

## Data Working Group 05 – Switching Programme: Data Management Design Principles

DRAFT FOR DISCUSSION

This document details the Switching Programme's Data Management Design Principles which should be utilised as a reference when assessing the design of end to end products produced within the Design, Build and Test phase of the Switching Programme. This document should be utilised as a reference to the individual developers of each Switching Data Service, the System Integrator and the Programme Coordinator.

These principles have been defined to ensure that the end to end design, which is predicated on the development of new services, such as the CSS; but also existing industry services, such as MPAS and UKLink; realises the policy objectives of Ofgem in the most efficient manner by constraining the scale of change and ensuring consistency of change related to data across the end to end architecture.

- a) Unambiguous requirements, constraints and business rules related to data management should be clearly defined and accessible to all Market Participants within controlled documentation, which may take the form of Documented Technical Specifications such as XML Schemas (XSD), data catalogues or descriptive business rules.
- b) Open Standards for the definition, management and messaging of data should be utilised where practically possible.
- c) Switching related data (data defined within a CSS interface documented in the REC Technical Specification Schedule), which has a direct mapping to existing industry data, as defined within the *UKLink Manual*, *electricity Data Transfer Catalogue* or *SPAA Data Flow Catalogues*, should be described with identical attributes (such as name (or reference to a synonym if alternate names or definitions are utilised), data type and field length) within the CSS physical Messaging Model.
- d) All data in existing systems at the point of the commencement of Transition Stage 1 will be in accordance with industry agreed standards and procedures, such as those defined within the MRA, SPAA and UNC; and therefore assumed to be fit for the purpose of CSS as defined in Design Baseline 4 (DB4).
- e) Data cleanse activity is being undertaken by industry during Transition Stage 0 to ensure that existing data conforms to the required current industry standards. It is expected that all data being migrated into the CSS has been cleansed by Data Masters to quality standards as required by (d) in advance of Transition Stage 1.
- f) Only those Registrable Measurement Points (RMPs) that are related to physical network connections and/or Energy Locations that exist or are planned should be set to a state that allows those RMPs to migrate to the CSS; thus preventing any erroneous migration of obsolete RMPs and the potential creation of invalid Retail Energy Locations by the CSS.
- g) The gas and electricity Switching Data Services (UKLink and MPAS), shall continue to master existing data (e.g. Metering Points, Agent Appointments) and shall not be required to change the Data Type Format of the data at rest within their services.

- h) Equally, for data which is changing Data Mastership from existing Switching Data Services to the CSS (e.g. Registrations) and which is still required to be provided to those services for existing end to end operations out of scope of switching (e.g. settlements); the CSS should present the data in a Data Type Format, via a CSS message, which is equal to the existing data industry standards as defined with the *UKLink Manual* and *Data Transfer Catalogue*.
- i) The design of applications, such as the Central Switching Service, should not be constrained by existing industry data for CSS Day-1 live operations. Where opportunities are identified pertaining to wider industry future proofing, efficiencies and market innovation, these should be considered for future CSS releases and evaluated through the Change Control process for adding to the CSS future roadmap of releases.
- j) However, to ensure that the delivery of the new switching arrangements scope (for Day-1 Live Operation) is focused to the realisation of Ofgem's related published policies, the scope of the future proofed design shall be limited to the CSS; this shall be accomplished by adhering to the constraints of existing industry Data Type Format standards via CSS messages which shall conform to new Data Exchange Formats and Data Exchange Semantics.
- k) A Roadmap of future switching and wider industry improvements which require switching data (as defined within the REC Technical Specification Schedule) should be developed so that these opportunities are visible and can be evaluated in a timely manner. This Roadmap should be controlled by the Switching Programme Coordinator until such point that the REC Code Manager is operational.
- l) Data provided over switching Interfaces should be constrained in scope so that only services which have a requirement to process specific data receive that data. In addition, certain data which is commercially confidential or is classified as personal sensitive should be controlled in a manner that is compliant with the security requirements of the new arrangements.
- m) Each data item will be assigned a *Meta Data Owner*, a *Data Master* and certain data items will be assigned one or more *Data Responsible Users*.

## Glossary of terms

**Message** – Is defined as an electronic communication between Market Participants as defined within the logical Message Model (Abacus) or once the physical design is complete the REC Technical Specification Schedule.

**Interface** - The shared boundary across which two or more separate services exchange information.

**Data Migration** - The process of transferring data between data storage systems, data formats or computer systems.

**Data Transformation** – The process of converting data from one Data Type Format or Data Exchange Format to another Data Type Format or Data Exchange Format.

**Data Cleanse** – The process of identifying incomplete, incorrect or inaccurate data and replacing, modifying or deleting such data.

**Open Standards** – Are technical specifications, controlled by a formal body (such as the World Wide Web Consortium “W3C” or the International Standards Organisation “ISO”), which either make those standards available for use without any conditions or on terms that are based on reasonable and non-discriminatory (RAND) contract. Some examples of implementation of open standards are: XML, JSON, ISO-11179 etc.

**Messaging Model** - The specification document which defines the Data Exchange Formats, Data Exchange Semantics and Data Type Formats related to Messages sent between Market Participants. The logical model is defined within the Detailed Design Repository (Abacus) and the physical design defined within the REC Technical Specification Schedule.

**Meta Data Owner** - The relevant Code Manager (e.g. the BSC, REC and UNC) responsible for the control of the meta data associated to each data item. Changes to the meta data is administered in a controlled manner via the Change Management Procedure associated to that Code, in conjunction with the REC Code Manager administration of the REC Data Catalogue within the REC Change Management Schedule.

**Data Master** - The Market Participant responsible for the stewardship of the data quality for that data item, leading on the cleansing of that data and in most cases the only party responsible for the creation and update of the data item value. The quality of data will be assured within the REC Performance Assurance Framework (PAF).

**Data Responsible User** - A Market Participant responsible for notifying the Data Master, on an ongoing basis, if they believe that the data quality for a data item can be improved or that the fitness for purpose of a data item has been compromised which is likely to result in a reduction of switching reliability. They are also required to support the Data Master in data cleansing activities. An example would be an Energy Supplier being the Data Responsible User for Premises Address data mastered by the Network Operator. Their performance in performing this role will be assured via the REC PAF.

**Data Exchange Format:** This refers to an interim data format for converting from one file or database structure ('data at rest') to another. Also called a "data interchange format". The source data is converted into the exchange format by one program, moved across ('data in motion'), and when received by the target, is converted (or transformed) to the target format by another program. XML is commonly used as a data exchange format. Other examples are: JSON, RAML, CSV, YAML, AXON etc.

**Data Exchange Semantic:** This refers to meaning of the data being transferred across various systems using a Data Exchange Format. The parsing logic (to convert/ transform the data) on both the source and target system should be sharing a common, shared vocabulary without which it will be impossible to have a meaningful exchange. For eg: The below is a set of data being passed using the XML Data Exchange Format:

```
<item>
  <title> Alice in wonderland </title>
  <note> Special Edition </note>
  <quantity> 1 </quantity>
  <price> 10.90 </price>
</item>
```

If the target system parser knows that this is 'XML' then it can do the syntax checks and detect syntactical issues. But that's only part of the communication. The source and target both need

to share a common semantic understanding of what is meant by the individual data items being passed. E.g.: Is the " " the quantity in stock or quantity ordered?

**Data Type Format:** This refers to the attributes of the data item's length and type i.e. whether a particular data item is a date, text, number etc. Also, it specifies the length of the text (character) or precision if a number. In the case of a date, it would mean how it's being expressed for e.g: 'DD-MM-YYYY', 'DD-MON-YYYY', 'DD-MM-YYYY HH24:MI:SS' and so on. Note that the Data Type Format can be either in the context of data in motion (i.e. used within a Data Exchange Format) or data at rest (i.e. a database repository of a source or target system).

DRAFT