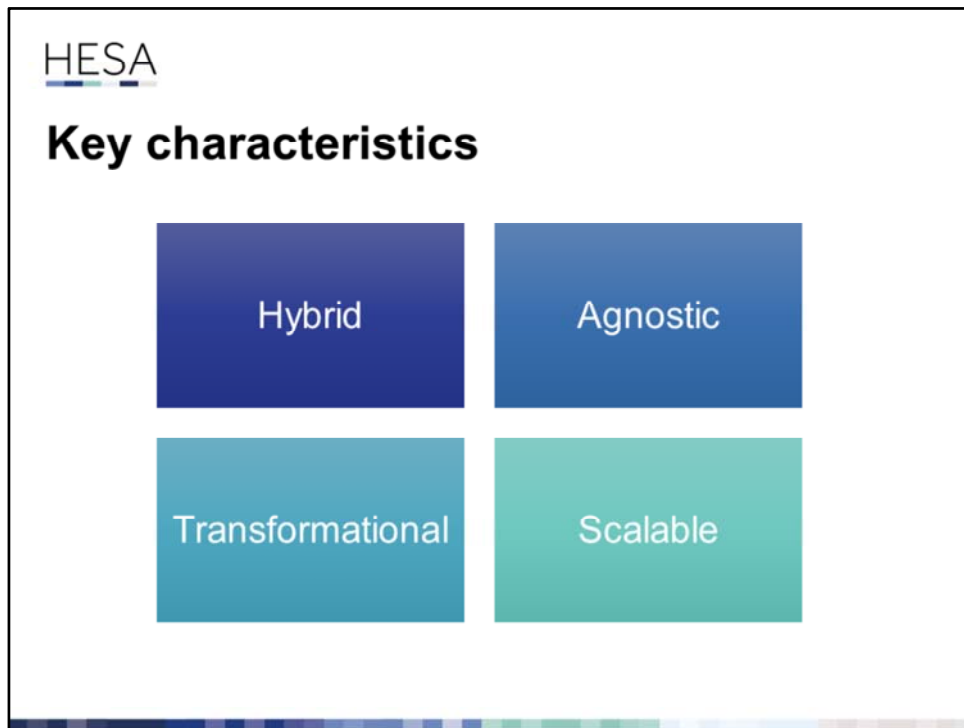The first round of workshops with HE providers has now taken place. Design phase V1 has now been presented.
We will be moving into a more detailed design phase in December 2016.
There will be further workshops in January 2017, to respond to the more detailed design.

When considering the future of data collection, we have defined four key characteristics to help us shape the design.

Hybrid - treats the demand (primarily statutory customers) and supply (HE Providers) as different customer groups. Demand side have requirements for the data collected by HESA either at an individual level or aggregated/shaped by HESA. Supply side needs to have a robust, high value approach to supplying the data for those requirements. This model respects those two groups and looks to provide specific services to both, while preserving what is liked about the current system.

Agnostic – While their will still be a dominance of Full Time Undergraduate provision for at least the next few years, it's important to design a collection system that supports the whole range of provision covering off non September start dates, part time, post grad, ITT and new offerings which are emerging in the HE landscape.

Transformational – Rather than just looking to satisfy the needs of the current requirements, the design must consider how collections might need to change over the next 10-20 years. This includes what services can be offered where the data is returned far closer to collection.

Flexible – Linked to the previous three points. Any collection design needs to be flexible

in both scale-up and scale-out dimensions. By this we mean, it has to have the potential to target any current or future requirements and be able to deal with an expansion in the number of providers and potentially of collectors.
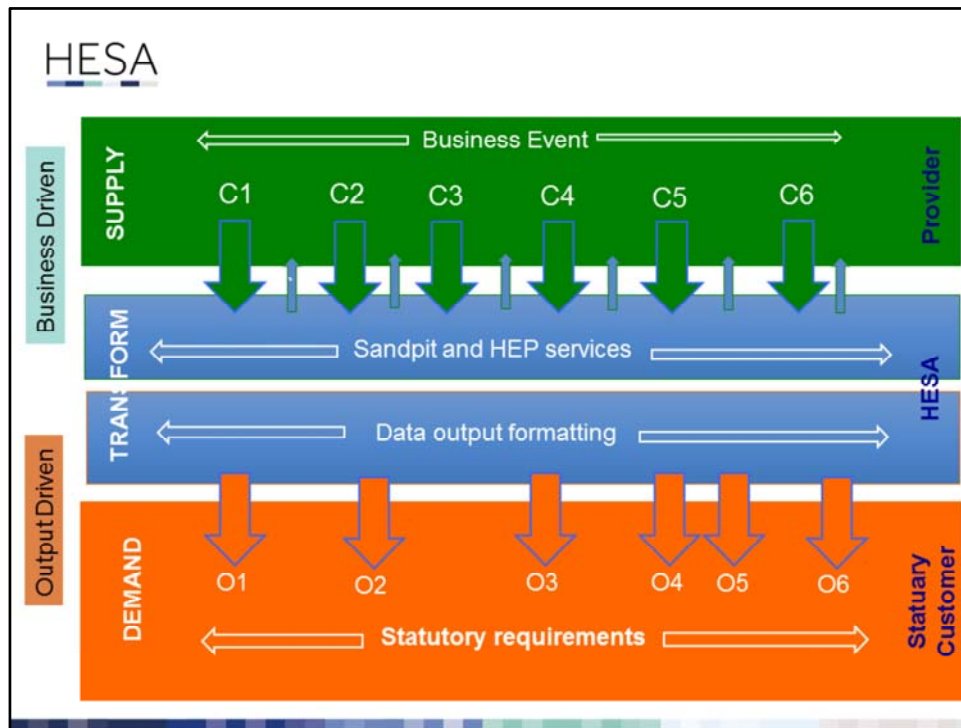
This diagram is best understood by considering each box in turn.

The providers are represented in the green box. Here we are proposing to define a number of collections (labelled here C1 to C6 but this is not a fixed number) that are tied to existing provider business events. The exact specification of these will follow later in the project but targets could include 'registration' , 'assessment', etc. Our current thinking is the collection will follow both the provider business process and the student journey in whatever academic provision they are under. Each of these events will return data to the top half of the blue box that is HESA. HESA will provide a number of value add services (to be defined but including early sight of potential issues for statutory submission). It's important to understand that the data in these collections is not specifically tied to a single output. And this data can and will change between collections. It is also important to confirm that the providers will still in control of when that data is released to statutory customers.

One way to consider this is that the provision of an individual collection can be thought of as provider operational data whereas when it is sent to the demand customer it can be considered sd statutory data.

This is where the second half of the blue box comes into play. We've purposely split HESA into two in this model to show how the hybrid model will work – so treating the

customer groups separately. For the statuary customers, we consider their requirements output driven represented by O1-O6 (again for explanation only, we are not fixing the outputs to only 6). Each of these outputs will be loosely coupled but not tightly linked to one or more of the collections. Therefore two rule sets will be available to the provider – one set for 'input / operational' and one set for 'output / statutory'. An example of an output might be the current HESES or SFC return.

HESA

Option 1: business driven

- No concept of a 'reporting year'
- Follows supply (provider) side business logic
- Based on administrative and academic events as collection triggers
- Collections are operationally discreet but statutory interdependent
- Not entirely 'natural' as will still require transform to HESA schema
- HESA provides many services for early sight data
  - QA 'sandpit'
  - Early sight MI
  - etc
- Current output deadlines do not change

Business driven does not mean every event recorded by the provider. Here we're talking about batching up data within known 'business' event. So as this existing event takes place, the provider 'extends' their business process to return the data to HESA to get it into the sandpit at an appropriate time. They can also return data at other times outside of these events.

Their 'time to return' will be based on what data is needed to meet a known output (e.g. HESES return mentioned before) but we want to leave flexibility for all types of provision, so there will be no fixed dates around collection.

This option would be aligned to reporting outputs. Each cut of the data would be sufficient in terms of coverage and quality to satisfy one or more reporting deadlines. As the academic year progressed more detail would be provided overwriting any previous data