

LE DATA MANAGEMENT PLAN COST CALCULATOR : L'AUTOMATISATION AU SERVICE DES CHERCHEURS DE L'EPFL

Antoine Masson

Fondateur et directeur de la société Youmi Lausanne

Eliane Blumer

Coordinatrice de l'équipe de gestion des données de recherche
à la bibliothèque de l'École polytechnique fédérale de Lausanne (EPFL)

L'origine du « Data Management Plan (DMP) Cost Calculator » provient d'un besoin exprimé par nos chercheurs lorsqu'ils doivent remplir leur DMP. En effet, de nombreux DMP comportent une partie budgétaire dans laquelle les chercheurs doivent fournir le montant lié aux coûts des données de la recherche pour tout le cycle de vie (de l'acquisition à l'archivage). C'est souvent difficile pour les chercheurs d'obtenir ces informations, car ils doivent aller chercher différents fournisseurs de logiciel, trouver les informations, les calculer, etc. De plus, beaucoup de professeurs ne savent pas à quels fournisseurs s'adresser, ni quelles sont les opportunités qui leur sont offertes.

Pour combler ce manque, nous avons opté pour un projet de site internet accessible à tous, qui centralise un ensemble de fournisseurs liés aux données de la recherche. Ces fournisseurs peuvent être à l'intérieur de l'École polytechnique fédérale de Lausanne (EPFL), comme les services IT (*Information Technology*), ou des fournisseurs externes, privés ou non. L'objectif est principalement de couvrir tous les fournisseurs importants pour le contexte d'un chercheur à l'EPFL.

FOCUS SUR LES SERVICES POUR L'UTILISATEUR

Le Cost Calculator se présente comme une seule page web dynamique, c'est-à-dire qu'elle s'adapte en fonction des choix de l'utilisateur.

L'utilisateur y accède en entrant l'URL du site internet : <https://costcalc.epfl.ch/>. Une introduction sommaire figure en haut de la page ; puis les catégories de fournisseurs, qui reprennent un peu l'ordre du cycle de vie des données : stockage actif, LabNotebook, publication, etc. Se trouvent ensuite une possibilité d'exporter ses calculs, et enfin une aide plus complète des différentes options.

Le principe de fonctionnement est assez simple : dans chaque catégorie, l'utilisateur choisit un fournisseur. Il peut ensuite calculer le coût lié à l'utilisation de ce fournisseur (par exemple, en spécifiant les options dont il a besoin, la durée, etc.). Il est possible également de rentrer un fournisseur à la main, de convertir des devises automatiquement

et d'ajouter autant de fois (ou de retirer) une même catégorie. Le montant de chaque fournisseur est ensuite additionné et affiché en bas de la page.

L'utilisateur peut ensuite exporter un tableau final, résumant l'ensemble de ses choix, pour le coller dans son document DMP. Chaque catégorie et chaque fournisseur possèdent un bouton d'aide qui permet à l'utilisateur d'avoir des informations supplémentaires. L'idée de ce bouton est multiple: elle permet au chercheur de contacter le fournisseur, mais également de se former sur les données de la recherche, s'il le souhaite.

■ **Figure 1. Interface principale du Cost Calculator**

The screenshot displays the main interface of the Cost Calculator. At the top, there are input fields for 'Project Name', 'Project Duration' (set to 5 years), and 'Change Currency' (set to EUR). A blue circle with the number '1' highlights the currency selection area. Below this is a table with columns for 'Line controls', 'Category', 'Provider information', and 'Cost'. The first row shows '1. Active Storage' with 'Google Drive Educ : Online Storage' and costs of 0 CHF and 0 EUR. A tooltip indicates that Google Storage is not recommended as data is stored outside of Switzerland. Below the table, there are options for 'Select a provider' (Google Drive Educ) and 'Options' (Cloud Based illimite...). A 'My Comments' field is also present. The bottom section shows a summary of 'Total Cost for 5 years' with 57750 CHF and 50745 EUR, highlighted by a blue circle with the number '3'. At the very bottom, there are buttons for 'Export HTML', 'Export HTML Source', 'Export Markdown', and 'Export CSV', with a blue circle and the number '4' highlighting the export options.

Line controls	Category	Provider information	Cost
	1. Active Storage	Google Drive Educ : Online Storage	0 CHF 0 EUR
	1. Electronic LabBook	EPFL-SV-IT : SLIMS	56250 CHF 49427 EUR
	1. Database	EPFL-VPSI : MySql	0 CHF 0 EUR
	1. Data Repository	Dryad : Dyrad	1500 CHF 1318 EUR
	1. Collaborative Repository	GitHub : GitHub	0 CHF 0 EUR

Total Cost for 5 years

Total Cost	57750 CHF
Total Cost	50745 EUR

Export HTML | Export HTML Source | Export Markdown | Export CSV

Légende

1. Possibilité de spécifier un nom de projet, une durée, et de changer la monnaie par défaut.
2. Présentation des différentes catégories et choix d'un fournisseur.
3. Total de tout le projet.
4. Possibilité d'export.

L'ENVERS DU DÉCOR : CADRE MÉTHODOLOGIQUE ET ÉLÉMENTS TECHNIQUES

Le principe du calculateur

Dès la conception du projet, l'idée était de faire une application :

- Simple : l'application est une seule page où toutes les informations sont disponibles, mais également l'application est au final une simple page ne nécessitant qu'un serveur web ultrabasique.
- Adaptable : il est entièrement possible d'adapter le Cost Calculator pour une autre institution, de modifier les fournisseurs et les coûts, sans formation particulière ni réel développement.
- Ouverte : le code source est disponible et chacun peut y contribuer ou le réutiliser à sa convenance.

L'application utilise le framework JavaScript open source React¹ qui permet de faire des pages web dynamiques facilement. L'application se compose d'une partie figée (l'application en elle-même) et d'un fichier de configuration. L'ensemble est téléchargé par le navigateur de l'utilisateur et l'application est exécutée et se dessine compte tenu du fichier de configuration. De ce fait, la même application peut être utilisée par une autre institution : il lui suffit de modifier le fichier de configuration par rapport à ses besoins.

Toute l'application s'exécute sur le navigateur de l'utilisateur (en local) ; le serveur ne fait que délivrer une page internet, ce qui ne nécessite finalement qu'un serveur web ultra-élémentaire, voire uniquement un dépôt de fichiers Git ayant la possibilité de créer un serveur web, par exemple GitHub (avec l'option Pages gratuite), si l'institution n'a pas accès à un serveur web.

Configuration et personnalisation

Le fichier de configuration est en quelque sorte le cœur de l'application, il permet de configurer l'apparence du Cost Calculator selon ses propres besoins. C'est un simple fichier texte écrit en JSON, il ne nécessite pas de compétence particulière pour l'utiliser. On peut ainsi configurer :

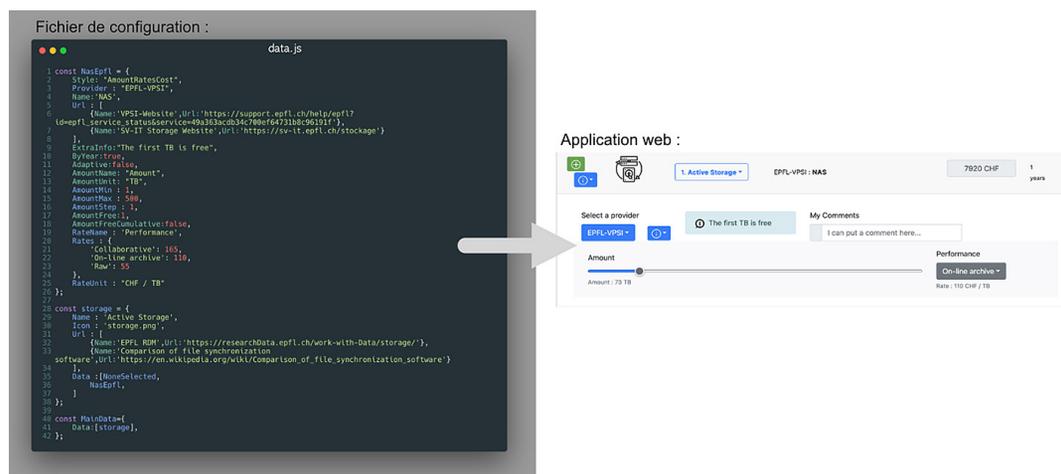
Les fournisseurs	Leurs noms, les boutons d'aides (liens) ainsi que leurs modèles économiques. L'application contient plusieurs modèles économiques, que l'on spécifie selon le fournisseur. Chaque modèle contient ensuite un ensemble de données qu'il faut fournir compte tenu du modèle (coût, catégories, etc.). Si le modèle économique n'existe pas, il est possible d'en implémenter un nouveau, mais cela nécessite une modification du code source.
Les catégories	Leur nom, leur icône, les boutons d'aides (liens) et, bien entendu, leur attribuer des fournisseurs.
L'application	On peut ainsi configurer la devise par défaut (par exemple « franc suisse »), les devises que l'on veut voir apparaître dans les conversions, le logo de l'institution, les e-mails notamment.

1 <https://reactjs.org/>

Une documentation complète est disponible² sur le dépôt GitHub ainsi que des exemples³ de fichiers de configuration.

La configuration étant manuelle, il faut que la personne responsable mette à jour régulièrement les coûts, si par exemple ceux-ci changent.

Figure 2. Principe de la configuration du Cost Calculator



Légende de la figure 2

- Le fichier de configuration (/js/data.js) à gauche, fonctionne comme le principe des poupées russes: on définit des fournisseurs, qui sont intégrés dans des catégories qui, elles, seront intégrées dans l'application à l'aide de la variable MainData. Le résultat de l'application obtenu grâce à ce fichier de configuration est affiché à droite de la figure.
- Dans notre exemple, on crée un fournisseur «EPFL_VPSI» dans la variable NasEPFL (le nom des variables n'a aucune importance). Cette variable est du style AmountRatesCost. Ce style créera une barre de progression verticale avec des catégories à coût variable.
- On définit ensuite les paramètres de la barre (valeur minimale, maximale, incrément, unité), les valeurs des catégories. On peut également spécifier des liens et informations qui seront affichés.
- Ensuite (dès la ligne numéro 28), on définit une catégorie dans la variable storage. Cette catégorie, possède un nom «Active Storage», une icône, des fournisseurs et des liens.
- La variable finale MainData (ligne 40) sera enfin lue par l'application.

Le fichier de configuration complet peut être obtenu ici⁴ et l'application correspondante ici⁵.

2 <https://Github.com/rdmepfl/costcalc/blob/master/docs/data.js-config.md>

3 <https://Github.com/rdmepfl/costcalc/blob/master/docs/data.js-example.md>

4 <https://Github.com/amasson84/costcalc/blob/master/js/data.js>

5 <https://amasson84.Github.io/costcalc/>

Installer le Cost Calculator

- **Serveur web**

Dans le cas de l'utilisation sur un serveur web attribué, l'installation consiste à copier les fichiers sur la racine du serveur web et à éditer le fichier de configuration.

- **Dépôt Git**

Dans ce cas, il faut créer un dépôt Git sur le repository, en important le dépôt du Cost Calculator. En activant la publication web (nommée GitHub Pages sur GitHub par exemple), le Cost Calculator sera disponible à l'adresse HTML ainsi générée.

La modification du Cost Calculator se fait en éditant le fichier de configuration.

DES PISTES DE DÉVELOPPEMENTS POUR LE CALCULATEUR

Le Cost Calculator est relativement complet et offre de nombreuses possibilités. La dernière version intègre une fonctionnalité de tracking (avec Matomo⁶, un logiciel libre de gestion de statistique web) qui permet de connaître plus en détail l'utilisation du calculateur, le nombre d'utilisateurs, un ensemble de statistiques – avec notamment ce que les utilisateurs utilisent le plus sur l'application comme le nom des catégories, les fournisseurs, les fonctionnalités, etc. L'idée était de voir quels sont les fournisseurs et catégories les plus demandés afin notamment d'adapter le contenu du Cost Calculator. L'utilisateur a le choix d'activer ou non ce tracking et aucune donnée plus sensible (comme les coûts, le nom du projet...) n'est de toute façon enregistrée.

Le code source du Cost Calculator est ouvert à tous et disponible en ligne sur un dépôt GitHub. Il est donc possible à tous d'utiliser le Cost Calculator, mais également de le modifier ou de proposer de nouvelles fonctionnalités.

Un développement possible serait de construire une interface « administrateur ». Cette interface permettrait, par exemple, de générer le fichier de configuration de manière simple et graphique.

Le Cost Calculator a également inspiré d'autres services de l'EPFL. En effet, il y a des demandes de la part du Research Office de développer un outil similaire pour planifier les budgets de la recherche et offrir ce service aux chercheurs.

ÉVOLUTIONS ET PROMOTION DU CALCULATEUR AUPRÈS DES CHERCHEURS

Depuis son déploiement, le Cost Calculator a vu plusieurs évolutions. Nous avons notamment rajouté plusieurs monnaies (EUR, CHF, etc.) et nous avons aussi continuellement mis à jour des sources de nouveaux services de l'EPFL.

Concernant la promotion de l'outil, nous avons mené plusieurs actions très diverses au fil des deux ans. Nous avons référencé le calculateur sur notre site web. Dans chaque formation autour des données de recherche, nous parlons du Cost Calculator et nous

6 <https://matomo.org/>

laissons le temps aux participants de se familiariser avec l'outil, si souhaité. Nous envoyons également le lien vers le calculateur à chaque demande de soutien – par exemple, dans un Data Management Plan – qui nous parvient par notre mailing list.

Une autre façon de promouvoir le Cost Calculator est que nous utilisons son format (open source, publique) comme un exemple des bonnes pratiques lors de nos formations pour la publication open data par exemple. En effet, le dépôt GitHub⁷ du projet est ouvert à tous, possède une documentation complète, un README et des métadonnées de citation. Cela permet de montrer un exemple concret lors de nos formations et de donner le lien du dépôt afin que les chercheurs et chercheuses l'utilisent comme exemple de bonne pratique.

Le service est uniquement utilisé sur un plan opérationnel. Actuellement, on ne peut pas parler d'impact sur le plan du pilotage politique.

CONCLUSION

Le Cost Calculator est né d'une demande des chercheurs et chercheuses de notre école polytechnique : comment trouver et calculer les coûts liés aux données de la recherche ? De là, nous avons conçu une application web regroupant différents fournisseurs, leurs informations, et également des moyens faciles de calculer tous ces coûts de manière assez précise. Il est, de plus, possible d'exporter son calcul, notamment pour l'intégrer dans un DMP. Le Cost Calculator est open source, c'est-à-dire que tout le monde peut le modifier, l'utiliser, et le partager.

Pour nous, le Cost Calculator était un bon moyen de démontrer que ce genre de développement d'outil de soutien trouve sa place dans une bibliothèque scientifique et aussi son utilité. Il nous a également permis de montrer que des tâches souvent très administratives qui entourent la préparation d'une demande de financement d'un projet de recherche peuvent être automatisées et rendues plus accessibles. Nous espérons que des outils de ce genre vont continuer à être développés dans les bibliothèques. Pour cela, il est certain qu'il faut une synergie entre de nombreuses compétences tant au niveau de l'informatique que des métiers liés aux spécificités des données de la recherche et du monde de la recherche en général. •

Quelques chiffres clés sur l'utilisation du calculateur

Nous n'avons pas fait une analyse statistique fine, nous ne connaissons donc actuellement pas le degré d'adoption, mais nous pouvons faire quelques estimations. Nous pensons qu'environ un dixième de chaque projet de recherche passe par le calculateur, avec une tendance croissante. Les statistiques issues des visites (voir point 4) montrent qu'il y en a eu 140 en 2020 et 240 en 2021 soit environ 380 visites sur deux ans (depuis le lancement).

De ces visites, 60 % entrent à l'adresse directement, 33 % par le site web de l'équipe des données de la recherche (<https://researchdata.epfl.ch>) et le reste y arrive via un moteur de recherche ou d'autres chemins.

7 <https://Github.com/rdmepfl/costcalc>

Dans les catégories les plus utilisées sur les deux ans, nous voyons :

- Active storage (41 % des *events*) ;
- Data repositories (17 %) ;
- Database, LabNotebook et Code repo ont un score similaire (environ 14 % chacun).

Les fournisseurs les plus utilisés sont :

- VPSI (22 %), notre service central pour toutes ressources informatiques de l'EPFL ;
- Switch (9 %), un service d'infrastructure informatique en Suisse ;
- Zenodo (8 %), le dépôt de données du CERN.

À noter que 12 % ont utilisé l'option « Manual » permettant d'entrer manuellement un fournisseur.

Cela nous semble cohérent considérant le fait que ce sont tout principalement les services de stockage qui coûtent dans un projet de recherche et qui sont nécessaires pour tous les projets de recherche.

À noter que ces données ne représentent pas les personnes qui ont un bloqueur de trackers, ou qui renoncent au tracking.

Présentation des auteurs

Antoine Masson⁸ est citoyen suisse et fondateur de Youmi Lausanne⁹ (Suisse). Il est diplômé d'un doctorat en physique de l'université de Paris-11. Antoine Masson a effectué deux postdocs à l'EPFL et à l'université de Berkeley en Californie (États-Unis). Il a ensuite travaillé à la bibliothèque universitaire de l'EPFL, à Lausanne, en qualité de spécialiste des données de la recherche. Maintenant, il se consacre à son entreprise, spécialisée dans le développement de solutions et d'équipements pédagogiques, le développement de logiciel et la formation.

Eliane Blumer est spécialiste de système d'information formée à Genève (Bachelor of Science) et à Olten (Master of Science) en Suisse. Elle a une expérience dans diverses bibliothèques et services IT ainsi que dans des projets scientifiques. Elle a donné des formations autour de nombreux sujets liés aux sciences de l'information, comme le web sémantique, les tests d'utilisabilité ou la gestion des données de recherche. Depuis 2017, elle travaille à la bibliothèque de l'EPFL où elle coordonne l'équipe de gestion des données de recherche. Elle est membre de la Research Data Alliance (RDA), de la communauté du soutien Research Data Management (RDM) en Suisse, et de Bibliosuisse.*

Résumé

Les chercheurs recevant des financements publics pour mener leurs travaux sont soumis à un ensemble d'obligations. Ainsi, il leur faut notamment compléter des Data Management Plans (DPM), lesquels incluent une partie budgétaire. Mais, rassembler des données disséminées dans une multiplicité de silos constitue une difficulté supplémentaire pour les chercheurs : où trouver l'information, comment effectuer le calcul

8 <https://www.youmi-lausanne.ch/wp-content/uploads/2021/11/cv-fr.pdf>

9 <http://www.youmi-lausanne.ch>

du coût ? Pour répondre au besoin des chercheurs, l'équipe de la bibliothèque de l'École polytechnique fédérale de Lausanne (EPFL) a conçu un DMP Cost Calculator, se présentant sous la forme d'un site internet accessible à tous, qui centralise un ensemble de fournisseurs liés aux données de la recherche.