**Q-ITV: Enhancing Data Quality for Digital Twins**

**By NJJ Hendriks**

During a discussion with national stakeholders for digital twins it became clear that there is a need for a visual tool to quickly assess the quality of a dataset or connection and how it can be used within a digital twin. This tool should be placed directly at the source so that users know what they are receiving and the quality of the data validation behind it.

In both the scientific and commercial sectors, there are many examples of processes used to identify the quality of data and APIs. However, there is a lack of a clear 'stamp' that immediately shows the quality of the information, how often it is updated, the validation of the data, and the exact connection being made—a sort of ‘Energy label' label for connections. To address this, I have developed a model that addresses these points in the context of a QR code.

**Q-ITV**, or 'Quick-Information, Time, and Validation,' consists of three parameters to make the quality of information from a source visible and match it with the requester's needs. Additionally, it provides a gauge of the data quality itself, allowing for the pursuit of higher quality datasets.



**Parameters:**

1. **Interface:**
	* Describes the interface being called, making it clear what to expect for your digital environment. Given the many developments in interfaces, this is an open text field to include specifications or version numbers.
2. **Information:**
	* Indicates the level of information with four levels: Static -> Line & Points -> Object -> Dynamic. Static refers to an image, number, or text without interaction. Line & Points are geographical indications for use on a map. Object is a visualization of a model, and Dynamic is a combination of the previous levels or new information that does not fit the above categories.
3. **Time:**
	* Indicates the information's refresh rate with four levels: Irregular -> Monthly -> Daily -> Minutes. This parameter includes two considerations: the debate over Real-Time and Near Real-Time, with 'Minutes' used as an indication of fast refresh rates, and 'Irregular' indicating that the information is updated so infrequently that it cannot be assumed to be updated at a fixed time.
4. **Validation:**
	* Indicates the level of validation with four levels: None -> Quantitative -> Qualitative -> Standardized. This parameter distinguishes between experimental data and verified data. Initially, data from unknown sources is not validated but can become quantitatively validated over time through pattern recognition and the mass of results from previous measurements. Qualitative validation is based on comparable measurements from third parties, such as satellite measurements validated with a national institute data source/ research.

**Examples:**

|  |  |  |  |
| --- | --- | --- | --- |
| WCS |   |   |   |
|   |   |   |
|   |   |   |
|   |   |   |
| I | T | V |

***Ground Lever raster***

* **AHN4 – 0.5M Ground Level Raster:** Geographical raster data with highly variable measurement intervals (4 to 10 years), but fully quality-checked by the government with a complete description of how it was tested.

|  |  |  |  |
| --- | --- | --- | --- |
| B3DM |   |   |   |
|   |   |   |
|   |   |   |
|   |   |   |
| I | T | V |

***3D Mesh File***

* **3D Environment:** A photorealistic 3D mesh file created from Lidar measurements. The 3D model can be purchased when the calculations are complete and is not further validated.

|  |  |  |  |
| --- | --- | --- | --- |
| WFS |   |   |   |
|   |   |   |
|   |   |   |
|   |   |   |
| I | T | V |

***Traffic Data***

* **Traffic Data:** Traffic flow is measured every 15 minutes between two reference points on the road. The system has been running for several years and provides a good estimate of traffic density on various road segments.

**Next Steps:**

With Q-ITV, we can make an initial inventory of all public data sources and see what works well for use in digital twins. We also have a validation tool to get an overview of where we might want to invest resources to improve the quality of public datasets on the portal.