**PROBLEM STATEMENT**

Current data integration implies consuming data from different data services and integrating the results while meeting users’ quality requirements. Such requirements include the data that is retrieved and integrated, but also the properties of the data, its producers and the conditions in which such data is produced and processed. For example, whether the user accepts to pay for data, its provenance, veracity and freshness and how much is the user ready to pay for the resources necessary for integrating her expected result. Data services provide data according to specific APIs that specify method headers with input parameters describing the data to be retrieved and the type of results they can produce. Moreover data provision can be done by services according to different data quality measures. Such measures describe the conditions in which a service can provide or process data. These measures can be expressed in a service level agreement (SLA). An SLA states, what the user can expect from a service or system behaviour. For example, whether it implements an authentication process, if it respects data consumer’s privacy and the quality of the data the service can deliver, like freshness, veracity, reputation and other non-functional conditions like the business model that controls data delivery.

Data provision and data processing services may need a considerable amount of storage, memory and computing capacity that can be provided by cloud architectures. Furthermore, users could need to integrate the data provisioned by services in a homogeneous and general result. The integration process can also require important computing resources that can be obtained from the cloud too. Data provision and processing services can be deployed in the cloud. Their SLA includes the measures about the cloud services that they require to execute their requests. The cloud, itself exports a general SLA that specifies the conditions in which users can access the services (infrastructure, platform and software) deployed in it. A user willing to use the cloud services establishes a contract with the cloud provider guided by an economic model that defines the services she can access, the conditions in which they can be accessed (duplication, geographical location) and their associated cost. Different cloud providers have different possible contracts to establish with users (i.e., platinum, silver, gold, ivory users). Thus, for a given requirement, a user could decide which cloud services (from one or several cloud providers) to use for retrieving, processing and integrating data according to the type of contracts she can establish with them.

In consequence, data consumption is determined by quality constraints specified by the user (data consumer) and different contracts (i.e., SLA’s) of the clouds providing the required services. User’s constraints define the storage and computing capacity of the device that consumes the data, the data transmission bandwidth and cost, whether the data consumption is critical (time constraints) and energy consumption. The SLA of the cloud determines the type of quality a user can expect from its services, according to the contract (subscription) signed between her and the cloud provider. The user profile, her quality requirements and her execution context determine the conditions in which she is expecting to consume data by using such cloud services.

Thus, **the first challenge** is to compute what we call an *integrated SL*A that matches the user’s integration preferences (including quality constraints and data requirements) with the SLA’s provided by cloud services, given a specific user cloud subscription. The user may have general preferences depending on the context she wants to integrate her data such as economic cost, bandwidth limit, free services, and storage and processing limits. The SLA’s associated to the cloud services can be of different types: user - data service, data service - cloud provider, data provision service - data processing service, and cloud provider - cloud provider. In this context, matching the user integration preferences with the services that can contribute to produce a result can lead to search and identify in the chain of SLAs. Probably it is possible to find an incompatibility between the preferences and a SLA in the chain, in this case it is necessary to propose a strategy to solve the problem.

Furthermore, in order to fulfill requirements and satisfy user expectations, it is possible to have a collaboration between different clouds. This collaboration implies the agreement through SLAs between services deployed in different cloud providers. In consequence, matching user preferences can deal with SLA’s can lead to deal with heterogeneous SLA specifications (different schemata, different measures semantics and granularities). Computing an *integrated SLA* can imply dealing with heterogeneous SLA specifications and SLA-preferences incompatibilities.

The **second challenge** is to guide data integration taking into consideration the *integrated SLA*. Here, the data integration process includes looking for services that can be used as data providers, and for services required in order *(i)* to retrieve the data; *(ii)* to build an integrated result; and *(iii)* to deliver it to the user considering the user quality requirements, her context and resources consumption.