Єдиний вхід до Sketch Engine через Microsoft Entra ID, SimpleSAMLphp та eduGAIN

Інструкція для адміністраторів університетів, які використовують Microsoft 365 і хочуть надати студентам та працівникам доступ до Sketch Engine (та інших сервісів eduGAIN) з університетським обліковим записом.

---

0. Огляд

0.1. Як це працює

```
Користувач → Sketch Engine (Shibboleth SP)
                  |  вибір установи (eduGAIN)
                  ▼
            SimpleSAMLphp (IdP університету, https://<IDP_HOST>/idp)
                  |  проксі-автентифікація
                  ▼
            Microsoft Entra ID (облікові записи Microsoft 365)
```

SimpleSAMLphp виконує дві ролі одночасно:

- **SP (Service Provider) для Microsoft Entra ID** — перенаправляє користувача на вхід Microsoft і отримує від нього дані облікового запису;
- **IdP (Identity Provider) для федерації** — перетворює дані Microsoft на стандартні атрибути eduPerson/SCHAC і видає підписані SAML-відповіді сервісам eduGAIN, зокрема Sketch Engine.

0.2. Еталонна інфраструктура

Інтеграцію розгорнуто й перевірено в Київському національному лінгвістичному університеті на такій інфраструктурі:

Компонент	Версія / параметр
ОС	Ubuntu 24.04 LTS
Вебсервер	Apache 2.4 з mod_php (без PHP-FPM)
PHP	8.3
SimpleSAMLphp	гілка 2.x, встановлено в `/var/www/simplesamlphp`
Сервер	3,8 ГБ RAM, без swap
DNS / проксі	Cloudflare (проксування увімкнене)
Облікові записи	Microsoft 365 / Microsoft Entra ID (один тенант, домен університету)
Федерація	PEANO (координатор — Асоціація URAN), експорт в eduGAIN
Сервіс	Sketch Engine (SP на Shibboleth, entityID `https://auth.sketchengine.eu/entity`)

На іншій інфраструктурі (Debian, Nginx + PHP-FPM, інша версія PHP) логіка налаштування та конфігураційні файли SimpleSAMLphp не змінюються, відрізняються лише команди встановлення пакетів і налаштування вебсервера.

0.3. Плейсхолдери

Усі значення, специфічні для вашого університету, позначено плейсхолдерами. У розділі 1.11 є команда, яка підставляє їх у всі конфігураційні файли одразу.

Плейсхолдер	Що означає
`<IDP_HOST>`	DNS-ім'я сервера IdP, наприклад піддомен `idp.` вашого домену
`<DOMAIN>`	Основний домен університету, він же домен облікових записів Microsoft 365 і scope
`<TENANT_ID>`	Tenant ID вашого тенанту Microsoft Entra ID (GUID)
`<ADMIN_EMAIL>`	Адреса технічного контакту IdP (потрапить у публічні метадані)

`<ORG_NAME_UK>`	Повна назва університету українською
`<ORG_NAME_EN>`	Повна назва університету англійською
`<ORG_SHORT>`	Коротка назва (абревіатура)
`<ORG_URL>`	Адреса головного сайту університету
`<PRIVACY_URL>`	Адреса сторінки політики конфіденційності сервісу (див. розділ 3.2)

0.4. Рекомендований порядок виконання

1. Розділ 1.1–1.5: встановлення SimpleSAMLphp, базова конфігурація, ключі IdP.
2. Розділ 2: застосунок у Microsoft Entra ID.
3. Розділ 1.6–1.12: зв'язка з Entra ID, роль IdP, внутрішнє тестування.
4. Розділ 3: реєстрація у PEANO / eduGAIN і завантаження метаданих федерації.
5. Розділ 4: підключення до Sketch Engine.

---

1. Налаштування SimpleSAMLphp

1.1. Пакети

```
sudo apt update
sudo apt install -y apache2 libapache2-mod-php8.3 \
  php8.3 php8.3-cli php8.3-xml php8.3-mbstring php8.3-curl php8.3-intl \
  php8.3-zip php8.3-gmp php8.3-sqlite3 composer curl openssl
```

Розширення `php8.3-sqlite3` обов'язкове: без нього `composer` відмовиться встановлювати модулі SimpleSAMLphp (розділ 3.5).

1.2. Встановлення SimpleSAMLphp

Завантажте останній реліз гілки 2.x (архів `simplesamlphp-2.x.y-full.tar.gz`) зі сторінки https://github.com/simplesamlphp/simplesamlphp/releases.

```
cd /tmp
# підставте актуальну версію
curl -LO https://github.com/simplesamlphp/simplesamlphp/releases/download/v2.X.Y/simplesamlphp-2.X.Y-full.tar.gz
sudo tar xzf simplesamlphp-2.X.Y-full.tar.gz -C /var/www/
sudo mv /var/www/simplesamlphp-2.X.Y /var/www/simplesamlphp

cd /var/www/simplesamlphp
sudo cp config/config.php.dist config/config.php
sudo cp config/authsources.php.dist config/authsources.php
sudo chown -R www-data:www-data /var/www/simplesamlphp
```

1.3. Apache і TLS

Файл `/etc/apache2/sites-available/simplesamlphp.conf`:

```
<VirtualHost *:443>
    ServerName <IDP_HOST>

    SetEnv SIMPLESAMLPHP_CONFIG_DIR /var/www/simplesamlphp/config
    Alias /simplesaml /var/www/simplesamlphp/public

    <Directory /var/www/simplesamlphp/public>
        Require all granted
    </Directory>

    SSLEngine on
    SSLCertificateFile    /etc/ssl/certs/<IDP_HOST>.crt
    SSLCertificateKeyFile /etc/ssl/private/<IDP_HOST>.key
</VirtualHost>
```

```
sudo a2enmod ssl
sudo a2ensite simplesamlphp
sudo apache2ctl configtest && sudo systemctl reload apache2
```

Якщо домен проксується через Cloudflare, увімкніть режим SSL **Full (strict)** і використовуйте на сервері сертифікат Cloudflare Origin або Let's Encrypt.

1.4. config.php

Згенеруйте хеш пароля адміністратора й сіль:

```
cd /var/www/simplesamlphp
sudo -u www-data php bin/pwgen.php      # введіть пароль, скопіюйте хеш
openssl rand -hex 32                    # значення для secretsalt
```

Змініть у `config/config.php` такі параметри (вони вже є у файлі, змінюйте наявні рядки, а не додавайте нові):

```
'baseurlpath'            => 'https://<IDP_HOST>/simplesaml/',
'auth.adminpassword'     => '<ХЕШ_З_PWGEN>',
'secretsalt'             => '<РЕЗУЛЬТАТ_OPENSSL_RAND>',
'technicalcontact_name'  => '<ORG_SHORT> IT',
'technicalcontact_email' => '<ADMIN_EMAIL>',
'enable.saml20-idp'      => true,
```

Значення `secretsalt` після запуску **не змінюйте**: від нього залежать псевдонімні ідентифікатори (eduPersonTargetedID) усіх користувачів.

Блок модулів (справжній блок `'module.enable' => [`, **не** приклад у коментарі над ним):

```
'module.enable' => [
    'core'        => true,
    'admin'       => true,
    'saml'        => true,
    'metarefresh' => true,   // знадобиться в розділі 3.5
    'cron'        => true,
],
```

Джерела метаданих:

```
'metadata.sources' => [
    ['type' => 'flatfile'],
    ['type' => 'flatfile', 'directory' => 'metadata/metarefresh-edugain'],
],
```

Важливо: вимкніть глобальний `core:AttributeLimit`. У стандартному `config.php` в блоці `'authproc.idp'` є рядок `50 => 'core:AttributeLimit',`. Його треба закоментувати:

```
// 50 => 'core:AttributeLimit',   // замінено кроком 95 у saml20-idp-hosted.php
```

Інакше цей фільтр спрацює раніше за перетворення імен атрибутів на OID, не знайде збігів зі списком атрибутів Sketch Engine і видалить усі атрибути з відповіді.

```
sudo -u www-data php -l config/config.php
```

1.5. Ключ і сертифікат IdP

```
cd /var/www/simplesamlphp/cert
sudo openssl req -newkey rsa:3072 -x509 -days 3652 -nodes \
  -subj "/CN=<IDP_HOST>" \
  -keyout idp.pem -out idp.crt
sudo chown www-data:www-data idp.pem idp.crt
```

```
sudo chmod 600 idp.pem
sudo chmod 644 idp.crt
```

Федерація вимагає ключ щонайменше 2048 біт.

1.6. Сертифікат Microsoft Entra ID

Виконується **після** розділу 2.4. Збережіть сертифікат підпису SAML, завантажений з Entra (формат Base64), як:

```
/var/www/simplesamlphp/cert/azure_idp.crt
```

```
sudo chown www-data:www-data /var/www/simplesamlphp/cert/azure_idp.crt
head -n1 /var/www/simplesamlphp/cert/azure_idp.crt   # має бути -----BEGIN CERTIFICATE-----
```

1.7. authsources.php

Додайте до масиву `$config` у `config/authsources.php` (запис `admin` залиште):

```
    // SP-частина: вхід через Microsoft Entra ID
    'microsoft-entra-id' => [
        'saml:SP',
        'entityID' => 'https://<IDP_HOST>/simplesaml/module.php/saml/sp/metadata/microsoft-entra-id',
        'idp'      => 'https://sts.windows.net/<TENANT_ID>/',
    ],

    // Тимчасовий тестовий SP для перевірки ролі IdP (видалити після запуску)
    'idp-selftest' => [
        'saml:SP',
        'entityID' => 'https://<IDP_HOST>/simplesaml/module.php/saml/sp/metadata/idp-selftest',
        'idp'      => 'https://<IDP_HOST>/idp',
    ],
```

У SimpleSAMLphp 2.x параметр `entityID` для `saml:SP` обов'язковий.

1.8. metadata/saml20-idp-remote.php

Метадані віддалених IdP: Microsoft Entra ID та (для самотесту) ваш власний IdP.

```
sudo tee /var/www/simplesamlphp/metadata/saml20-idp-remote.php > /dev/null <<'EOF'
<?php

/**
 * SAML 2.0 remote IdP metadata
 */

// Microsoft Entra ID
$metadata['https://sts.windows.net/<TENANT_ID>/'] = [
    'entityid' => 'https://sts.windows.net/<TENANT_ID>/',
    'SingleSignOnService' => [
        [
            'Binding'  => 'urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect',
            'Location' => 'https://login.microsoftonline.com/<TENANT_ID>/saml2',
        ],
    ],
    'SingleLogoutService' => [
        [
            'Binding'  => 'urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect',
            'Location' => 'https://login.microsoftonline.com/<TENANT_ID>/saml2',
        ],
    ],
    'certificate' => 'azure_idp.crt',
];
```

```
// Власний IdP — лише для самотесту (idp-selftest), видалити після запуску
$metadata['https://<IDP_HOST>/idp'] = [
    'SingleSignOnService' => [
        [
            'Binding'  => 'urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect',
            'Location' => 'https://<IDP_HOST>/simplesaml/module.php/saml/idp/singleSignOnService',
        ],
    ],
    'certificate' => 'idp.crt',
];
EOF
```

Зверніть увагу на дві вимоги SimpleSAMLphp 2.x до файлів метаданих:

- файл має заповнювати змінну `$metadata['<entityID>'] = [...]`; формат `return [...]` **не працює** (дає помилку `MetadataNotFound`);
- кожен ендпоінт (`SingleSignOnService`, `AssertionConsumerService` тощо) задається масивом з `Binding` і `Location`; скорочений запис одним рядком дає помилку «Endpoint ... is not an array».

1.9. metadata/saml20-idp-hosted.php

Головний файл ролі IdP: ідентифікатор, ключі, перетворення атрибутів Microsoft на атрибути федерації, опис для каталогу eduGAIN.

```
sudo tee /var/www/simplesamlphp/metadata/saml20-idp-hosted.php > /dev/null <<'EOF'
<?php

/**
 * SAML 2.0 IdP університету (проксі до Microsoft Entra ID)
 */

$metadata['https://<IDP_HOST>/idp'] = [
    'host'        => '__DEFAULT__',
    'privatekey'  => 'idp.pem',
    'certificate' => 'idp.crt',
    'auth'        => 'microsoft-entra-id',

    'scope' => ['<DOMAIN>'],
    'attributes.NameFormat' => 'urn:oasis:names:tc:SAML:2.0:attrname-format:uri',

    'authproc' => [
        // Claims Microsoft Entra ID -> короткі імена
        10 => [
            'class' => 'core:AttributeMap',
            'http://schemas.xmlsoap.org/ws/2005/05/identity/claims/name'         => 'eduPersonPrincipalName',
            'http://schemas.xmlsoap.org/ws/2005/05/identity/claims/emailaddress' => 'mail',
            'http://schemas.xmlsoap.org/ws/2005/05/identity/claims/givenname'    => 'givenName',
            'http://schemas.xmlsoap.org/ws/2005/05/identity/claims/surname'      => 'sn',
            'http://schemas.microsoft.com/identity/claims/displayname'           => 'displayName',
            'http://schemas.microsoft.com/identity/claims/objectidentifier'      => 'entraObjectId',
        ],
        // Пускати лише облікові записи домену університету (відсікає гостей тенанту)
        15 => [
            'class' => 'core:PHP',
            'code'  => '
                $eppn = strtolower($attributes["eduPersonPrincipalName"][0] ?? "");
                if (substr_count($eppn, "@") !== 1 || str_contains($eppn, "#")
                    || !str_ends_with($eppn, "@<DOMAIN>")) {
                    throw new \\SimpleSAML\\Error\\Exception("Access restricted to <DOMAIN> accounts");
                }
            ',
        ],
        // Статичні атрибути установи
        20 => [
            'class' => 'core:AttributeAdd',
```

```
            'schacHomeOrganization' => ['<DOMAIN>'],
            'eduPersonAffiliation'  => ['member'],
        ],
        // member -> member@<DOMAIN>
        30 => [
            'class'           => 'core:ScopeAttribute',
            'scopeAttribute'  => 'eduPersonPrincipalName',
            'sourceAttribute' => 'eduPersonAffiliation',
            'targetAttribute' => 'eduPersonScopedAffiliation',
        ],
        // Стабільний псевдонім для кожної пари користувач–сервіс
        40 => [
            'class' => 'core:TargetedID',
            'identifyingAttribute' => 'entraObjectId',
        ],
        // commonName = displayName (його запитує Sketch Engine)
        45 => [
            'class' => 'core:AttributeCopy',
            'displayName' => 'cn',
        ],
        // Короткі імена -> urn:oid:...
        90 => ['class' => 'core:AttributeMap', 'name2oid'],
        // Обмеження атрибутів ПІСЛЯ перетворення на OID.
        // SP зі списком атрибутів у метаданих отримує лише свій список;
        // SP без списку — лише неідентифікуючі атрибути нижче.
        95 => [
            'class'   => 'core:AttributeLimit',
            'default' => true,
            'urn:oid:1.3.6.1.4.1.5923.1.1.1.9',
            'urn:oid:1.3.6.1.4.1.5923.1.1.1.1',
            'urn:oid:1.3.6.1.4.1.5923.1.1.1.10',
            'urn:oid:1.3.6.1.4.1.25178.1.2.9',
        ],
    ],

    'OrganizationName'        => ['uk' => '<ORG_NAME_UK>', 'en' => '<ORG_NAME_EN>'],
    'OrganizationDisplayName' => ['uk' => '<ORG_SHORT>', 'en' => '<ORG_SHORT>'],
    'OrganizationURL'         => ['uk' => '<ORG_URL>', 'en' => '<ORG_URL>'],
    'UIInfo' => [
        'DisplayName'         => ['uk' => '<ORG_NAME_UK>', 'en' => '<ORG_NAME_EN>'],
        'Description'         => ['uk' => 'Вхід для студентів і працівників <ORG_SHORT>', 'en' => 'Login for <ORG_SHORT> students
and staff'],
        'InformationURL'      => ['en' => '<ORG_URL>'],
        'PrivacyStatementURL' => ['en' => '<PRIVACY_URL>'],
    ],
    'contacts' => [
        ['contactType' => 'technical', 'givenName' => '<ORG_SHORT> IT', 'emailAddress' => 'mailto:<ADMIN_EMAIL>'],
        ['contactType' => 'support',   'givenName' => '<ORG_SHORT> IT', 'emailAddress' => 'mailto:<ADMIN_EMAIL>'],
    ],
];
EOF
```

Пояснення до окремих кроків:

- **Крок 10.** Назви claims — стандартні для SAML-застосунку Entra ID. Якщо у вашому тенанті claims змінено, звірте назви з результатом тесту `microsoft-entra-id` (розділ 1.12).
- **Крок 15.** Відхиляє облікові записи, що не належать домену університету, зокрема гостей B2B (`…#EXT#@…onmicrosoft.com`). Подвійні зворотні слеші правильні: код міститься в PHP-рядку в одинарних лапках.
- **Крок 40.** Не додавайте сюди `'nameId' => true`. SimpleSAMLphp 2.x сам загортає eduPersonTargetedID у NameID, і з цим параметром значення загортається двічі, що робить атрибут некоректним.
- **Крок 95.** Має стояти після кроку 90. Sketch Engine та інші SP вказують запитувані атрибути у форматі `urn:oid:…`, тому порівнювати їх можна лише після перетворення імен.

Якщо назва університету містить апостроф, екрануйте його в цьому файлі як `\'` (рядки PHP в одинарних лапках).

1.10. metadata/saml20-sp-remote.php

Метадані тестового SP. Список атрибутів збігається з тим, що запитує Sketch Engine, тож самотест показує саме ту відповідь, яку отримає Sketch Engine.

```
sudo tee /var/www/simplesamlphp/metadata/saml20-sp-remote.php > /dev/null <<'EOF'
<?php

/**
 * SAML 2.0 remote SP metadata
 */

// Тимчасовий тестовий SP (видалити після запуску)
$metadata['https://<IDP_HOST>/simplesaml/module.php/saml/sp/metadata/idp-selftest'] = [
    'AssertionConsumerService' => [
        [
            'Binding'  => 'urn:oasis:names:tc:SAML:2.0:bindings:HTTP-POST',
            'Location' => 'https://<IDP_HOST>/simplesaml/module.php/saml/sp/saml2-acs.php/idp-selftest',
        ],
    ],
    'attributes' => [
        'urn:oid:1.3.6.1.4.1.5923.1.1.1.6',
        'urn:oid:2.16.840.1.113730.3.1.241',
        'urn:oid:1.3.6.1.4.1.5923.1.1.1.9',
        'urn:oid:1.3.6.1.4.1.5923.1.1.1.10',
        'urn:oid:0.9.2342.19200300.100.1.3',
        'urn:oid:2.5.4.3',
        'urn:oid:2.5.4.42',
        'urn:oid:2.5.4.4',
        'urn:oid:1.3.6.1.4.1.25178.1.2.9',
    ],
];
EOF
```

1.11. Підстановка значень, права, перевірка

Заповніть змінні своїми значеннями (без символів `|` і `&`) і виконайте:

```
IDP_HOST="..."
DOMAIN="..."
TENANT_ID="..."
ADMIN_EMAIL="..."
ORG_NAME_UK="..."
ORG_NAME_EN="..."
ORG_SHORT="..."
ORG_URL="..."
PRIVACY_URL="..."

cd /var/www/simplesamlphp
FILES="config/config.php config/authsources.php metadata/saml20-idp-hosted.php metadata/saml20-idp-remote.php metadata/saml20-sp-remote.php"

sudo sed -i \
  -e "s|<IDP_HOST>|${IDP_HOST}|g" \
  -e "s|<DOMAIN>|${DOMAIN}|g" \
  -e "s|<TENANT_ID>|${TENANT_ID}|g" \
  -e "s|<ADMIN_EMAIL>|${ADMIN_EMAIL}|g" \
  -e "s|<ORG_NAME_UK>|${ORG_NAME_UK}|g" \
  -e "s|<ORG_NAME_EN>|${ORG_NAME_EN}|g" \
  -e "s|<ORG_SHORT>|${ORG_SHORT}|g" \
  -e "s|<ORG_URL>|${ORG_URL}|g" \
  -e "s|<PRIVACY_URL>|${PRIVACY_URL}|g" \
  $FILES
```

```
grep -n "<[A-Z_]*>" $FILES      # має нічого не знайти

sudo chown www-data:www-data $FILES
sudo chmod 640 $FILES
for f in $FILES; do sudo -u www-data php -l "$f"; done
sudo systemctl reload apache2
```

Перевірка завантаження метаданих:

```
cd /var/www/simplesamlphp
sudo -u www-data php -r '
require "vendor/autoload.php";
$h = \SimpleSAML\Metadata\MetaDataStorageHandler::getMetadataHandler(\SimpleSAML\Configuration::getInstance());
foreach (["saml20-idp-hosted","saml20-idp-remote","saml20-sp-remote"] as $s) {
  echo $s, ": "; var_export(array_keys($h->getList($s))); echo PHP_EOL;
}'
```

Очікуваний результат:

- `saml20-idp-hosted`: `https://<IDP_HOST>/idp`;
- `saml20-idp-remote`: `https://sts.windows.net/<TENANT_ID>/` і `https://<IDP_HOST>/idp`;
- `saml20-sp-remote`: URL `idp-selftest`.

Метадані IdP (їх URL передається у федерацію):

```
curl -s https://<IDP_HOST>/simplesaml/module.php/saml/idp/metadata | head -n 20
curl -s https://<IDP_HOST>/simplesaml/module.php/saml/idp/metadata | grep -o '<shibmd:Scope[^<]*</shibmd:Scope>'
```

У метаданих має бути `entityID="https://<IDP_HOST>/idp"` і `<shibmd:Scope regexp="false"><DOMAIN></shibmd:Scope>`. Без scope Shibboleth на боці Sketch Engine відкине ePPN і eduPersonScopedAffiliation.

1.12. Тестування

Адмін-панель `https://<IDP_HOST>/simplesaml/` → **Test** (Test authentication sources).

Тест 1: `microsoft-entra-id`. Перевіряє зв'язку SimpleSAMLphp ↔ Entra ID. Після входу Microsoft має відкритися сторінка з claims Entra (`…/claims/name`, `…/claims/emailaddress` тощо). Звірте назви з кроком 10 у `saml20-idp-hosted.php`.

Тест 2: `idp-selftest`. Перевіряє повний ланцюжок: тестовий SP → ваш IdP → Entra ID → відповідь IdP. Перед тестом натисніть **Logout**. Має з'явитися рівно 9 атрибутів:

Атрибут	OID	Очікуване значення
eduPersonPrincipalName	`urn:oid:1.3.6.1.4.1.5923.1.1.1.6`	UPN користувача `…@<DOMAIN>`
schacHomeOrganization	`urn:oid:1.3.6.1.4.1.25178.1.2.9`	`<DOMAIN>`
eduPersonScopedAffiliation	`urn:oid:1.3.6.1.4.1.5923.1.1.1.9`	`member@<DOMAIN>`
eduPersonTargetedID	`urn:oid:1.3.6.1.4.1.5923.1.1.1.10`	хеш у NameID, однаковий між сесіями
mail	`urn:oid:0.9.2342.19200300.100.1.3`	пошта користувача
givenName	`urn:oid:2.5.4.42`	ім'я
sn	`urn:oid:2.5.4.4`	прізвище
displayName	`urn:oid:2.16.840.1.113730.3.1.241`	повне ім'я
cn	`urn:oid:2.5.4.3`	повне ім'я

Службових claims Microsoft (`tenantid`, `identityprovider`, `authnmethodsreferences`) і атрибута `entraObjectId` у відповіді бути не повинно. Виконайте тест і службовим, і звичайним обліковим записом користувача.

---

2. Налаштування Microsoft Entra ID

Потрібні права Global Administrator, Cloud Application Administrator або Application Administrator.

2.1. Створення застосунку

1. Entra admin center → **Enterprise applications** → **New application** → **Create your own application**.
2. Назва, наприклад `<ORG_SHORT> SimpleSAMLphp IdP`; варіант **Integrate any other application you don't find in the gallery (Non-gallery)**.
3. Відкрийте застосунок → **Single sign-on** → **SAML**.

2.2. Basic SAML Configuration

Поле	Значення
Identifier (Entity ID)	`https://<IDP_HOST>/simplesaml/module.php/saml/sp/metadata/microsoft-entra-id`
Reply URL (ACS)	`https://<IDP_HOST>/simplesaml/module.php/saml/sp/saml2-acs.php/microsoft-entra-id`
Sign on URL	`https://<IDP_HOST>/simplesaml/` (необов'язково)

Identifier має **посимвольно** збігатися з `entityID` джерела `microsoft-entra-id` в `authsources.php`, а Reply URL — з адресою ACS у метаданих SP:

```
curl -s https://<IDP_HOST>/simplesaml/module.php/saml/sp/metadata/microsoft-entra-id \
  | grep -o '<md:AssertionConsumerService[^>]*>'
```

Натисніть **Save** і дочекайтеся підтвердження збереження.

2.3. Attributes & Claims

Залиште стандартний набір claims застосунку:

Claim	Джерело
`…/claims/givenname`	`user.givenname`
`…/claims/surname`	`user.surname`
`…/claims/emailaddress`	`user.mail`
`…/claims/name`	`user.userprincipalname`
Unique User Identifier	`user.userprincipalname`

Entra додатково надсилає `displayname`, `objectidentifier`, `tenantid` та інші службові claims. `displayname` і `objectidentifier` використовуються в кроці 10, решту IdP відсікає.

Claim `…/claims/name` має містити саме UPN виду `ім'я@<DOMAIN>`: він стає eduPersonPrincipalName. Значення ePPN не можна перепризначати іншій людині, тож UPN звільнених або відрахованих користувачів не слід повторно видавати новим.

2.4. Сертифікат підпису SAML

У блоці **SAML Certificates** завантажте **Certificate (Base64)** і збережіть його на сервері як `cert/azure_idp.crt` (розділ 1.6).

Сертифікат Entra має обмежений строк дії (за замовчуванням 3 роки). Перед його заміною в Entra оновіть файл на сервері, інакше вхід припиниться.

Tenant ID для плейсхолдера `<TENANT_ID>` наведено на сторінці **Overview** тенанту, а також у полі **Microsoft Entra Identifier** (`https://sts.windows.net/<TENANT_ID>/`).

2.5. Доступ користувачів

Застосунок → **Properties** → **Assignment required?**

- **No** — вхід доступний усім обліковим записам тенанту. Рекомендовано для університетського IdP: не треба вручну додавати кожного нового студента. Гостьові облікові записи відсікає крок 15 у `saml20-idp-hosted.php`.
- **Yes** — вхід лише для призначених у **Users and groups**. Призначення груп потребує ліцензії Microsoft Entra ID P1 (є в освітніх планах A3/A5, немає в A1).

2.6. Типові помилки Entra ID

Код	Причина	Рішення
`AADSTS700016`	Identifier у Basic SAML Configuration не збігається з `entityID` SP	Вирівняти значення (2.2)
`AADSTS50011`	Reply URL із запиту не додано до застосунку	Додати точний Reply URL і зберегти (2.2)
`AADSTS50105`	Увімкнено Assignment required, користувача не призначено	Призначити користувача/групу або вимкнути Assignment required (2.5)

---

3. Реєстрація у федерації PEANO та eduGAIN

3.1. Передумови

- Українська федерація ідентифікації для науки та освіти — **PEANO**, координатор — Асоціація **URAN** (https://peano.uran.ua). Через PEANO установа потрапляє в міжнародну інтерфедерацію **eduGAIN**, в якій зареєстрований Sketch Engine.
- Виступати як IdP у PEANO можуть лише URAN та її члени. Кожна установа може мати лише один IdP.
- Перевірте членство університету в URAN до подачі документів.

3.2. Документи

1. **Політика використання IdP.** Хто отримує обліковий запис, на підставі чого (наказ про зарахування, прийом на роботу), як перевіряється особа, коли запис блокується (відрахування, звільнення), як захищені облікові записи (MFA, політики паролів).
2. **Політика конфіденційності сервісу**, опублікована за адресою `<PRIVACY_URL>`. Ця адреса вказана в метаданих IdP, тож сторінка має реально існувати до подачі заявки. Що в ній описати: які атрибути передаються, кому і з якою метою, правову основу (Закон України «Про захист персональних даних», GÉANT Data Protection Code of Conduct), строки зберігання журналів, права користувачів і контакт.

3.3. Заявка до PEANO

Лист координатору федерації, підписаний уповноваженою особою університету, містить:

- адміністративний і технічний контакти (ПІБ, посада, email, телефон);
- підтвердження ознайомлення з Policy, Technical Rules і Metadata Registration Practice Statement федерації та згоду з ними;
- **явну згоду на експорт метаданих в eduGAIN** (без неї установа не буде видимою для Sketch Engine);
- технічні дані IdP.

Шаблон технічних даних:

Параметр	Значення
Організація	`<ORG_NAME_UK>`
EntityID	`https://<IDP_HOST>/idp`
URL метаданих	`https://<IDP_HOST>/simplesaml/module.php/saml/idp/metadata`
Scope	`<DOMAIN>`
Протокол	SAML 2.0 Web Browser SSO
Програмне забезпечення	SimpleSAMLphp 2.x
Джерело облікових записів	Microsoft Entra ID (Microsoft 365, домен `<DOMAIN>`)
Ключ підпису	RSA 3072 біт
Формат імен атрибутів	`urn:oasis:names:tc:SAML:2.0:attrname-format:uri`
Технічний контакт	`<ADMIN_EMAIL>`

До заявки додайте таблицю атрибутів з розділу 1.12, а також опишіть, що:

- передача атрибутів обмежена для кожного SP окремо, а SP без погодженого переліку отримують лише неідентифікуючі атрибути;
- вхід дозволено лише обліковим записам домену `<DOMAIN>`;
- значення eduPersonPrincipalName не перепризначаються іншим особам.

Після тестів сумісності та публікації координатор надішле посилання на запис вашого IdP у каталозі eduGAIN (met.refeds.org).

3.4. Модуль metarefresh

Щоб IdP довіряв Sketch Engine, його метадані завантажуються з підписаного агрегату eduGAIN.

```
cd /var/www/simplesamlphp
sudo -u www-data COMPOSER_HOME=/tmp/composer-www composer require simplesamlphp/simplesamlphp-module-metarefresh
ls modules/ | grep -E 'cron|metarefresh'     # мають бути обидва
```

Модуль `cron` входить до складу SimpleSAMLphp 2.x, окремо його встановлювати не потрібно.

3.5. Сертифікат підпису eduGAIN

```
cd /var/www/simplesamlphp/cert
sudo curl -sSL -o edugain-mds-v2.pem https://technical.edugain.org/mds-v2.cer
openssl x509 -in edugain-mds-v2.pem -noout -fingerprint -sha256
sudo chown www-data:www-data edugain-mds-v2.pem
```

Виведений SHA-256 відбиток має **точно** збігатися з опублікованим на https://technical.edugain.org/metadata. Якщо не збігається, не продовжуйте.

3.6. Конфігурація metarefresh і cron

Агрегат eduGAIN дуже великий, тому з нього завантажуються лише потрібні SP через `whitelist`. Нові сервіси додаються в цей список за потреби.

```
cd /var/www/simplesamlphp
sudo -u www-data mkdir -p metadata/metarefresh-edugain

sudo tee config/module_metarefresh.php > /dev/null <<'EOF'
<?php

$config = [
    'sets' => [
        'edugain' => [
            'cron'         => ['hourly'],
            'sources'      => [
                [
                    'src'          => 'https://mds.edugain.org/edugain-v2.xml',
                    'certificates' => ['/var/www/simplesamlphp/cert/edugain-mds-v2.pem'],
                    'whitelist'    => [
                        'https://auth.sketchengine.eu/entity',   // Sketch Engine
                    ],
                ],
            ],
            'types'        => ['saml20-sp-remote'],
            'expireAfter'  => 60 * 60 * 24 * 4,
            'outputDir'    => 'metadata/metarefresh-edugain/',
            'outputFormat' => 'flatfile',
        ],
    ],
];
EOF

CRONKEY=$(openssl rand -hex 16)
sudo tee config/module_cron.php > /dev/null <<EOF
<?php

\$config = [
    'key'           => '${CRONKEY}',
    'allowed_tags'  => ['hourly', 'daily'],
    'debug_message' => true,
```

```
    'sendemail'     => false,
];
EOF

sudo chown www-data:www-data config/module_metarefresh.php config/module_cron.php
sudo chmod 640 config/module_metarefresh.php config/module_cron.php
for f in config/config.php config/module_metarefresh.php config/module_cron.php; do sudo -u www-data php -l "$f"; done
sudo systemctl reload apache2
```

3.7. Перший запуск і автоматичне оновлення

Розбір агрегату eduGAIN потребує 1–2 ГБ пам'яті. Через веб-URL модуля cron він падає з помилкою `Allowed memory size … exhausted`, тому оновлення запускається з командного рядка з окремим лімітом:

```
cd /var/www/simplesamlphp
time sudo -u www-data php -d memory_limit=2G -d max_execution_time=0 -r '
require "vendor/autoload.php";
print_r((new \SimpleSAML\Module\cron\Cron())->runTag("hourly"));
'
ls -la metadata/metarefresh-edugain/
```

У папці має з'явитися `saml20-sp-remote.php`. Перевірте, що Sketch Engine завантажено разом зі списком запитуваних атрибутів:

```
sudo -u www-data php -r '
require "vendor/autoload.php";
$h = \SimpleSAML\Metadata\MetaDataStorageHandler::getMetadataHandler(\SimpleSAML\Configuration::getInstance());
var_export(array_keys($h->getList("saml20-sp-remote")));
echo PHP_EOL;
var_export($h->getMetaData("https://auth.sketchengine.eu/entity", "saml20-sp-remote")["attributes"] ?? "NO ATTRIBUTES");
echo PHP_EOL;'
```

Щогодинне оновлення (новий агрегат публікується о 45-й хвилині кожної години):

```
( sudo crontab -u www-data -l 2>/dev/null; echo '50 * * * * cd /var/www/simplesamlphp && php -d memory_limit=2G -d
max_execution_time=0 -r '"'"'require "vendor/autoload.php"; (new \SimpleSAML\Module\cron\Cron())->runTag("hourly");'"'"' >
/dev/null 2>&1' ) | sudo crontab -u www-data -
sudo crontab -u www-data -l
```

Без регулярного оновлення метадані Sketch Engine вважатимуться застарілими через 4 дні (`expireAfter`).

---

4. Налаштування Sketch Engine

4.1. Вимоги Sketch Engine

Інституційний вхід (SSO) доступний установам — членам eduGAIN. Атрибути:

Тип	Атрибути
Обов'язкові	eduPersonPrincipalName, schacHomeOrganization, eduPersonScopedAffiliation
Рекомендовані	givenName, surname, displayName, commonName, mail, eduPersonTargetedID

Доступ до ліцензії Sketch Engine визначає за eduPersonScopedAffiliation (`member@<DOMAIN>`).

4.2. Звернення до Sketch Engine

Коли IdP опубліковано в eduGAIN, напишіть команді Sketch Engine англійською і вкажіть:

- назву університету й те, що IdP зареєстровано в eduGAIN через федерацію PEANO;
- IdP entityID: `https://<IDP_HOST>/idp`;
- URL метаданих IdP і посилання на запис у met.refeds.org;
- scope: `<DOMAIN>`;

- перелік атрибутів, які IdP видає Sketch Engine;
- правило доступу: `eduPersonScopedAffiliation = member@<DOMAIN>`;
- технічний контакт;
- прохання прив'язати інституційну підписку до IdP (або описати, що для неї потрібно).

4.3. Тест атрибутів

Sketch Engine попросить пройти тестову сторінку:

1. Відкрийте у вікні інкогніто `https://auth.sketchengine.eu/Shibboleth.sso/Login?target=https://auth.sketchengine.eu/shib_auth_test`.
2. Натисніть **Full access login**.
3. Оберіть свій університет і увійдіть обліковим записом Microsoft 365.
4. Відкриється сторінка з таблицями атрибутів. Надішліть її скріншот команді Sketch Engine.

Очікуваний результат:

- **Mandatory attributes**: усі три заповнені (зелена таблиця);
- **Recommended**: givenName, surname, displayName, commonName, mail заповнені;
- **eduPersonTargetedID** може показувати `None`, бо Shibboleth на боці Sketch Engine не розпізнає формат, у якому SimpleSAMLphp 2.x надсилає цей атрибут. Атрибут необов'язковий, на доступ це не впливає. Не намагайтеся виправити це параметром `'nameId' => true` (див. розділ 1.9, крок 40);
- **Deprecated attributes** (червона таблиця) зі значеннями `None` — норма.

4.4. Перевірка інституційного входу

1. Відкрийте https://auth.sketchengine.eu/#login (або https://app.sketchengine.eu/) → **INSTITUTIONAL LOGIN**.
2. Оберіть свій університет і увійдіть.
3. Має відкритися сторінка **New account**.

Підтвердіть Sketch Engine, що вхід працює. Останній крок — оформлення інституційної підписки, після чого користувачі можуть працювати із Sketch Engine.

4.5. Прибирання після запуску

Видаліть тестовий SP `idp-selftest`:

- запис `'idp-selftest'` з `config/authsources.php`;
- запис `https://<IDP_HOST>/idp` з `metadata/saml20-idp-remote.php`;
- запис `idp-selftest` з `metadata/saml20-sp-remote.php` (файл можна залишити з порожнім вмістом `<?php`).

```
cd /var/www/simplesamlphp
for f in config/authsources.php metadata/saml20-idp-remote.php metadata/saml20-sp-remote.php; do sudo -u www-data php -l "$f";
done
sudo systemctl reload apache2
```

---

5. Типові проблеми

Симптом	Причина	Рішення
`MetadataNotFound` для Entra ID	Файл метаданих у форматі `return [...]`	Використовувати `$metadata['<entityID>'] = [...]` (1.8)
`Could not retrieve the required option 'entityID'`	Не задано `entityID` у `authsources.php`	Додати `entityID` (1.7)
`Endpoint of type … is not an array`	Ендпоінт задано рядком	Масив з `Binding` і `Location` (1.8, 1.10)
`The module 'saml' is not enabled`	У справжньому `module.enable` стоїть `'saml' => false`	Повернути `true`; не правити `config.php` командами `sed` за шаблоном `'saml' =>`, бо він трапляється і в `debug`
Порожній список атрибутів у відповіді IdP	Глобальний `50 => 'core:AttributeLimit'` у `config.php`	Закоментувати його (1.4)
Зайві claims Microsoft у	Немає обмеження атрибутів	Крок 95 у `saml20-idp-hosted.php`

відповіді		
eduPersonTargetedID вкладений сам у себе	`'nameId' => true` у кроці 40	Прибрати параметр
`Allowed memory size … exhausted` при оновленні метаданих	Агрегат eduGAIN не вміщується в 128 МБ	Запуск через CLI з `memory_limit=2G` (3.7)
`composer: requires PHP extension ext-pdo_sqlite`	Немає розширення	`sudo apt install php8.3-sqlite3`
`composer: Could not find package …module-cron`	Модуль `cron` вбудований	Встановлювати лише `metarefresh`
`Failed to restart php8.3-fpm.service: Unit not found`	PHP працює через mod_php	Перезапускати лише Apache
ePPN і scopedAffiliation відсутні на боці SP	Немає `<shibmd:Scope>` у метаданих IdP	Параметр `'scope'` у `saml20-idp-hosted.php` (1.9)

---

6. Обслуговування

- **Сертифікат Entra ID** (`cert/azure_idp.crt`) — оновлювати перед закінченням строку дії в Entra.
- **Сертифікат IdP** (`cert/idp.crt`) — діє 10 років. Заміну узгоджувати з федерацією заздалегідь: новий сертифікат має потрапити в метадані eduGAIN до того, як IdP почне ним підписувати.
- **Метадані eduGAIN** — оновлюються cron щогодини. Перевірка: `ls -la /var/www/simplesamlphp/metadata/metarefresh-edugain/` (дата файлу не старша за кілька годин).
- **Нові сервіси eduGAIN** — додати їхній entityID у `whitelist` модуля metarefresh.
- **Журнали** — для діагностики можна тимчасово ввімкнути `'logging.level' => SimpleSAML\Logger::DEBUG` і `'debug' => ['saml' => true, …]`, але після діагностики обов'язково повернути `NOTICE` і `false`: у режимі налагодження в журнал потрапляють персональні дані.
- **Резервні копії** — перед кожною зміною копіювати файл: `sudo cp <файл> <файл>.$(date +%Y%m%d-%H%M%S).bak`.