

ARIANE - compte rendu du l'atelier chaîne éditoriale

La première réunion de l'atelier sur les chaînes éditoriales a eu lieu fin septembre en visio afin pour faire un premier état des lieux. Il a réuni une dizaine de personnes, dont des chercheurs et chercheuses mais aussi des ingénieurs issu de pôles d'appui à la recherche ou de laboratoires. En deux mois il y a eu assez peu d'avancées mais un point d'étape est envisagé au début de l'année 2024, possiblement en présentiel (encore à organiser).

On est assez liés à l'atelier scripts et d'ailleurs on verra si à terme on va vers un fusion des groupes (en fonction de comment on avance).

A ce stade je me contenterai de dresser l'inventaire des principales questions soulevées et de faire un premier bilan sur les discussions qu'on peut regrouper en 5 points :

- Question des outils
- Question de l'impact environnemental
- Question de l'édition numérique
- Question humaine
- Question de l'éparpillement

Question des outils

Parler de chaîne éditoriale soulève très vite des questions liées aux outils qui permettent de construire cette chaîne.

Il en **existe beaucoup, vraiment beaucoup, potentiellement complémentaires et parfois franchement concurrents**, qui sont utilisés ou utilisables dans le cadre d'une chaîné éditoriale, qu'ils soient développés au sein d'institutions de recherche ou non, qu'ils soient privés ou open sources, qu'ils soient génériques ou développés au cas par cas en fonction des besoins et des projets. Quoi qu'il en soit, on constate souvent une **certaine confusion dans leur périmètre** et leurs fonctionnalités et l'un des besoins exprimé pendant cette première réunion est celui de **mieux référencer ces outils pour mieux s'y repérer mais aussi pour mieux éclairer leur utilisation**, leur limitations ou leurs possibilités.

Malgré cela, la question de **ne pas se focaliser sur l'outil est très vite ressortie**. Ce n'est pas l'outil qui doit être le plus **important mais le choix des standards, des schémas et de la modélisation**.

Et pour cela un point important qui est beaucoup ressorti c'est **l'utilisation de la TEI** comme vocabulaire commun, bien que celle-ci implique des schémas spécifiques à chaque projet éditorial.

Alors, on a quand même eu une question sur l'intérêt ou non de la TEI notamment dans le cadre de projets très spécifiques mais la majeure partie du groupe a souligné l'importance au minimum d'une conversion vers la TEI dans le cadre des dépôts sur des entrepôts, pour favoriser la réutilisation.

Force est de constater que nous n'avons que très très peu envisager la question de l'IA dans le cadre des chaînes éditoriales, ce qui est peut-être un problème mais cela nous permettra peut-être de nous rapprocher d'un autre atelier ou GT dans le cadre du consortium.

Question de l'empreinte écologique de nos pratiques

Parlant d'IA justement, la question des outils a entraîné au sein du groupe des remarques sur **l'empreinte écologique de nos pratiques**. Bien sûr, on pourra toujours arguer qu'une édition en TEI n'est rien par rapport à un voyage commercial dans l'espace. Mais on peut aussi se dire que c'est le rôle des sciences humaines de se poser ce genre de questions et de réfléchir à des solutions.

On en revient aux outils : tous ne se valent pas en terme de consommation, et on sait par exemple que l'entraînement de modèles HTR est particulièrement énergivore.

Mais sans aller vers le plus criant avec l'HTR ou l'IA, on peut aussi se demander, par exemple, si ça vaut le coup de maintenir un site dynamique pour une édition qui, presque par définition, devrait être la plus statique possible pour pouvoir citée. Et puis une interface de consultation basée sur un logiciel, c'est potentiellement une durée de vie limitée (maintenance des serveurs, des outils, etc.)

Ces questions continueront à être creusées, mais quelques pistes ont déjà été évoquées comme le fait de **favoriser l'édition sous l'angle d'un dépôt (par exemple dans nakala)** avec une documentation détaillée et un schéma, ou encore à rendre les site web statiques au maximum en créant autant de **pages html en dur que** pages appelées dynamiquement.

Questions des humains

Autre point évoqué lors de la réunion : les humains, qui travaillent à la réalisation et à la conservation des ces éditions numériques.

Déjà, une **chaîne éditoriale complexe implique des compétences nombreuses**. Apprentissage du vocabulaire TEI et du XML, des schémas, utilisation de XPATH, de scripts XSLT, Xquery ou Python, puis pour la mise en ligne, de serveurs web (les fameuses VM chez huma-num), pourquoi pas d'intégration continue sur Gitlab, et si rajoute l'ATR en entrée de chaîne, les lignes de commandes. Bref, un panel de compétences de plus en plus large, chacune n'étant pas forcément quelque chose qu'on apprend sur un coin de table (en tout cas pour le faire bien).

Ceci pose la **question du coût d'entrée dans l'univers de l'édition numérique pour un bétien**. Sauf que la question ne se pose pas de la même manière dans le cas où l'éditeur/trice dispose ou non d'un pôle d'appui, ou encore si l'édition est le coeur de son étude ou un accompagnement à son étude.

Par ailleurs **trop de complexité et de technicité implique une plus grande difficulté à maintenir ce qui a été produit**. Et en ce qui concerne la maintenance, on retombe sur la question des personnes capables de s'en occuper sur le long terme. La maintenance, plus que la phase de réalisation, est clairement dépendante du nombre de postes d'appui à la recherche sur contrats longs (i.e. titulaires).

Pour rester sur cette question des postes, on sait qu'aujourd'hui les humanités numériques tournent majoritairement avec des postes d'appui temporaires ou précaires. Sans nous éloigner de la question centrale de cet atelier, c'est quand même un aspect important à prendre en compte et à interroger quand on fait des sciences humaines.

Et on se rend compte aussi que ces situations temporaires créent beaucoup de turn over, donc beaucoup d'énergie dépenser à former et à re-former sans qu'on puisse vraiment s'appuyer sur des compétences qui restent et qui s'enrichissent.

Enfin, en 10 ou même 20 ans, l'édition numérique a énormément changé, et elle continue à le faire avec l'IA.

Pour toutes ces raisons, le groupe souhaite s'intéresser à une **meilleure définition des métiers et des compétences dans l'édition numérique**. Le fait que pas mal de chercheurs soient de plus en plus touchés à tout veut-il dire que désormais on attend ce niveau de compétence de la part de chacun ?

Ou au contraire, faut-il mieux définir ce qui relève de la recherche et de l'ingénierie, mais sans pour autant cloisonner les métiers.

Et puis bien sûr **insister sur la formation**

Question de l'édition en elle même

Alors bien sûr quand on aborde la question de la chaîne éditoriale on pense aussi à l'édition numérique qui en découle, et on se demande comment on peut définir cela au mieux. A quel moment le travail éditorial commence-t-il et se termine-t-il, et comment bien ou lieux définir les différentes étapes. Ceci rejoint la question humaine, en tout cas celle des compétences, puisque à chaque étape peuvent être sollicités des compétences différentes et, sur l'ensemble de la chaîne, très variés.

A été également soulevée la **question des bonnes pratiques et des recommandations**, ce qui nous ramène à l'idée d'"éditions numériques de qualité". Plusieurs personnes ont évoqué le cas de nombreuses éditions en cours, dans un stade intermédiaire, pas vraiment 100% dans les clous mais qui mériteraient sans doute d'être diffusées plutôt que de rester dans le tiroirs, ou plutôt les disques durs. Peut-on envisager plusieurs niveaux de qualité, plusieurs niveaux de dépôts et de diffusion ?

Là aussi, un précédent groupe de travail au sein du consortium CAHIER avait réalisé un travail sur cette question des différents niveaux de transcription, d'édition, etc. travail sur lequel nous avons prévu de nous appuyer.

Question de l'éparpillement

Cette dernière remarque me permet de raccrocher au dernier point, celui de l'éparpillement, qui a été soulevé par plusieurs participant.e.s

Tout d'abord, le groupe réunissait des personnes ayant déjà participé à des groupes relativement proches, qui avaient déjà travaillé sur ces questions ou qui le faisaient actuellement. On peut citer le groupe de travail Scripto du RnMSH, ceux du cluster 5 de biblissima+ dont le travail fait écho à cet actuel GT3 du consortium ARIANE. Si on élargit la focale, on retrouvera aussi l'ancien consortium Cahier (ancêtre d'Ariane), l'ancien GT TEI/Nakala ou encore le GT OLIO - ceci pour les groupes à échelle nationale - mais aussi ce qui a été fait dans le cadre de MSH comme du côté du

PDN et de Métopes à Caen, de la MSH Val-de-Loire ou plus récemment de la MISHA en Alsace ou encore dans le cadre de recherches plus individuelle notamment dans le cadre de thèse (Pierre Yves Buard, Ariane Pinche, et plus récemment Floriane Chiffolleau ou Alyx Chagué).

On a donc beaucoup de gens qui se sont déjà posés ces questions, où en tout cas, dont les questions font en partie écho à celle qu'on se pose aujourd'hui. Et ces personnes ou groupes ont également produit beaucoup de documentation.

En tout cas on peut légitimement s'interroger sur la multiplication de ces groupes. Sommes nous trop nombreux sur les mêmes questions pour travailler efficacement : c'est possible, mais dans ce cas il convient d'éviter l'effet silo, ce qui n'est pas évident lorsque les agenda se remplissent d'initiatives et de groupes de travail parallèles. Dans ce cas comment nous organiser efficacement ? (je ne sais pas) Il y a t-il trop de documentation pour qu'on soit en mesure de se la réapproprier ? Le risque en tout cas c'est de créer des groupes concurrents, que cela soit ou non conscient, ce qui apparaissait relativement négatif aux personnes présentes ce jour là qui affichaient plutôt l'envie de travailler dans un dynamique commune.

Éparpillement des groupes donc, de la documentation mais aussi des outils comme dit plus haut

conclusion

Je terminerai cet inventaire en listant les quelques points d'actions ou livrables qui ont été proposés par le groupe :

- constitutions d'une bibliothèque partagée pour faire un état de l'art (sur Zotero)
- faire cet état de l'art, compiler la documentation pour rebondir sur l'existant
- élaborer une cartographie des outils utilisés dans les chaînes de production (pour y voir plus clair), ce qui complète l'atelier sur les scripts
- cartographier, recenser les différentes activités métiers compétences en lien avec l'édition numérique

Et bien sûr, nous pourrions compléter ou réviser ces objectifs au cours de la discussion qui suivra le prochain exposé.