ESAP Auth options for storage access.

Stefano's slides:
https://docs.google.com/presentation/d/e/2PACX-1vSeuOB0rFbe2jy3eKM1EuX6z4h89VVs5uHOzwRAWpBrrbindyydBfmg2hqMelBgIUggPC9SMox4lkn4/pub

What is data ?
Stefano : Working assumption, "data is one or more files somewhere".

Ian: Reason for not choosing 4 over 2 is tracebility of who did what. Need some kind of logging on the server side to trace who did what.
Stefano : In both 3 and 4 platform can still provide asomething to say who the users is. In all cases, when data is sent to the task, we always rely on the task to control what happens to the data. We have to trust the platform because the platform can send user x data to user y.

Matthias : The external token for case 3 is from IAM ? In 3 the storage can verify the user credentials, why can't we apply the same checks in the platform.
Stefano Once the data is sent to the task, we have to trust the task what it does with it. With 3 and 4, it is possible for the platform to send the wrong data to the wrong task. In this regard, 1 is always the safest form, because the task has to identity itself.

Pierre: Use case is access to data owned by the user, which has been solved by OATH2. Where does this fit in the options ?
Stefano : Yes, OATH2 is equivalent to option (3).

John: Always assumed we are in option (1). If I were a storage provider I would be nervous about options (2), (3) and (4). Is there a way we can acheive this using some kind of wrapper.

Ian: Part of the problem with (1) is the integration of different platforms and storage services. With many cloud compute in distributed environment, and things like Rucio, the data access is hidden from you by thingslike (2), (3) and (4).
Ian : A task accessing th DataLake, access is mediated using tokens issued by IAM, equivalent to (2) or (3). Many platforms do this now using certificates and are moving to tokens.
Ian: reason for this is storage platforms wanted to see the identity of the user (e.g. certificate).

Problem with 2 is that the user gives the platform their certificate or token. So the platform can impersonate that user and do other things. Stefano note: not token, actually, which is public.. just certifcate. Which is instead secret and where problems arise.

Ian: In delegated certificates means only the user's data is at risk, because actions are linked to their certificate. The storage system itself is not at risk.

Hugh: ESAP is currently option(1). Problem is how to make (1) less clunky. Copying token into a notebook is clunky. Work in progress to put the token into user environment ? Current state of the art is that we have a mixture of x509 and tokens for different storage systems.

Ian: Goal is to replace x509 completly. Reason is entirely historical. Issue is several storage platforms haven't completed the work yet.

Hugh: If IAM is the identity provider, it is separate.

Gareth: In work package 2 RES are being chased to make sure they accespt tokens rather than certificates. Dirac currently only accepts certificates. Wor around is to accept certificates from RCAuth project. Goal is that everything will accept tokens.

Matthias : User belong to groups, e.g. cta-users During the lifetime of the project a user might jion and leave different groups. Same user, different groups over time. Confirmation does this always work in all of these options ?
Stefano Yes. If you do use authentication you are always free to do this kind of mapping.
In (1) and (2) the storage has to handle the mappings.

Matthias : Group membership is external to this.

Hugh: The plan (for ESCAPE) was for IAM to mange the group assignments.