ZipFile(io.BytesIO(quiz_data))
quiz_data_zip.extractall(r'destination directory')
```

8. A Google Places API search link

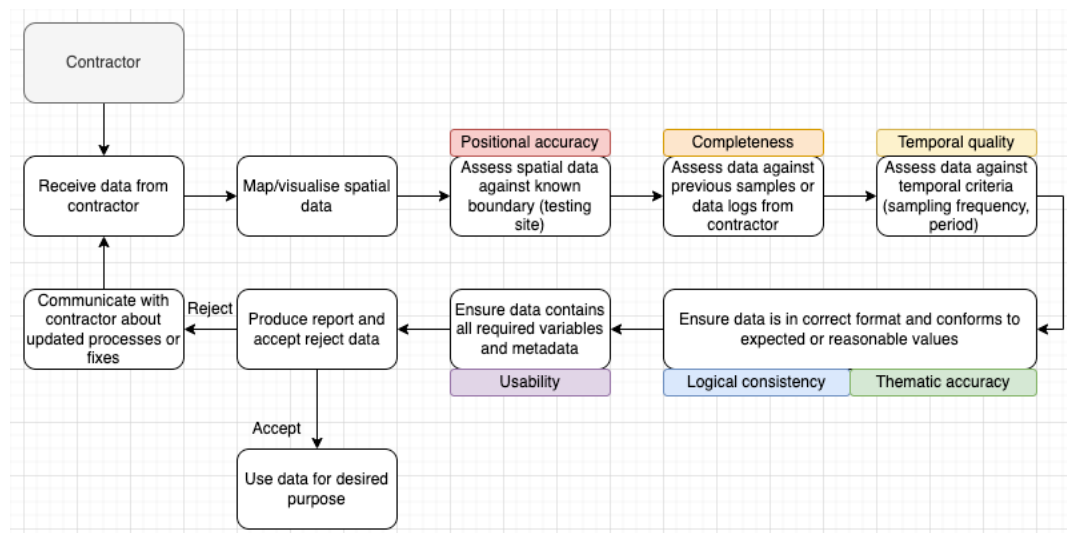
9. Batch or stream

- LiDAR dataset – batch-based pipeline style, as you have a large quantity of data to process relatively infrequently
- Washing machines – stream-based pipeline style, as you have many continuous instances of small datasets
- Census data – batch-based pipeline style, as you have a large quantity of data to process relatively infrequently (every ~10 years for census data)

10. Volume, velocity, veracity, variety

11. Data quality:

- Positional accuracy – the accuracy of features within a spatial reference system
- Completeness – the presence/absence of features and their attributes and relationships compared to what is expected or known to exist
- Temporal quality – the quality of temporal attributes and temporal relationship of features (do the data reflect the whole sampling period at regular intervals)
- Thematic accuracy – classification correctness related to features and their attributes (comparison to another source to ensure data is accurate to 'real life')
- Logical consistency – Adherence to logical rules of data structure, attribution, and relationships (checking for outliers, known ranges etc)
- Usability – the data adhering to the user requirements for its intended use



12.

13. I would envision Positional Accuracy, Temporal Quality and Usability being wholly or largely automated, with Completeness, Thematic Accuracy and Logical Consistency requiring more manual input or visual inspection. Sampling could be used to speed up the process on most steps, although Positional Accuracy should be checked for all spatial data.
14. Data quality is dependent on the use of the dataset. The relative importance of each of the above data quality elements is entirely related to the purpose of the dataset. While some are important always, like completeness or positional accuracy, thematic, temporal, and logical consistency are incredibly hard to define without knowing the way that data will be processed and used. Usability is literally the definition of comparing the data to the requested or desired purpose, which is definitionally linked to the desired purpose. So, while you could make a rough assessment of a dataset's quality without knowing the purpose (looking for obvious outliers, sample size, patterns of collection etc), it would be near-on impossible to define good data quality without knowing what the data would be used for.

#### Class Feedback

1. 3  
I'm not great at conceptual stuff, so some things haven't "clicked" just yet. I'm sure I'll get there; it might just take a little bit. Missing a class didn't help – the student notes are basically just regurgitations of the slides, and the lack of a recorded session really throws a wrench in things (no disrespect meant to my note-taking colleagues)
2. The lectures are engaging and interesting, and the group exercises are enjoyable. The feeling of achieving something in the labs is wonderful, if only because Python can feel uncooperative when you're still finding your feet
3. I wish there was a little more connection (or at least more obvious of a connection) between the labs and the lectures. Sometimes I'll struggle to understand why lecture content was useful or related to the lab at all – but this also may just be because my knowledge of Python isn't advanced enough to understand the inner workings of it all. And as I said, I'm not very good at conceptual thinking.

Also, a recording of the lectures would be super useful. I know it's another administrative/technological hurdle to surpass but it does feel almost impossible to replicate the learning experience of being in the classroom just through the slides and the notes. I actually enjoy the lectures largely *because* the slides are light on text, excellently supplemented by your/Bryan's engaging instruction, but that focus on in-person engagement ends up being a hinderance if you're trying to catch up (I hope this didn't come off to rude lol)