

Fondation SAPA, les arts de la scène suisses et Wikidata : point de situation

Durant la première moitié de l'année 2024, plus de 11'000 nouvelles entrées concernant les productions des arts de la scène suisses ont été créées dans Wikidata. Cela a été rendu possible par un [partenariat](#) établi en 2023 entre la [Fondation SAPA, Archives suisses des arts de la scène](#) et l'association [Wikimedia CH](#). Il s'agit de moins 20 % de l'ensemble des productions décrites par la Fondation. Il s'agit donc d'une première étape. Nous allons présenter ici le contexte de cet import, les objectifs de la démarche et les difficultés rencontrées.

Contexte

On peut estimer que les institutions patrimoniales au sens large (archives, bibliothèques et musées) créent et prennent soin de deux types de métadonnées : celles sur les documents présents dans leurs collections et celles sur les contextes qui concourent à l'existence de ces derniers. Par contextes, on entend habituellement les personnes, institutions, événements et lieux ayant un lien avec les documents. Les deux ensembles de données sont inséparables. On ne peut décrire l'un sans l'autre. Toutefois, si le premier groupe a vocation à décrire un document conservé, à terme, sous forme unique dans une seule institution d'archives, le second groupe de métadonnées est partagé avec de nombreuses institutions (et pas seulement patrimoniales). Cette dimension partagée n'a toutefois pas pu vraiment être vécue comme telle jusqu'ici. En effet, les institutions patrimoniales, à quelques exceptions près, ont concentré leurs efforts sur la gestion des métadonnées concernant les documents. Le soin apporté aux métadonnées des contextes est toujours resté limité faute de ressources. Or, les renouvellements technologiques pourraient permettre à présent de changer la donne, en ouvrant la possibilité d'enrichir celles-ci au travers d'une gestion mutualisée de ces dernières. Les données ouvertes liées (*Open Linked Data*), et Wikidata en particulier, semblent en ce sens prometteuses. Les travaux sur les données des contextes engagés par les institutions ne seraient pas forcément supplémentaires mais c'est la forme de leur mise en œuvre qui serait modifiée. C'est en tout cas, le pari que prend la Fondation SAPA qui s'est engagée dans une expérimentation à grande échelle dont nous allons rendre compte ici des premières étapes.

La Fondation SAPA

La Fondation SAPA est à la fois un centre de compétences et une institution de conservation d'archives sur les arts de la scène en Suisse. Constituée en 2017, elle est l'héritière de collections rassemblées par des institutions précédentes, que sont la Collection suisse du théâtre et la Collection suisse de la danse. Elle gère des métadonnées sur ses collections de documents et elle effectue aussi une veille afin de disposer de métadonnées

représentatives sur ce qu'elle a appelé le Panorama des arts la scène suisses, à savoir : quelle production des arts de la scène est jouée par qui, sous quel titre, durant quelle saison et dans quel lieu. Ces données sont collectées, pour les productions professionnelles, au sein d'une base de données appelée le "Repertorium Profi".

Le pré-projet

La Fondation SAPA a identifié Wikidata, dès le lancement en 2017 des travaux de modélisation RDF de ses métadonnées (voir le [rapport d'étape](#) paru à ce sujet début 2024), comme une base de référence externe incontournable. Elle a donc entrepris, depuis cette date, un long et patient travail d'alignement (*matching* en anglais) entre les notices dont elles disposaient sur des personnes ou des groupes et leur équivalent dans Wikidata. Ce travail fait à présent partie des processus obligatoires lors de la création de nouvelles notices. Elle a ainsi référencé dans sa base RDF, au moyen de la propriété [owl:sameAs](#) les correspondances trouvées, mais elle a aussi introduit dans les notices Wikidata un lien vers sa base au moyen de la propriété "identifiant SAPA" ([P8974](#)), créée à sa demande. Plus de 20'000 notices Wikidata disposaient d'une déclaration P8974 avant le lancement du nouveau projet d'import réalisé en 2024 et que nous allons présenter à présent.

Les données

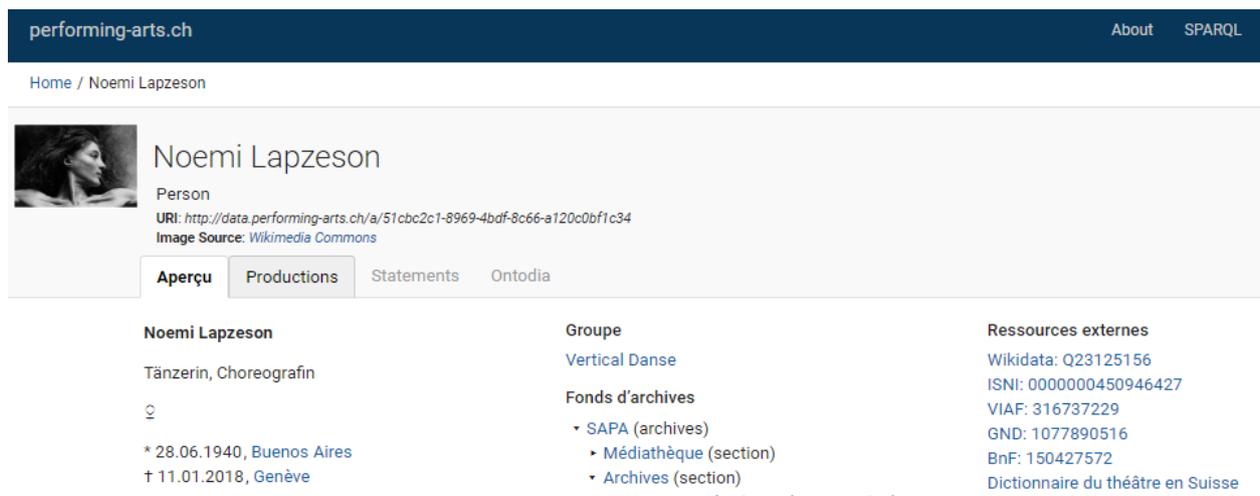
Les données identifiées par la Fondation SAPA comme bonnes candidates à la gestion mutualisée sont celles contenues dans son Repertorium Profi. Cette base de données est le résultat d'un travail de longue haleine de documentation effectué par la Fondation, avec l'aide précieuse de bénévoles. Ce travail est réalisé à la suite de la Collection suisse du théâtre qui l'avait mise en place dans les années 1970. Ce répertoire couvre l'ensemble des représentations données dans une sélection d'une cinquantaine d'institutions suisses représentatives des arts de la scène. Il est réalisé sur la base des programmes, des fiches de spectacle et des recensions parues dans les journaux collectés par la Fondation. Il s'agit d'un répertoire des premières représentations de chaque production dans une institution donnée. Il comporte à jour environ 70'000 entrées et porte sur les années allant du milieu du XIXe siècle à ce jour, avec une couverture plus complète sur les représentations à partir de la seconde moitié du XXe siècle. Il est tenu à jour dans une base FileMaker. Un export de ces données sous forme RDF (licence CC0) est mis à disposition à travers le nouveau portail géré par la Fondation : [performing-arts.ch](#).

Le projet

Les institutions patrimoniales n'ont pas les ressources pour créer et maintenir des métadonnées riches sur les contextes (personnes, institutions, événements et lieux) qui entourent la production des documents qu'elles conservent. Ces données sont intrinsèquement multiformes, foisonnantes et perpétuellement incomplètes. Celles utilisées jusqu'ici par les institutions sont régulièrement pauvres et insatisfaisantes. La mutualisation trouve donc ici tout son sens. Chaque institution apportant une part de l'information, le

résultat au final dépasserait largement ce qui pourrait être atteint de façon isolée. La contribution via Wikidata étant ouverte à toutes et tous, on peut espérer aussi que des particuliers ou les groupes concernés y ajoutent des informations.

L'utilisation de ces informations mutualisées par les institutions peut se faire de multiples manières. La plus simple est l'affichage de ces dernières (sans stockage) dans les notices publiées en ligne par l'institution, au moyen d'une requête établie "à la volée". Ce principe est déjà utilisé dans le portail performing-arts.ch pour afficher des images issues de Wikimedia Commons. Voir par exemple la notice de la chorégraphe [Noemi Lapzeson](#) ci-dessous.



performing-arts.ch About SPARQL

Home / Noemi Lapzeson

 **Noemi Lapzeson**
Person
URI: <http://data.performing-arts.ch/a/51cbc2c1-8969-4bdf-8c66-a120c0bf1c34>
Image Source: [Wikimedia Commons](#)

Aperçu Productions Statements Ontodia

Noemi Lapzeson Tänzerin, Choreografin 📍 * 28.06.1940, Buenos Aires † 11.01.2018, Genève	Groupe Vertical Danse Fonds d'archives <ul style="list-style-type: none">• SAPA (archives)<ul style="list-style-type: none">▸ Médiathèque (section)▸ Archives (section)	Ressources externes Wikidata: Q23125156 ISNI: 0000000450946427 VIAF: 316737229 GND: 1077890516 BnF: 150427572 Dictionnaire du théâtre en Suisse
--	--	--

La requête fonctionne ici via la propriété "image" ([P18](#)) de la notice Wikidata de la personne concernée. Par exemple celle de Noemi Lapzeson : [Q23125156#P18](#).

Suivant ce principe, la Fondation SAPA, a décidé de tester la mutualisation des données sur les productions des arts de la scène en Suisse dont elle conserve les traces. Il s'agit paradoxalement d'un ensemble sur lequel, elle dispose, grâce à son travail de collecte, d'informations très riches. Il s'agit en effet des éléments au cœur de son travail patrimonial. La mise en commun de ce "trésor" lui permet à la fois de mettre à disposition de toutes et tous, de façon facilitée, les informations qu'elle conserve, mais également de poursuivre le travail de collecte à leur sujet via la gestion mutualisée. Elle espère ainsi encourager d'autres institutions ou personnes à compléter ces données et bien sûr à les réutiliser.

Les objectifs

Le premier objectif de la Fondation est que ces données mises en partage via Wikidata puissent être réutilisées par d'autres. Une manière de constater cela sera de suivre l'ajout d'identifiants externes sur les notices créées. Le second objectif est que ces notices puissent être complétées par d'autres, afin qu'en retour la Fondation puisse enrichir ses propres notices de ces nouveaux ajouts. Les principaux ajouts recherchés par la Fondation

sont des images libres de droit pour chacune de ces productions, par exemples les affiches ou des photos des représentations. Voici comme exemple la notice Wikidata de la production Penthesilea (Q122913309) avec une image vers dans Wikimedia Commons à partir d'une copie (<http://data.performing-arts.ch/r/712a8e0a-4f24-4bb4-8015-392d228d60e7>) tirée du fonds d'archives du photographe Fred Erismann conservé par la Fondation SAPA.

Penthesilea. In einem Aufzug (Q122913309)...

Élément [Discussion](#)

Produktion der darstellenden Künste des Stadttheater Bern in der Spielzeit 1935-1936

► [Recoin: Propriétés manquantes les plus pertinentes](#)

► [Plus de langues](#)

Déclarations

nature de l'élément

 production des arts de la scène ...

► [1 référence](#)

image





1553-FO-28-3-DC-2.tif

Import dans Wikidata

Les données RDF issues du Repertorium Profi présentent dans le portail performing-arts.ch sont une sélection des informations originales stockées dans FileMaker. Pour la duplication de ces données dans Wikidata, il a été décidé de débiter par une sélection d'institutions productrices. Un choix des informations facilement mises en partage (voir le modèle de données plus bas) a été effectué en leur sein.

La sélection des données et le modèle de celles-ci pour l'import ont été établis par l'équipe de la Fondation. Un [premier lot](#) avait pu être versé par ses soins en 2023 pour tester la procédure. Les travaux d'import du nouveau lot ont été réalisés par Nicolas Vigneron et Léa Lacroix, sur mandat de Wikimedia CH, entre janvier et juin 2024. Il faut relever que le travail

d'import dans Wikidata d'un lot de plusieurs milliers de données n'est jamais un exercice aisé et la Fondation est très heureuse d'avoir pu bénéficier de cette aide externe. Une [page de présentation du projet](#) a été publiée sur Wikidata.

Sélection

L'import dans Wikidata a concerné les informations collectés par la Fondation sur les productions des cinq institutions suivantes :

- [Stadttheater Basel](#) (période couverte 1834-1968) : 7'861 productions
- [Schauspielhaus Zürich](#) (période couverte 1921->) : 2'550 productions
- [Theatre de Vidy](#), Lausanne (période couverte 1989->) : 766 productions
- [Théâtre de Carouge - Atelier de Genève](#) (période couverte 1958->) : 664 productions
- [Teatro sociale](#) Bellinzona : 64 productions

Modèle de données

Pour établir le modèle de données, la Fondation a suivi les recommandations du [WikiProjet Arts de la scène](#). Le modèle a été ensuite soumis à l'évaluation des participantes et participants de ce WikiProjet. Le modèle de données retenu est le suivant :

- La nouvelle entité A :
 - dispose d'un **libellé**;
 - dispose de plusieurs **descriptions**;
 - **est** (P31) une production des arts de la scène ([Q43099500](#));
 - **a pour titre** ([P1476](#)) une suite de caractères (string)
 - **réalisé(e) ou mis(e) en scène par** ([P57](#)) une entité B;
 - cette entité B est une être humain ([Q5](#));
 - **a pour société de production** ([P272](#)) une entité C;
 - cette entité C **est** un groupe des arts de la scène ([Q105815710](#));
 - **a eu lieu au cours** ([P2348](#)) d'une entité D;
 - cette entité D **est** une saison de théâtre européenne ([Q40008090](#));
 - **est maintenu par le wikiprojet** ([P6104](#)) WikiProjet Arts de la scène ([Q59956276](#))
 - **a comme identifiant SAPA** ([P8974](#)) une suite de caractères (string)

Le libellé est reproduit en quatre langues (français, allemand, italien, anglais). Les descriptions sont basées sur une phrase type traduite dans les mêmes quatre langues utilisées pour les libellés. Concernant la metteuse en scène ou le metteur en scène,

l'information n'est présente, sauf exception, que dans les cas où l'identifiant wikidata était inclus dans les données RDF à importer, c'est-à-dire si un travail de matching avait déjà pu avoir lieu. L'identifiant SAPA permet d'accéder à la notice plus complète présente sur le portail performing-arts, et de pouvoir avoir un suivi des imports réalisés.

Exemples de productions créées selon ce modèle :

- "Leonce und Lena" ([Q124469818](#)) mise en scène par Barbara Frey et produite par le Schauspielhaus de Zürich durant la saison 2011/12.
- "Vive l'Europe" ([Q125021250](#)) mise en scène par Laurence Calame et produite par le Théâtre de Vidy-Lausanne durant la saison 1992/93.
- "Chanta, o unda. Un progetto multimediale con canti romanci e marocchini" ([Q123938228](#)) mise en scène par Sarah Derendinger et produite par le Teatro sociale durant la saison 2012/13.
- "Le Journal d'un fou" ([Q125913992](#)) mise en scène par Françoise Courvoisier et produite par le Théâtre de Carouge durant la saison 1995/96.

Difficultés

La principale difficulté rencontrée est celle des doublons. Il existe deux types de doublons possibles : ceux créés car la production existait déjà dans Wikidata avant l'import, et ceux présents dans les données à importer elles-mêmes. Wikidata offre des moyens simples et très efficaces de fusionner les entités mais il est toujours préférable de limiter ces travaux en prévenant la création de doublons.

Concernant les productions existantes déjà dans Wikidata, il est à noter que sur la même sélection d'institutions productrices, d'autres imports de productions avaient déjà été réalisés avant celui de la Fondation SAPA. Le premier était connu de la Fondation. Il a été piloté par Beat Estermann à partir d'un autre jeu de données. Il s'agissait d'environ 700 productions du Schauspielhaus de Zurich (voir [plus d'informations à son propos](#)). Aucun doublon n'a été créé dans ce cas. Le second groupe de productions préexistantes est issu d'un export très complet réalisée depuis la base "data.kunsten.be" ([Q57323933](#)). Ces dernières sont facilement identifiables via leur déclaration "identifiant Kunstenpunt d'une production" ([P5935](#)). Les notices concernées sont issues de coproductions internationales ayant été jouées en Belgique. Ici les doublons n'ont pas pu être évités car les titres ne correspondaient souvent pas. Les travaux de fusion devront encore être entrepris manuellement. Cela ne concerne toutefois qu'une cinquantaine de productions. Ces deux imports ont été faciles à détecter par le fait qu'ils utilisent la même propriété "a pour société de production" ([P272](#)) que les données importées par la Fondation. Toutefois, il faudrait élargir les contrôles, en analysant d'autres propriétés comme "lieu de la première représentation" ([P4647](#)), ce qui permettrait de corriger aussi l'emploi de cette propriété avec une valeur appartenant à un groupe des arts de la scène et non à bâtiment ou à une salle comme attendu.

Ce qui a été plus compliqué à gérer était la présence de doublons dans les données à importer elles-mêmes. En effet, la base Filermaker "Repertorium Profii" est un recensement

des premières **représentations** dans une sélection d'institutions suisses. Or, l'import dans Wikidata a été fait au niveau des **productions**. Le modèle du répertoire a été établi il y a plus de 40 ans, les coproductions étaient alors pratiquement inexistantes. En ce sens, il était rare qu'une même production se retrouve donnée dans une autre institution sélectionnée par le répertoire. Il n'existait donc dans la base pratiquement qu'une seule représentation pour une production. Ce n'est plus le cas avec les données récentes. Il existe à présent souvent dans le répertoire plusieurs représentations recensées pour une même production. En effet, les coproductions entre grandes institutions sont la norme aujourd'hui. Les données sont donc relevées dans le répertoire lors de la première dans l'institution X et lors de la première dans l'institution Y. Parfois les représentations d'une même production présentent des variantes orthographiques dans le titre. On peut prendre pour exemple "Small g" mise en scène par Anne Bisang durant la saison 2019/20. Les représentations de cette même production se trouvent trois fois (avec deux variantes de titre) dans le Repertorium Profi FileMaker et dans son export RDF dans la plateforme performing-arts.ch. Cette difficulté a été mal identifiée au départ et, malgré des efforts certains de l'équipe mandatée, certaines productions se sont retrouvées dédoublées dans Wikidata. Il faudra établir une stratégie pour les fusionner et pour éviter de reproduire le problème lors des prochains imports.

Par ailleurs, la démarche de la Fondation a pu soulever la question du "contrôle" sur les métadonnées qui ont fait l'objet d'un don à Wikidata. Il faut préciser que ce contrôle par la Fondation n'existe pas *stricto sensu*. En effet, les données transmises sont des doubles. Les données originales restent conservées par la Fondation. Il s'agit d'une relation de complémentarité qui est envisagée. Qu'une notice vienne à disparaître n'aurait aucun impact sur le travail de la Fondation, et celle-ci pourrait être facilement recréée. Il faut aussi préciser que Wikidata proposent des outils de suivi si nécessaire. Au final, l'expérience démontre que les risques principaux dans Wikidata sont les doublons et non les suppressions.

Conclusions

Il existe à ce jour plus de 35'000 notices Wikidata disposant d'une déclaration de correspondance (P8579) avec la base performing-arts.ch. Sur cet ensemble, [plus de 13'000](#) sont des productions qui ont été créées à partir des données de la Fondation.

La Fondation SAPA ne peut que recommander de confier ces travaux d'import à des externes ou à des wikimédien-ne-s en résidence. En effet, l'exercice demande une réelle expérience et de pouvoir s'y plonger de façon concentrée et de façon continue. L'une et l'autre chose manquent habituellement cruellement aux équipes fixes ou temporaires dans les institutions.

Par ailleurs, le premier résultat de cet import est qu'il a permis d'effectuer un contrôle qualité sur les données elles-mêmes de la Fondation. Ce résultat n'avait pas été attendu mais il s'est avéré très précieux. Dès la préparation des données, l'équipe externe a transmis à la Fondation des demandes d'éclaircissements qui ont mis à jour des incongruités dans celles-ci. Puis, suite à l'import, la richesse des données Wikidata, a permis à l'équipe externe d'effectuer certaines vérifications, amenant des mises à jour dans les données d'origine.

Vis-à-vis des objectifs fixés, il faudra attendre les retours suite aux démarches de

communication sur ces premiers imports. L'idée étant d'encourager les institutions partenaires et les groupes des arts de la scène concernés à utiliser ces données et à les compléter. Un vrai bilan pourra être tiré dans une année par exemple.

En faisant don des informations de contexte qu'elle avait rassemblées, la Fondation SAPA espère encourager d'autres institutions à faire de même. Toutes auraient à gagner à enrichir et réutiliser ces données partagées. Les données ouvertes liées et Wikidata, nous offrent, enfin, des outils particulièrement adaptés pour cela.

Septembre 2024

Baptiste de Coulon, data archiviste, baptiste.decoulon@sapa.swiss

Fondation SAPA, Archives suisses des arts de la scène