We thank the reviewers for carefully reading our submission and providing very thoughtful comments. We address the reviewer comments below categorized into various issues.

Regarding the Threat/Trust Model *(Reviewer* 1 *and* 4*):* The aggregator needs to collect D_s points that closely match D_v in MMD distance. Therefore the aggregator at least sees $D_s + D_v$ points in this framework. This forms a natural lower bound on how many points the aggregator has to access. We define a ρ-parsimonious aggregator who sees ρ times the minimum required. On page number 3, we define this and our approach yields a $K^{1/3}$-parsimonious aggregator. For a data source i, all communication to the data source should be differentially private with respect to data points from all other data sources $\bigcup_{j \neq i} D_j$ - i.e. knowing all but one point in the union as side information must not reveal much about the missing point given the communication (standard informed adversary model with respect to union of other datasets). Preservation of differential privacy across data sources constrains the aggregator to collect more points than necessary. Hence, quality of the summary and parsimonious nature of the aggregator is traded-off with privacy requirements of the data sources (with respect to each other).

Discussion on Incentivization (*Reviewer 2*): The idea is that every data source would be able to monetize their contribution in proportion to the value they provide to the summary. After the protocol ends, value of a data source's contribution could be deemed proportional to the sum of winning marginal bids from the source. Value attribution based on this would be a incentive for data holders to participate. We will mention this. However, we only consider the privacy aspects and not the incentivization in this work.

Regarding the Trust in Aggregator (*Reviewer* 4): The aggregator needs to train a downstream task on a test distribution that is similar to D_v. That is why D_s points (much larger set) are being collected for training. In fact you could think of aggregator paying for the D_s points. So the best set of points (up to the approximation guarantee in our algorithm) that the aggregator could have is through this protocol. There is no incentive for aggregator to cheat since it has to pay for the points it collects. The data providers are happy to provide a set as long as they are compensated and other data sources do not know about their data (in a differential privacy sense.) We will make this point clear in our draft.

Regarding Federated Learning vs Our Approach: (*Reviewer* 4) This is a good point – in fact, we do point out the distinction of our approach vs federated learning in the experimental section (lines 289-295). We will add a note in the related work section as well. Indeed, here we deal with a transfer learning problem. The validation set distribution is distinct from each of the individual data source distributions. We are focused on a setting involving significant covariate shifts between validation set and training data sources. In fact in our experiments, uniform sampling (which is a proxy for gradient updates in a federated style algorithm that works by uniformly sampling points across datasets) has a poorer performance compared to our method. We are collecting points that closely resemble the validation data. Federated learning would assume a training distribution that is typically uniform or mixed in a specific ratio using different sources. Transfer learning component or the distirbutional shift between D_v and each individual D_i makes the problem non-trivial.

Regarding Rahimi&Recht and Missing Citations (*Reviewer* 4) : We do cite Sarwate's paper (Chaudhuri, Monteleoni and Sarwate, 2011) – citation number 3, on line number 79. However, we miss citing Rubinstein's paper, which we will rectify. Although, Rahimi-Recht's method has been used in privacy before, we use it in a novel way in combination with MWEM and a private auctioning protocol, to solve the distributed data summarization problem under covariate shifts and differential privacy constraints.

Differentiation between our Private Auction and the Exponential Mechanism (*Reviewer* 4): In Step 6, each data source selects its "local marginally best point" and submits its value as a bid to the aggregator. It is not necessary that such a point will be chosen by the aggregator. In our private auction, if a point is chosen by a data source τ times the source would submit the point to the aggregator to be included into the final summary (it is easy for the aggregator to verify that a point has been chosen τ times using the corresponding bids submitted by the data source). Although, there is a superficial resemblance with the exponential mechanism, our private auction is significantly different. First note that there is no "loss" in value of the best point chosen in our mechanism, indeed the probability of choosing the best point is $e^{-\epsilon_{auc}(1-1)} = 1$. The second key difference is in the way we prove the guarantees in our mechanism. While the exponential mechanism selects one approximately "best" point, we flip a coin for every bid whose bias has an exponential decreasing relationship to the position of the bid in sorting order. Then, we choose multiple of them (instead of one) and a key proof point is to show that we can restrict the number of the points chosen overall. In fact the bias probabilities do not even depend on the bid value (i.e., "score") while it would be the case for exponential mechanism. We were short on space and could not add the insights on Thm 2 but we will do it in the camera ready if accepted.

Explanation of the Protocol (*All Reviewers*): We will add explanations for various steps of our protocol to make different parts clear. We will also address all the grammatical and typographical errors in our submission.