Évolution de la Fédération d'identités

Geoffroy Arnoud

Évolution de la Fédération d'identités

Geoffroy ARNOUD
GIP RENATER
1A Avenue de Belle Fontaine
35576 Cesson-Sévigné Cedex

Résumé

La Fédération Éducation-Recherche met en relation des organismes proposant des services avec des fournisseurs d'identités. Elle permet à la fois de réduire le nombre de mots de passe à connaître et de mieux maîtriser la diffusion des données à caractère personnel.

Lors des JRES 2019, nous avons exposé nos objectifs pour améliorer la qualité de la Fédération Éducation-Recherche.

Dans cette présentation, nous exposerons le chemin parcouru et les évolutions à venir.

Le fonctionnement de la Fédération Éducation-Recherche repose principalement sur deux outils développés par RENATER :

- *le Guichet de la Fédération – outil principal à disposition de la communauté ;*
- *la Supervision Métier de la Fédération qui permet de vérifier le respect du cadre technique, et le démarrage d'un dialogue avec nos membres pour la résolution des problèmes détectés.*

La mise à jour du Guichet de la Fédération a permis de faciliter notre travail d'opérateur. En effet, les saisies étant effectuées par chaque membre, un guichet plus fiable génère moins de support et une meilleure expérience utilisateur.

Nous avons également ajouté d'autres fonctionnalités :

- *support du Single-Logout ;*
- *conformité R&S ;*
- *renforcement des fonctionnalités de fédérations locales.*

La Supervision Métier de la Fédération a permis de réduire fortement le nombre de problèmes constatés, au prix d'une certaine rigidité vis-à-vis de notre communauté.

Ce travail nous amène à faire évoluer nos outils pour mieux accompagner ces usages, problématiques à l'échelle de la Fédération Éducation-Recherche, mais tout à fait anodins ou légitimes du point de vue d'un organisme.

Mots-clefs

Fédération, Fédération Éducation-Recherche, SAML, MDQ, métadonnées

1 Introduction

Une fédération d'identité est un cercle de confiance, qui permet d'étendre la délégation d'authentification en dehors d'une organisme.

Celle-ci ne peut vraiment fonctionner que si des « règles du jeu » existent, et que leur respect est vérifié.

Le rôle de RENATER en tant qu'opérateur de fédération, est de permettre l'interopérabilité entre les briques techniques en fournissant un certain nombre d'outils, et en mettant à disposition des métadonnées SAML, dans lesquelles les participants peuvent avoir confiance.

Lors des JRES 2019, un article sur l'amélioration de la confiance dans la Fédération Éducation-Recherche a été présenté [1]. Notre volonté alors était de sensibiliser la communauté sur l'importance de maintenir la fédération dans un état de salubrité minimum sur les problématiques de :

- sécurité ;
- fiabilité des données ;
- interopérabilité.

Pour ce faire, un certains nombre d'objectifs ont été fixés, notamment :

- mise en œuvre d'une supervision proactive ;
- application plus stricte des règles de nos cadres contractuel et réglementaire – ainsi que leur mise à jour ;
- refonte du guichet de la fédération.

Dans cet article, nous verrons le travail effectué sur les outils jusqu'en 2021, les évolutions prévues pour 2022, ainsi qu'un focus plus particulier sur la mécanique de diffusion des métadonnées.

Petit point de terminologie pour cet article :

- Le terme « Fédération d'Identité » fait référence au service rendu par RENATER pour la mise en œuvre de plusieurs fédérations SAML. Ce service inclut un certains nombre d'outils (guichet, outils de test, supervision…) ;
- Le terme « Fédération Éducation-Recherche » fait référence à la fédération SAML nationale opérée par RENATER ;
- Les termes « IdP » et « SP » sont des acronymes signifiant respectivement « *Identity Provider* » et « *Service Provider* » (fournisseur d'identité et fournisseur de service). Ce sont les deux types d'entité SAML présents dans les fédérations SAML. Nous utiliserons les acronymes IdP et SP dans cet article.

2 Le chemin parcouru jusqu'en 2021

2.1 Refonte du Guichet de la Fédération

La refonte annoncée du Guichet de la Fédération a eu lieu. Le code de celui-ci est à présent disponible publiquement sur SourceSup, sous licence AGPL [2].

Cette refonte a porté le numéro de version 3.0. Celle-ci a été mise en service en février 2021, et a été suivie par d'autres tout au long de l'année 2021 :

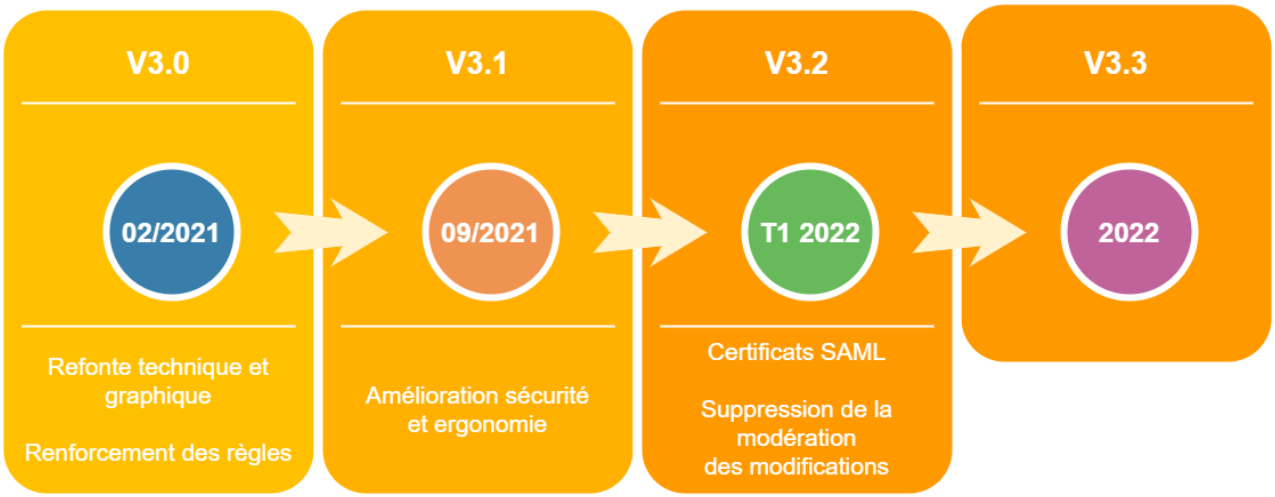

Figure 1 – Versions du Guichet de la Fédération

Pour le détail des améliorations des versions voir [3][1].

En voici une liste non exhaustive :

- **Refonte graphique et technique :** cette refonte nous permet de repartir sur une base de code plus propre et d'envisager de fournir des évolutions plus fréquemment ;
- **Règles de pré-enregistrement dans une fédération :** les demandes d'enregistrement dans une fédération sont soumises à des règles qui sont à présent paramétrables de manière différenciée pour chaque fédération – au lieu d'être liées au code source. Nous reviendrons plus en détail sur cette fonctionnalité plus loin ;
- **Amélioration de la sécurité :** la reconnaissance et l'autorisation des utilisateurs est à présent basée sur l'attribut *eduPersonPrincipalName* (ePPN) et non plus sur l'adresse email – qui devient elle un identifiant **transitoire** pour les nouveaux utilisateurs. L'utilisation de l'ePPN nous permet d'être en phase avec nos propres préconisations [4] ;
- **Certificats SAML** : depuis son origine, le Guichet de la Fédération ne permettait de déclarer qu'un seul certificat SAML par entité, alors qu'un SP ou un IdP Shibboleth en crée systématiquement deux lors de son installation. Cette limite pose des problèmes à la mise en place du Single Logout à l'échelle de la Fédération Éducation-Recherche. Nous reviendrons plus loin sur cette évolution ;
- **Suppression de la modération des modifications sensibles :** Au niveau des entités SAML, certaines données sont considérées comme sensibles. Jusqu'ici, leur modification était soumise à la modération de la part d'un responsable de l'organisme. Nous avons fait le choix de retirer cette modération pour des raisons que nous expliquerons dans la suite de cet article.

2.1.1 Gestion des fédérations

Le Guichet de la Fédération permet d'opérer une vingtaine de fédérations, certaines publiques (fédération de test, Fédération Éducation-Recherche, eduGAIN) et d'autres privées (réservées à certains organismes ou à des groupes d'organismes partageant des services).

1 La version 3.2 n'ayant pas encore été mise en service au moment de la rédaction de cet article, les notes de cette version peuvent ne pas être encore disponibles.

La mise en place d'une fédération privée (également appelée « fédération locale ») est ouverte à tous les organismes de la fédération [5].

Actuellement, presque un tiers des IdP sont inscrits dans une fédération locale (en plus de la Fédération Éducation-Recherche et eduGAIN), et l'ensemble des fédérations locales permettent d'interconnecter plus de 300 services (contre environ 1200 services dans la Fédération Éducation-Recherche).

Cette hétérogénéité de cas d'usage nous a conduit à ajouter des contrôles de pré-enregistrement d'une entité SAML dans une fédération. Par exemple, imposer une description en anglais, pour un service destiné à un publique francophone a peu d'intérêt. Par contre ceci est obligatoire pour pouvoir s'enregistrer dans l'inter-fédération eduGAIN.

Les contrôles peuvent porter sur toutes les informations disponibles concernant une entité. Les prédicats applicables variant en fonction du type de données :

- **Prédicat classique :** est égal, n'est pas égal, expression régulière ;
- **Prédicat sur les URLs :** publiquement joignable, publiquement résolvable ;
- **Prédicat vérifiant l'unicité** dans la fédération demandée.

2.1.2 Certificats SAML

Les échanges SAML peuvent être signés et chiffrés. Comme dans tout échange impliquant de la signature et du chiffrement, l'émetteur :

- chiffre le message avec la clé publique du destinataire, afin de garantir qu'il sera le seul à pouvoir lire le message ;
- signe le message avec sa propre clé privée, afin que destinataire s'assure de l'identité de l'émetteur du message.

Le processus d'installation classique d'un IdP ou d'un SP Shibboleth génère deux paires de clés distinctes. La configuration par défaut en affecte une à un usage de signature et l'autre à un usage de chiffrement.

Le Guichet de la Fédération ne gérant jusqu'à aujourd'hui qu'un seul certificat[2] par entité, les métadonnées publiées par RENATER exposent un unique certificat, pouvant être utilisé pour la vérification de signature et le chiffrement des messages (le type d'usage d'un certificat est spécifié dans les métadonnées).

Ce fonctionnement n'impacte pas le processus de connexion décrit en annexe.

Il n'empêche pas non plus le processus de déconnexion (décrit lui aussi en annexe) de fonctionner **à condition** qu'IdP et SP soient configurés pour tenir compte de la limite d'utiliser une unique paire de clés pour la signature et le chiffrement.

Cependant, si l'une des entités (SP ou IdP) a une configuration « par défaut »[3], la clé privée utilisée par une entité pour signer ne correspondra pas au certificat publié dans les métadonnées, utilisé par l'autre entité pour vérifier la signature.

Et réciproquement pour le chiffrement : le certificat publié dans les métadonnées pour chiffrer ne correspondra pas à la clé privée utilisée par l'entité recevant les données pour déchiffrer.

2 Pour des raisons historiques, les clés publiques des entités sont encapsulées dans des certificats x509, publiés dans les métadonnées SAML

3 On entend « par défaut », une configuration issue de l'installation basique d'un SP ou d'un IdP, pour lequel la configuration des clés n'a pas été modifiée.

C’est pour cette raison qu’après avoir ajouté la possibilité de déclarer les points d’accès de *Logout* pour les IdP et les SP dans la version 3.0, nous avons ajouté la possibilité de déclarer des certificats distincts pour des usages distincts dans la version 3.2[4].

Nous en profitons au passage pour améliorer deux fonctionnalités connexes :

- La saisie assistée des informations techniques, grâce aux métadonnées de l’entité : celle-ci va pouvoir prendre en compte les deux certificats distincts ;
- La procédure de renouvellement des certificats SAML est simplifiée, en laissant le Guichet de la Fédération gérer la suppression de l’ancien certificat, une fois que le délai de propagation des métadonnées est écoulé.

2.1.3 Suppression de la modération des données sensibles

Jusque-là, certaines informations étaient jugées sensibles. Leur modification était soumise à modération par les responsables de l’organisme, soit pour limiter les risques d’erreur pouvant mener à une interruption d’accès à un service, soit pour des raisons de sécurité. Il s’agit des informations suivantes :

- l’*entity ID* ;
- les attributs demandés par un SP ;
- les *scopes* d’un IdP.

Avec du recul, cette étape de modération apporte une grande complexité technique dans le Guichet de la Fédération, pour un gain très faible côté communauté, et mène parfois à des confusions.

C’est pourquoi cette modération est supprimée dans la version 3.2, et remplacée par :

- **l’immutabilité de l’*entity ID* :** si un changement d’*entity ID* est nécessaire, il est plus simple pour l’utilisateur de créer une nouvelle entité (avec les mêmes points d’accès le cas échéant). En effet, les deux définitions peuvent ainsi cohabiter dans les métadonnées, et permettre une bascule en douceur de l’un à l’autre ;
- des contrôles et des vérifications proactifs – réalisés hors du Guichet de la Fédération :
  - à l’instar de ce qui est fait sur TCS, vérifier qu’un *scope* déclaré sur un IdP appartient bien à l’organisme ;
  - échanger avec les responsables de services demandant un nombre important d’attributs afin de valider que le besoin est pertinent.

2.2 Supervision proactive

2.2.1 Présentation

Le but de la supervision proactive est de vérifier le respect des règles et du cadre d’usage de la Fédération Éducation-Recherche [6]. Bien que la supervision implémente plusieurs contrôles, les deux principales erreurs que nous avons voulu corriger en priorité étaient :

1. les entités inscrites à la fois dans la Fédération Éducation-Recherche et dans la fédération de test. Le Guichet de la Fédération n’autorisant plus ces inscriptions en double, il s’agit de résorber cette dette technique ;
2. les entités non joignables : services en adressage privé ou protégés par un pare-feu (et donc pas ouverts à l’ensemble de la communauté ESR) ;

4 À l’heure de la rédaction de cet article, la version 3.2 est encore en cours de réalisation.

Notre système nous permet :

- d'automatiser la création de notifications à destination des organismes. La décision d'envoyer ou non reste manuelle, afin de conserver un nombre de « tickets ouverts » raisonnable, nous permettant de traiter les réponses au fil de l'eau ;
- de consulter des tableaux de bord, afin de cibler les services devant être notifiés en priorité.

En cas d'absence de réponse, nous relançons systématiquement au bout de 15 jours. La deuxième relance s'accompagne d'une menace de désinscription de l'entité de la Fédération Éducation-Recherche, faute de retour. Jusqu'à présent, les seuls cas où cet ultimatum a été suivi d'effet correspondent à des services dont plus aucun contact n'était valide, ou avaient complètement manqué les avertissements précédents. En effet, notre but n'est pas d'exclure, mais de s'assurer de l'atteinte des objectifs suivants :

- pour les cas simples, obtenir une résolution dans un délai raisonnable ;
- pour les cas complexes, pouvoir dialoguer avec les administrateurs, afin de trouver une solution acceptable pour tous.

Cela nous permet aussi de détecter des services qui ont été décommissionés, mais dont les informations sont restées dans le Guichet de la Fédération.

2.2.2 Résultats obtenus

La Figure 2 compare le pourcentage d'entités inscrites uniquement dans la Fédération Éducation-Recherche et celles inscrites également dans la fédération de test.

Pour le second problème, nous sommes parti d'une situation encore une fois très décevante concernant l'utilisation de la Fédération Éducation-Recherche, puisqu'un nombre non négligeable de services utilisait cette fédération comme simple moyen d'ajouter un site interne dans le SSO de l'organisme. La Figure 3 montre l'évolution du nombre de SP inscrits dans la Fédération Éducation-Recherche non joignables pour des raisons de routage (nom d'hôte non résolvable ou IP privée). Cela n'inclut pas les SP disposant d'une IP publique, mais placés derrière un pare-feu empêchant toute connexion provenant de l'extérieur de l'organisme.

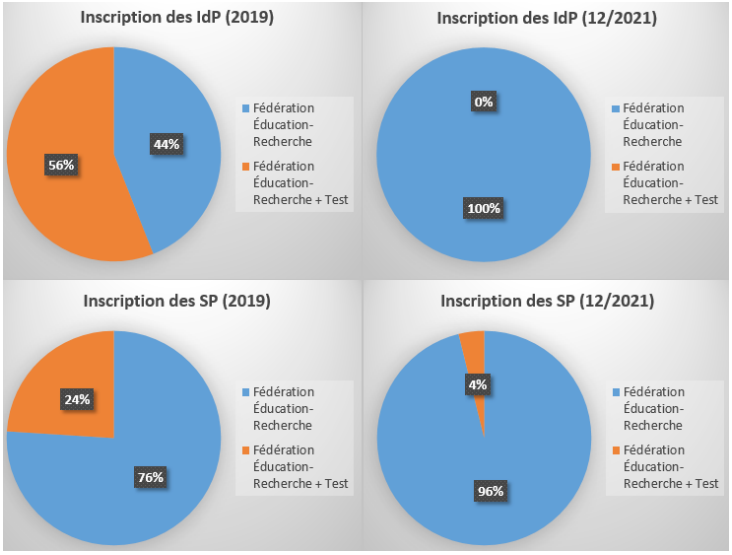

Figure 2 – Comparaison des doubles inscriptions

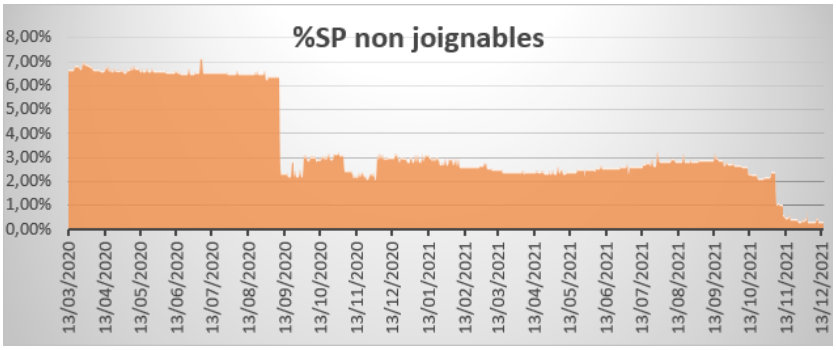

Figure 3 – Ratio de SP non joignables (IP privée ou hôte non résolvable)

Quand on analyse les Figures 2 et 3, on peut se dire que des bons résultats ont été obtenus sur ces deux non-conformités, mais ce n'est pas le cas. Autant la correction des doubles inscriptions a été plutôt efficace, avec une prise en compte de nos demandes dans un délai souvent correct de la part des administrateurs, autant les résultats concernant le second problème sont en demi-teinte.

En effet, les deux chutes dans le graphique sont artificielles :

- En septembre 2020, nous avons exclu les portails captifs Eduspot de ce contrôle, car « par construction », ils utilisent une IP privée, et ne doivent être joignables que depuis le réseau interne ;
- Nous avons identifié qu'il n'était pas efficace de solliciter individuellement les administrateurs de service dépendant de grosses structures. Nous leur avons donc remonté les problèmes globalement, et mis la supervision en sourdine pour ces services (novembre 2021).

3 Évolution du Guichet de la Fédération en 2022

La suppression de la modération des données sensibles (cf § 2.1.3) est un préalable à une refonte de la modération des actions sur le Guichet de la Fédération, et un renforcement du rôle du responsable de fédération.

Avant de détailler cette évolution, nous allons commencer par présenter les différents rôles des utilisateurs du Guichet de la Fédération ; préciser leur intervention dans le cycle de vie d'une entité ainsi que l'évolution de leur périmètre.

3.1 Rôles du Guichet de la Fédération

Un utilisateur (non administrateur) connecté sur le Guichet de la Fédération peut avoir un des rôles suivants :

- utilisateur : aucun droit particulier ;
- responsable d'entité(s) : l'utilisateur est identifié comme responsable d'au moins une entité ;
- responsable d'organisme : l'utilisateur est identifié comme responsable d'un organisme. À ce titre, il a les droits de validation d'un certain nombre de procédures, et les mêmes droits qu'un responsable d'entité, sur toutes les entités rattachées à son organisme ;
- responsable de fédération : l'utilisateur est identifié comme responsable d'au moins une fédération, et est notifié de chaque enregistrement d'une entité dans la fédération dont il est

responsable. Jusqu'à la version 3.2 incluse, il n'a aucun privilège sur les entités enregistrées dans sa fédération, et ne peut même pas en consulter les informations.

3.2 Cycle de vie d'une entité sur le Guichet de la Fédération

La Figure 4 présente le cycle de vie classique d'une entité SAML, de sa création à la publication de ses métadonnées dans une fédération.

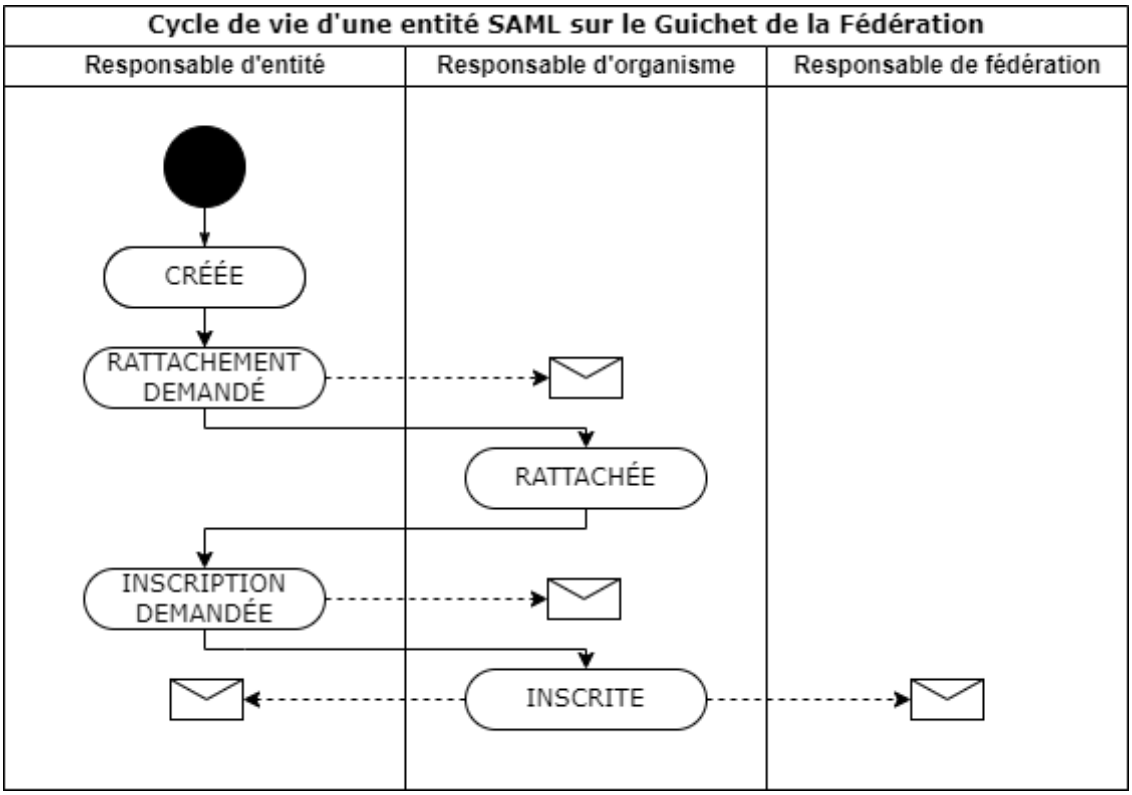

Figure 4 – Cycle de vie d'une entité SAML sur le Guichet de la Fédération

On voit que le responsable de la fédération n'est pas acteur de ce cycle de vie, alors qu'il semblerait naturel, qu'il puisse avoir un droit de regard.

3.3 Évolution du cycle de vie et du rôle de Responsable de Fédération

Ainsi en 2022, nous prévoyons de faire évoluer le Guichet de la Fédération afin que la modération de l'enregistrement dans une fédération soit effectuée par les responsables de cette fédération.

De plus, ces responsables auront aussi la possibilité de voir toutes les entités inscrites dans leur fédération et pourront *a minima* les désinscrire.

Dans le même temps, ils pourront également manipuler les paramètres de leur fédération afin de l'adapter à leurs usages :

- droits d'accès à leur fédération ;
- critères de pré-enregistrement (cf § 2.1.1).

En 2021, nous avons créé des fédérations spécifiques pour les usages de certaines grosses structures. Il nous semble donc normal de donner à ces structures les moyens d'être le plus autonome possible sur *leurs* fédérations.

Dans notre rôle d'opérateur de fédération nous sommes responsables des fédération Éducation-Recherche et eduGAIN. Ainsi, cette évolution va nous positionner en modérateur des enregistrements dans ces fédérations. C'est un bon moyen de s'assurer en amont du respect du cadre réglementaire et des conditions d'usage de ces fédérations.

4 Diffusion des métadonnées

4.1 Problématique

Le fonctionnement des fédérations que nous opérons repose sur la publication de métadonnées par nos soins, et leur consommation par les entités inscrites dans ces fédérations.

Depuis le début de la Fédération d'Identités, les métadonnées sont véhiculées par un fichier XML (trois par fédération : un complet, un contenant les IdP et un contenant les SP), mis à jour régulièrement ; et consommé par les entités à intervalles réguliers.

Le tableau suivant liste les versions de métadonnées mises à disposition par RENATER, ainsi que leur « popularité » respective :

Version	délai entre mise à jour du Guichet et publication	Taux d'usage[5]
preview	30 minutes	10 %
intermediate	6 heures	7 %
main	12 heures	83 %

La version correspond à un niveau de maturation. Ce mécanisme a été mis en place dans le passé pour palier à des fragilités de certains IdP/SP quant au format des métadonnées. L'objectif de tout administrateur étant de privilégier la stabilité, la majorité s'est orientée vers la version *main*. Ce fonctionnement présente actuellement peu d'intérêt, les métadonnées *preview* étant tout à fait stables.

La mise à disposition des métadonnées sous forme de fichier XML téléchargé par chaque entité a l'énorme avantage d'augmenter la résilience de nos fédérations, puisque l'intégralité des métadonnées est distribuée sur l'ensemble des participants. Les échanges lors des connexions ne faisant intervenir que le SP et l'IdP (au travers du navigateur), nous pourrions tout à fait arrêter de publier les métadonnées pendant plusieurs jours sans que cela impacte la communauté (hormis pour les ajouts et mises à jour).

Cette approche atteint cependant ses limites, pour deux raisons :

1. Le délai de propagation des métadonnées peut être long : le calcul du délai de propagation est détaillé en annexe. Ne pouvant pas faire de supposition ni sur le mode de récupération des métadonnées ni sur le paramétrage, nous annonçons toujours un délai de prise en compte d'au moins 24h [7]. Cette limite a toujours été présente ;
2. La taille des fichiers de métadonnées s'accroît constamment et commence à poser des soucis de consommation de ressource pour les entités – notamment au niveau de la mémoire vive. Des actions locales peuvent y remédier, mais nécessitent une très bonne maîtrise des métadonnées de la part des administrateurs système. Cette limite est de plus en plus

5 Calculé à partir des logs d'accès

prégnante, et nous a obligé côté RENATER à augmenter les ressources des IdP que nous hébergeons, et à changer le mode de consommation des métadonnées de nos SP.

Le tableau ci-dessous résume la taille des fichiers de métadonnées pour la Fédération Éducation-Recherche et eduGAIN :

	Fédération Éducation-Recherche	**eduGAIN**
Nombre d'IdP / SP	324 / 1206	4357 / 3419
Taille du fichier contenant les IdP et les SP	11 Mo	66 Mo
Taille du fichier contenant les IdP	2 Mo	43 Mo
Taille du fichier contenant les SP	8,4 Mo	24 Mo

Le volume de téléchargement de ces métadonnées atteint un total d'1 To par jour, soit une consommation de bande-passante moyenne de 100 Mbit/s.

D'après [8], ce volume génère un total annuel de 560 kg CO_2e.

Et pourtant, en moyenne un IdP est utilisé pour accéder à moins d'une centaine de services[6]. Pour un IdP inscrit dans la Fédération Éducation-Recherche et dans eduGAIN, il y a plus de 4500 SP disponibles. Un IdP va donc utiliser en moyenne 45 fois plus de ressources que nécessaire :

- espace mémoire ;
- consommation de bande-passante pour les transferts. Ce point pourrait être cependant mitigé à court-terme avec la mise à disposition de flux compressés : les entités supportant la compression tireraient partie d'un tel mécanisme ;
- espace disque ;
- CPU (parsing, validation des signatures…).

Nous allons dans la suite de cet article détailler le protocole MDQ, dont le but est de résoudre ce problème.

4.2 Le protocole MDQ

MDQ signifie Metadata Query Protocol, et est au statut de *draft* au niveau de l'IETF [9][10] - le premier *draft* ayant été proposé en 2013. Ce protocole permet à une entité SAML de récupérer « à la volée » les métadonnées d'une autre entité SAML, via une simple requête HTTP à un serveur disposant de ces métadonnées.

Ce changement de mode de récupération offre de nouvelles opportunités, mais nécessite aussi un peu de vigilance quant au paramétrage des entités, comme nous le verrons plus loin.

La diffusion des métadonnées via MDQ est déjà en place avec succès dans certains pays (États-Unis, Royaume-Uni, Hongrie…).

La mise en place de la diffusion des métadonnées de nos fédérations via MDQ n'est cependant pas à l'ordre du jour.

6 Statistique calculée à partir des IdP hébergés par RENATER pour le compte d'établissements de plus de 200 utilisateurs.

4.2.1 Description du protocole

En soit, le protocole MDQ est relativement simple : il est transporté par HTTP, avec un en-tête *Content-Type* valant *application/samlmetadata+xml*.

Lorsqu'une entité a besoin des métadonnées d'une autre entité pour une demande de connexion, elle va demander les métadonnées au serveur MDQ via une requête HTTP, comme montré Figure 5.

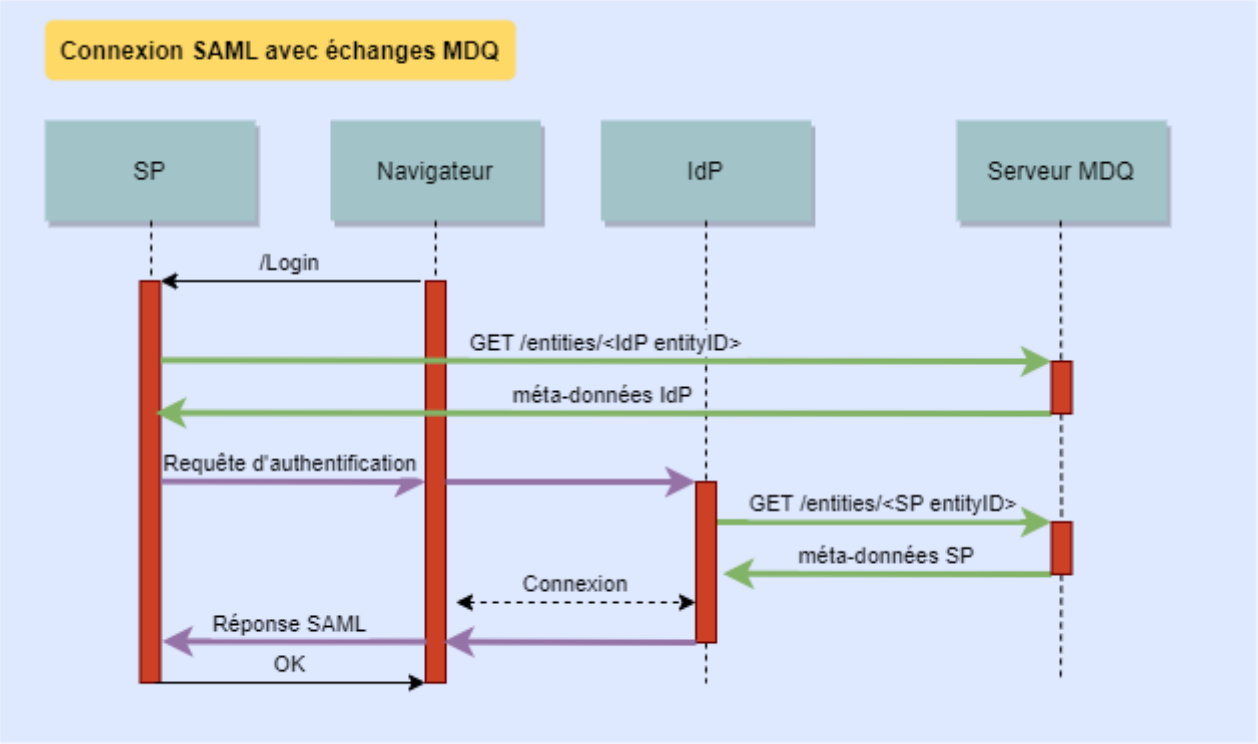

Figure 5 – Connexion SAML avec récupération des métadonnées via MDQ

Afin de limiter les appels, les traitements et augmenter la résilience, des mécanismes de cache sont implémentés côté client (entité) et côté serveur MDQ :

- Il est recommandé que les entités SAML implémentent un cache pour limiter les appels vers le serveur MDQ. L'IdP et le SP Shibboleth ont par défaut un cache de 6 heures. Ainsi un service appelé souvent ne générera que peu de trafic (ex : accès à Partage) ;
- Le SP Shibboleth dispose également d'un cache négatif, qui conserve l'information concernant l'absence de métadonnées pour une entité, ce qui évite de déclencher une requête vers le serveur MDQ, alors que celle-ci aura une forte probabilité d'avoir un résultat identique ;
- Le serveur MDQ doit fournir un en-tête de réponse *ETag* contenant une valeur associée à la ressource demandée. Lors des requêtes suivantes, la valeur est fournie par le client dans l'en-tête de requête *If-None-Match*, indiquant au serveur qu'il est inutile de renvoyer une réponse si la ressource n'a pas changé. Dans ce cas, le serveur renvoie le code 304 [11]. On peut imaginer l'utilisation d'une somme de contrôle en guise de valeur de l'en-tête *ETag* ;
- Le serveur MDQ peut fournir un en-tête de réponse *Cache-Control* incluant une directive *max-age* [12] permettant d'indiquer au client la durée au-delà de laquelle les métadonnées fournies peuvent être considérées comme périmées.

Il faut toutefois noter que l'IdP et le SP Shibboleth gèrent leur cache de manière autonome et ne font pas usage des en-têtes renvoyés par les serveurs MDQ. D'autres mécanismes de cache, au sein même des métadonnées peuvent être actionnés [13], et sont pris en compte par les IdP et SP Shibboleth.

4.2.2 L*'implémentation RENATER (et les autres implémentations disponibles)*

Il existe plusieurs implémentations du protocole MDQ. On peut en citer deux principales :

- **mdq-server** [14], qui est l'implémentation de référence – par l'auteur du protocole, et partage la même base de code que l'IdP Shibboleth ;
- **Pyff** [15], qui est historiquement un outil de manipulation des métadonnées SAML, et qui est également capable de jouer le rôle d'un serveur MDQ.

Nous avons jugé que ces deux implémentations ne convenaient pas à notre usage. Ceci nous a amenés à développer notre propre implémentation [16], pour des raisons suivantes :

- Un manque de visibilité sur la maturité de **mdq-server** : aucune version officielle n'est publiée ;
- Pour les deux solutions, un fonctionnement basé sur la consommation d'un ou plusieurs fichiers de métadonnées – ce qui ne donne aucune opportunité d'optimiser le délai de propagation des modifications ;
- Une consommation de ressources très importante rapportée par d'autres NREN sur l'utilisation de **Pyff** comme serveur MDQ, qui nous semble très exagérée pour quelque chose dont l'implémentation naïve se résume à un simple serveur web délivrant des fichiers statiques.

De fait, notre implémentation n'est composée que d'un script PHP chargé d'aller récupérer le fichier de métadonnées correspondant à l'entité demandée.

Ces données sur disque peuvent être assimilée à un cache. La mise à jour de ce cache est nécessairement décorrélée de leur publication, ce qui nous permet de réfléchir à plusieurs scénario de déploiement.

Le scénario le plus simple, actuellement en place, consiste à construire un fichier XML par entité à partir du fichier de métadonnées de chaque fédération.

Le résultat de ce simple scénario détaillé Figure 9 en annexe, est une mise à jour toutes les 30 minutes du serveur MDQ – produite à partir de nos fichiers de métadonnées les plus récents.

Ainsi, dans le pire des cas :

- Un nouveau SP sera visible au bout de 30 minutes ;
- Un SP modifié sera mis à jour au plus tard au bout de 6h dans les IdP ayant ses informations en cache.

4.2.3 Apports et points durs

On voit ainsi que la mise en place d'un service MDQ au sein de la Fédération d'Identité va pouvoir répondre aux difficultés actuelles liées aux métadonnées :

- La consommation de ressource par les IdP/SP pourra être réduite drastiquement. Sur certains IdP, des réductions de 50 % de l'empreinte mémoire ont été rapportées [17] ;
- Le délai de propagation des changements pourra être réduit, et même optimisé par les administrateurs de briques techniques.

D’un autre côté, faire reposer le fonctionnement de la Fédération d’Identité sur un système centralisé de diffusion des métadonnées amène un risque d’indisponibilité.

En effet, que se passe-t-il si https://mdq.renater.fr[7] ne répond plus ?

S’il est mis en œuvre, un tel service sera rendu hautement disponible, avec plusieurs instances, réparties sur plusieurs datacenters.

4.2.4 Perspectives

Si un serveur MDQ voyait le jour en 2022, des améliorations pourraient y être apportées dans le futur, notamment :

- Étudier la possibilité de réduire au minimum le délai entre une modification d’entité, et la mise à disposition des métadonnées ;
- Étudier le cas échéant, les moyens d’augmenter la disponibilité d’un tel service ;
- Authentifier l’accès aux métadonnées – celles-ci pouvant être considérées comme sensible d’un point de vue de la sécurité.

5 Conclusion

Depuis 2019, nous avons pu reprendre un cercle vertueux de mise à jour de nos outils.

Ceci nous permet de répondre aux problèmes que nous rencontrons sur la Fédération Éducation-Recherche – héritage de sa première décennie. Notre première approche a été de solliciter individuellement les administrateurs pour demander une mise en conformité. Cependant, le fait de demander des modifications sur des services fonctionnels (du point de vue des établissements) a vite montré ses limites : le nombre de corrections effectuées évolue très lentement.

Nous en tirons les conséquences pour le futur, en attribuant davantage de privilèges aux responsables de fédération – en plus de continuer à solliciter les responsables d’entité pour obtenir la résolution des problèmes.

Nous cherchons également à apporter des nouveautés à notre communauté, et c’est dans cet effort que s’inscrivent les évolutions du Guichet de la Fédération comme par exemple la simplification de la procédure de renouvellement des certificats, ainsi que la mise en place du service MDQ.

Annexe

Signature et chiffrement lors d’une connexion SAML

Lors d’une connexion SAML des requêtes sont échangées entre SP et IdP au travers du navigateur de l’utilisateur. Certaines sont chiffrées et/ou signées. La Figure 6 ci-dessous décrit les principes de chiffrement/signature dans le cas d’une connexion :

7 URL donnée à titre d’exemple

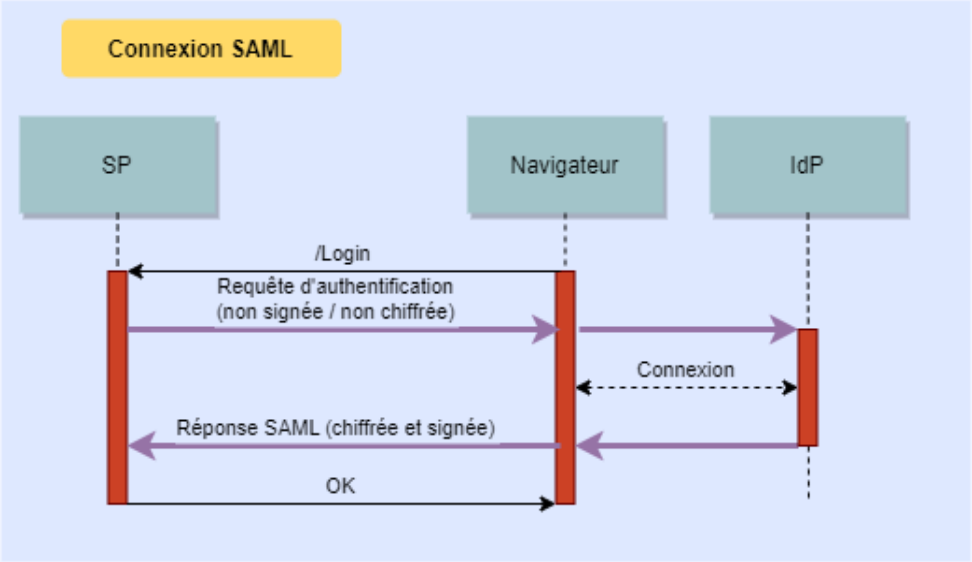

Figure 6 – Connexion SAML

Dans le cas général, seule la réponse SAML envoyée par l'IdP est signée et chiffrée :

- chiffrée avec la clé publique de chiffrement du SP ;
- signée avec la clé privée de signature de l'IdP.

Signature et chiffrement lors d'une déconnexion SAML

Lors d'une déconnexion, le SP est à l'origine du processus. Il va envoyer une requête SAML de déconnexion à l'IdP. L'IdP, quant à lui, peut être configuré pour propager la déconnexion vers les autres SP[8] pour lesquels une session a été créée :

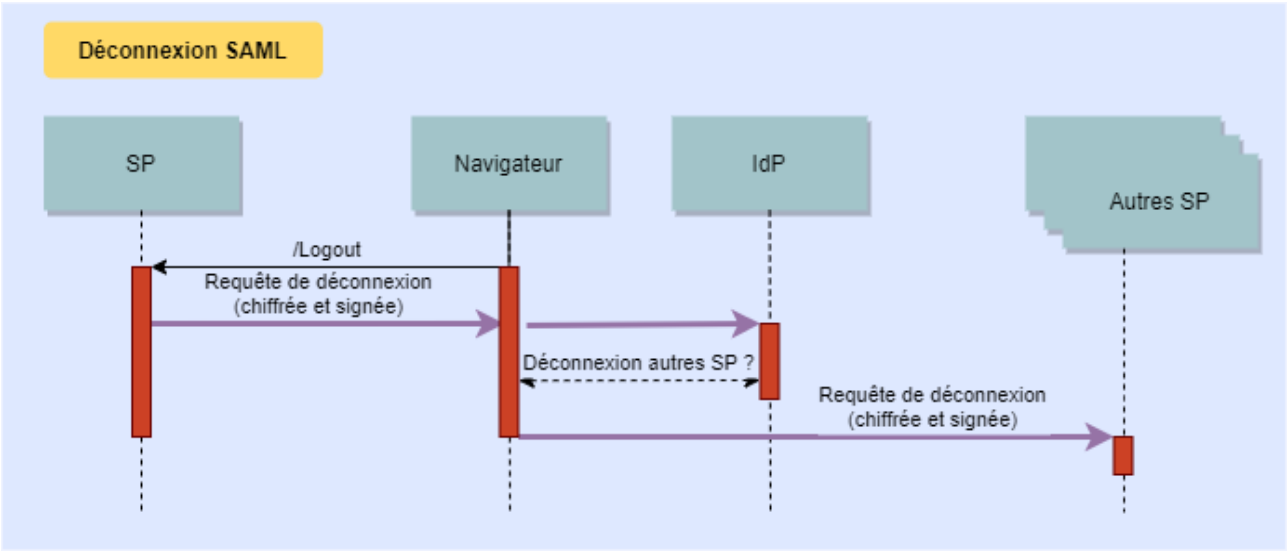

Figure 7 – Déconnexion SAML

La requête envoyée par le SP à l'IdP est elle aussi chiffrée et signée, mais cette fois :

8 La propagation de la déconnexion vers d'autres SP par l'IdP nécessite du paramétrage additionnel, qui sort du cadre ce cet article.

- chiffrée avec la clé publique de chiffrement de l'IdP ;
- signée avec la clé privée de signature du SP.

Délai de propagation d'une modifications

La Figure 8 illustre le délai de prise en compte de l'inscription d'un SP dans la Fédération Éducation-Recherche, pour 2 IdP avec une configuration par défaut :

- Un IdP A utilisant la version *preview* des métadonnées :
- Un IdP B utilisant la version *main* des métadonnées.

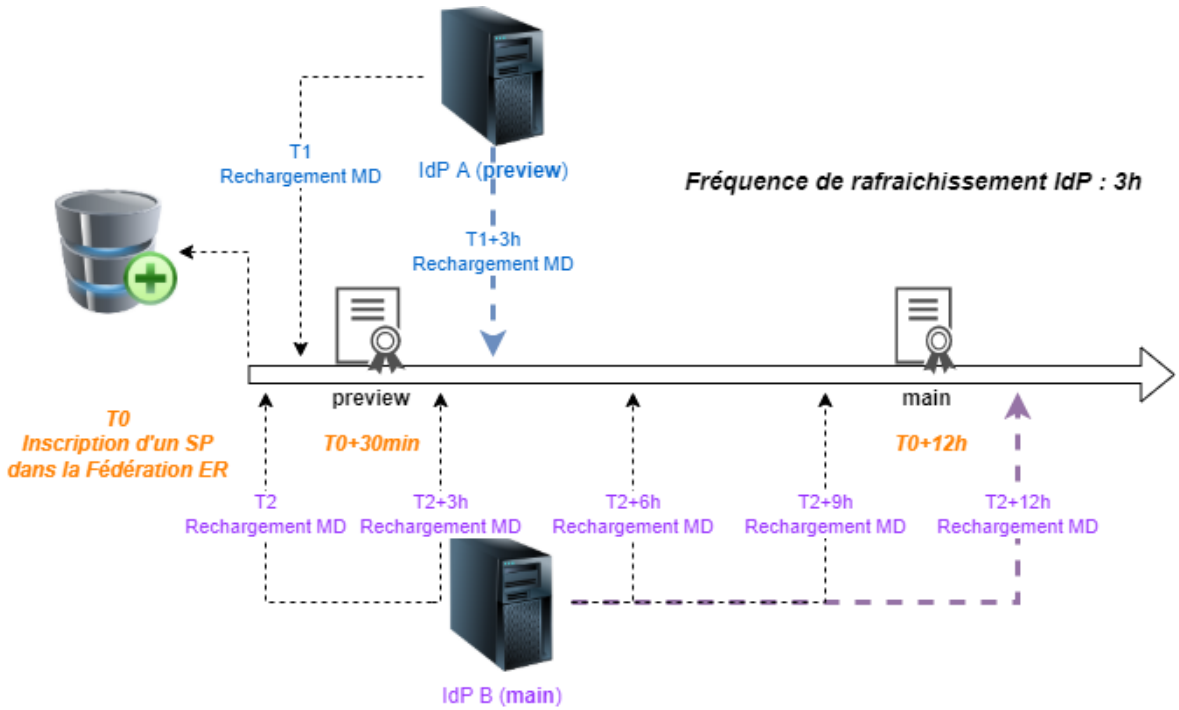

Figure 8 – Illustration de la propagation des métadonnées d'un SP

On peut ainsi calculer le délai minimum et maximum pour chaque IdP, entre l'inscription du SP (T0) et la prise en compte de ce SP par chaque IdP.

Pour l'IdP A, cela donne :

- $Dmin = 30\,min + \varepsilon$ Pour être dans cette condition, le dernier rechargement doit avoir eu lieu un peu moins de 3h avant la mise à jour de la version *preview* (ε) ;
- $Dmax = 3\,h + 30\,min - \varepsilon = 3\,h\,30 - \varepsilon$ Dans le pire des cas, le dernier rechargement a eu lieu très peu de temps (ε) avant la mise à jour de la version *preview.*

Pour l'IdP B, cela donne :

- $Dmin = 12\,h + \varepsilon$ Pour être dans cette condition, le dernier rechargement doit avoir eu lieu un peu moins de 3h avant la mise à jour de la version *main* (ε) ;
- $Dmax = 3\,h + 12\,h - \varepsilon = 15\,h - \varepsilon$ Dans le pire des cas, le dernier rechargement a eu lieu très peu de temps (ε) avant la mise à jour de la version *main.*

Dans le cas de MDQ, les délais sont drastiquement réduits, comme le montre la Figure 9 :

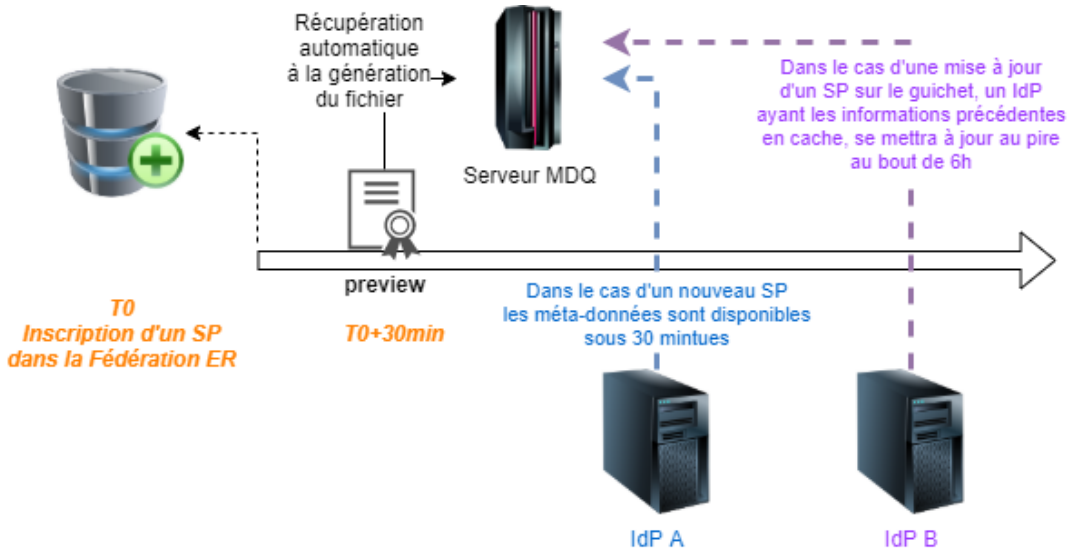

Figure 9 – Illustration de la propagation des métadonnées d'un SP avec MDQ

Bibliographie

[1] Arnoud G., Rousse G., Chabli A.. Amélioration de la confiance dans la Fédération Éducation-Recherche. Dans Actes du congrès JRES2019, Dijon, décembre 2019 ; https://conf-ng.jres.org/2019/document_revision_5284.html?download

[2] Collectif. Changements de la version 3.1 du Guichet de la Fédération. https://services.renater.fr/federation/outils/guichet-federation/v3/guichet-changes-v3.1#liberation_du_code

[3] Liste des changements apportés au Guichet de la Fédération. https://services.renater.fr/federation/outils/guichet-federation/changes

[4] Équipe Fédération. Choix d'un identifiant utilisateur. https://services.renater.fr/federation/documentation/generale/identifiant-utilisateur

[5] Collectif. Documentation des fédérations locales. https://services.renater.fr/federation/documentation/generale/federation-locale

[6] Équipe Fédération. Supervision métier de la Fédération Éducation-Recherche. https://services.renater.fr/federation/outils/sup-metier/index

[7] Équipe Fédération. Versions des fichiers de métadonnées. https://services.renater.fr/federation/documentation/generale/metadata/versions-metadata#delai_de_prise_en_compte_des_modifications

[8] Ficher M., Berthoud F., Ligozat A.-L., Sigonneau P., Tebbani B., Wisslé M.. Évaluation de l'empreinte carbone de la transmission d'un gigaoctet de données sur le réseau RENATER. https://ecoinfo.cnrs.fr/2020/12/04/quelle-est-lempreinte-carbone-de-la-transmission-d1-go-sur-le-reseau-renater/

[9] Young I. Metadata Query Protocol. https://datatracker.ietf.org/doc/draft-young-md-query/

[10] Young I. SAML Profile for the Metadata Query Protocol. https://datatracker.ietf.org/doc/draft-young-md-query-saml/

[11] Fielding R., Reschke J.. Hypertext Transfer Protocol (HTTP/1.1): Conditional Requests. https://datatracker.ietf.org/doc/html/rfc7232

[12] Fielding R., Nottingham M. Reschke J.. Hypertext Transfer Protocol (HTTP/1.1): Caching. https://datatracker.ietf.org/doc/html/rfc7234

[13] Metadata for the OASIS SecurityAssertion Markup Language (SAML) V2.0. https://docs.oasis-open.org/security/saml/v2.0/saml-metadata-2.0-os.pdf

[14] Young I. MDQ Server. https://github.com/iay/mdq-server

[15] Johansson L. the python Federation Feeder. https://pyff.io

[16] Équipe fédération. mdq-php. https://github.com/Renater/mdq-php

[17] Spaude P. Migrate to InCommon MDQ Shibboleth IdP. https://www.unicon.net/insights/blogs/migrate-to-incommon-mdq-shibboleth-idp