

Fixing the Transport Format

The PhUSE ET&T Alternative Transports Format Group

“A non extensible standard”

The SAS XPT v5 format was formalised as part of SAS v5 (in 1989)
A tabular based format with fixed-width columns
Limited to single byte characters
Limited to only Numeric and Char Types
.....

Criteria

Developed a range of technical and practical use cases

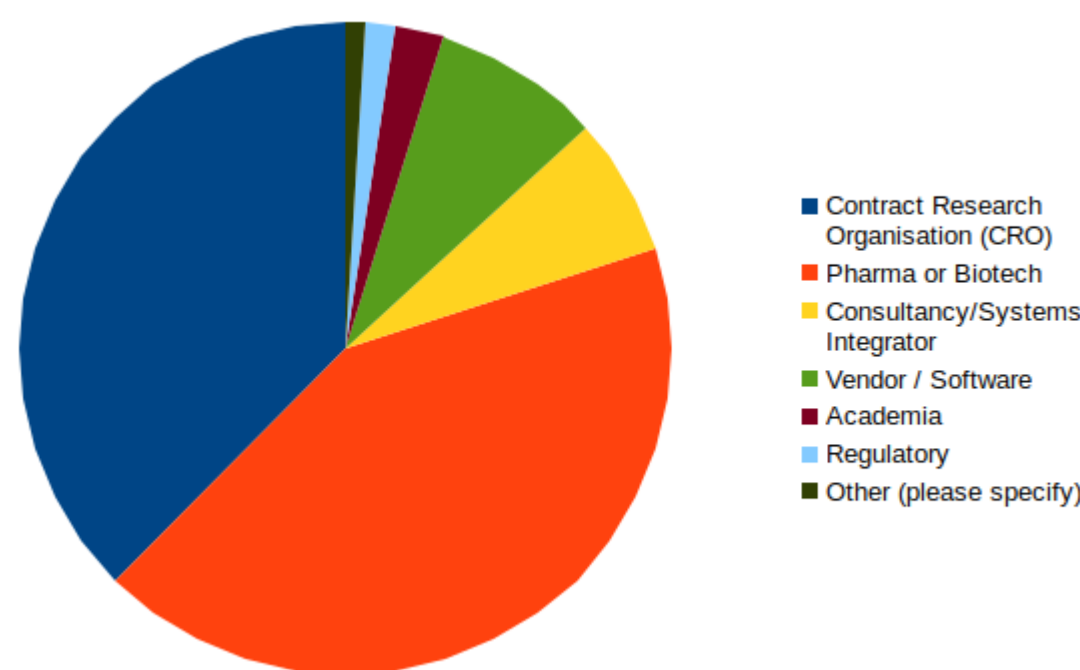
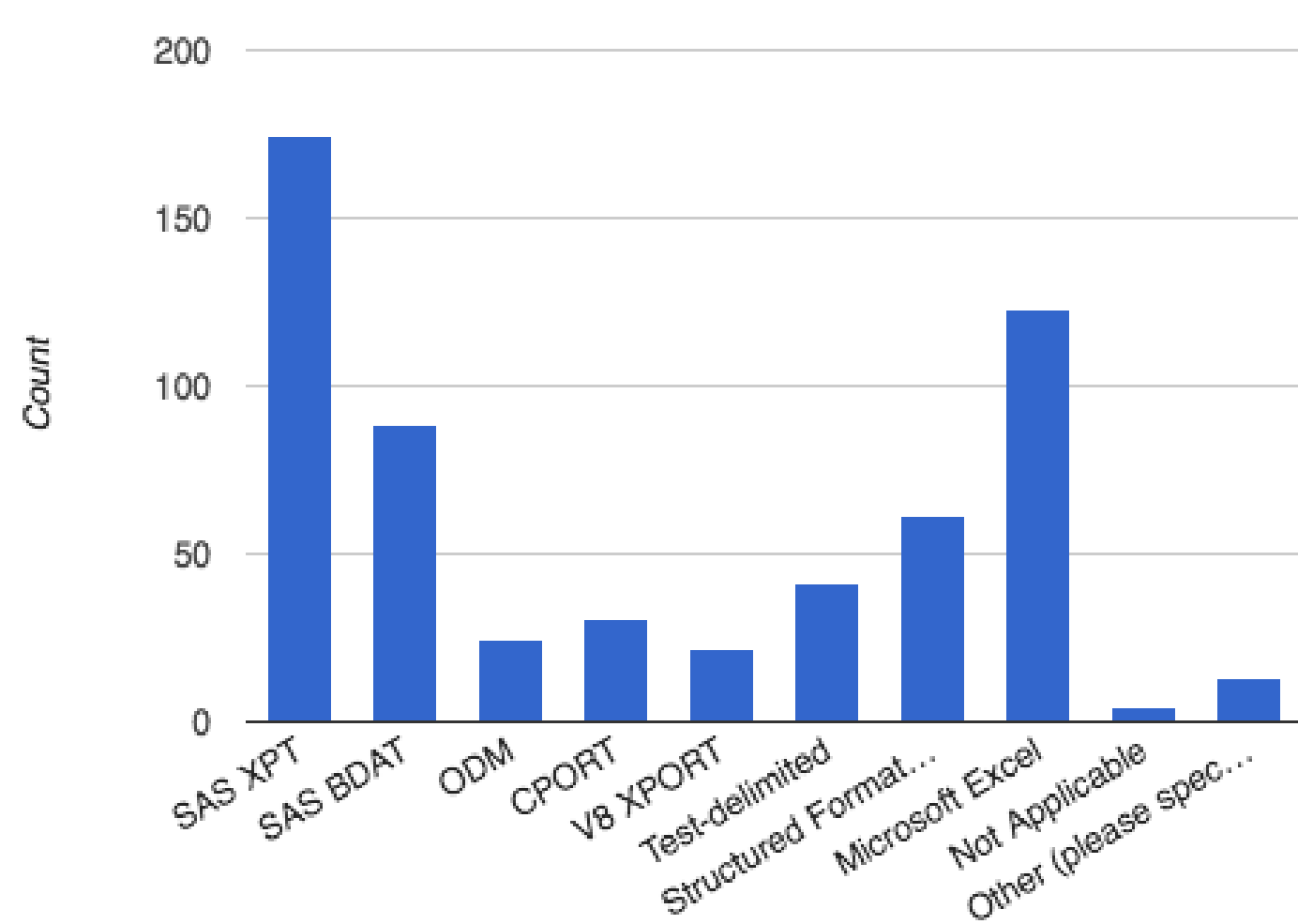
Categorised into three broad classes:
* **Implementation Characteristics**
* **Future Capabilities**
* **Process and Change**

Used Classes to be more prescriptive about responses recieved - some criteria crossed multiple classes

Data Collection

The Questionnaire was available for approximately five weeks after the Annual Conference.

There were around 210 responses to the questionnaire. The responses bore a good sample of the organisations in the industry:



Phase I

Identify important characteristics or criteria for a transport format for the future

Phase II

Evaluate different representations of the same dataset using the criteria

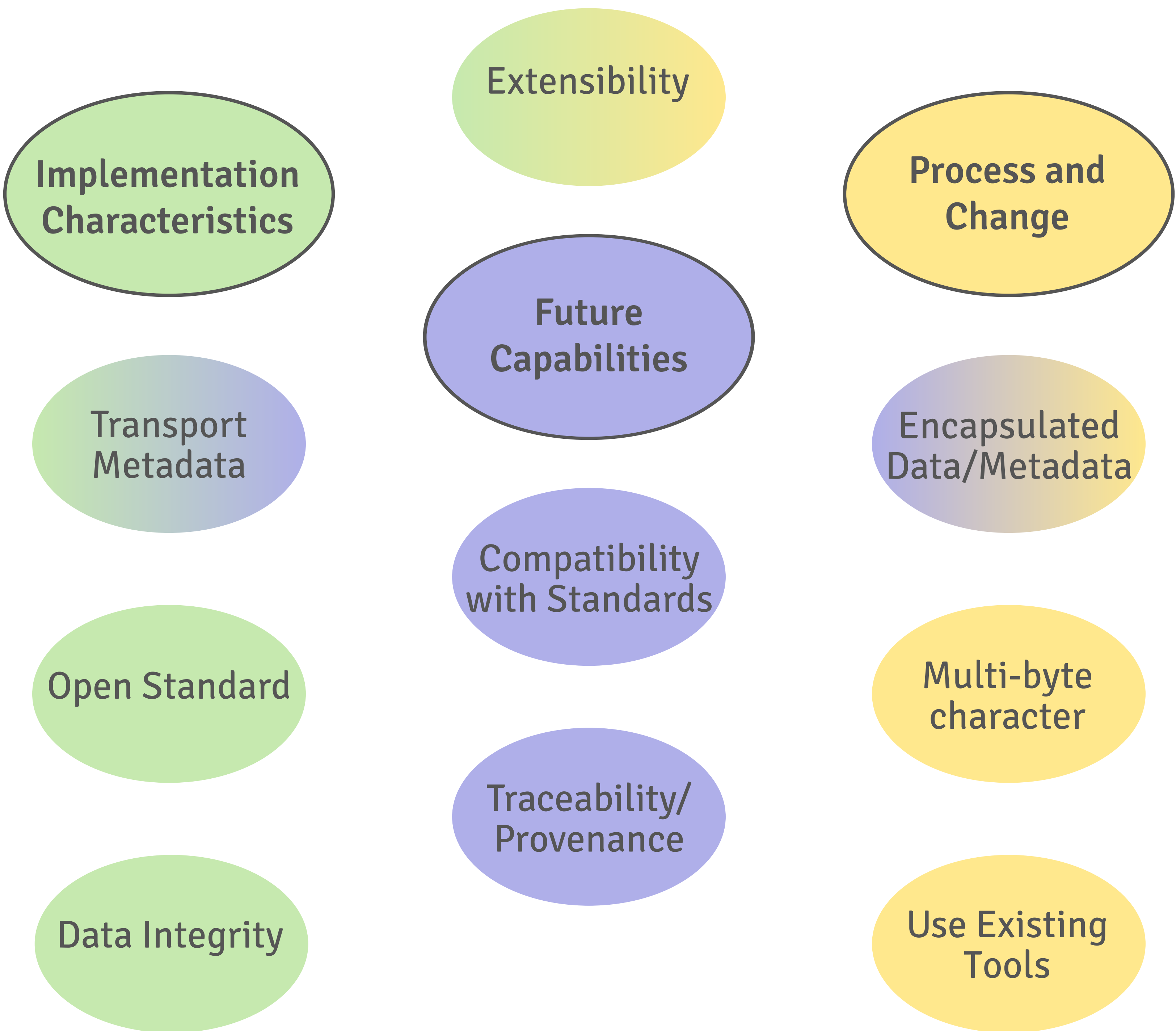
Methodology

Unable to get exact use cases against which previous candidates had been assessed

Elected to ape **Use Cases** by developing a set of criteria against which different representations could be assessed

Put Criteria out for rating by industry

Develop a **weighted index** of criteria reflecting the needs of the industry



Results

The top results for each of the classes are shown here; primary criteria include:
Encapsulating Data and Metadata (replacing define.xml)
Having the format be self-describing using Transport Metadata (e.g. allowing partitioning of datasets across transfers)
Having an extensible standard (eg supporting namespaces, etc)
Assisting with Traceability/Provenance
Supporting multi-byte characters

Next Steps

Phase II - use the weighted criteria to compare and contrast representations of standardised datasets

Put forward for formal evaluation

Questions

Many of the criteria identified could be addressed by changes to both transport and content models