



HAL
open science

Vers la science ouverte : de l'archivage à la diffusion des données de la recherche à partir de l'API de Nakala – Étude de cas en islamologie classique

Adrien de Jarmy

► To cite this version:

Adrien de Jarmy. Vers la science ouverte : de l'archivage à la diffusion des données de la recherche à partir de l'API de Nakala – Étude de cas en islamologie classique. Techniques au Top - Techniques émergentes appliquées au monde de la recherche en histoire et archéologie, 2024, 10.58079/w0yb . hal-04527738

HAL Id: hal-04527738

<https://hal.science/hal-04527738v1>

Submitted on 30 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Vers la science ouverte : de l'archivage à la diffusion des données de la recherche à partir de l'API de Nakala. Étude de cas en islamologie classique

Résumé : Le développement croissant des humanités numériques pousse les chercheurs à visualiser davantage les données de leurs travaux. Dans le domaine des études historiques et critiques, telles que l'islamologie classique ou l'étude du droit musulman (fiqh), la création automatique de chaînes de transmission du hadith permet de produire de nombreux diagrammes complexes, qui sont difficiles à insérer dans un format papier traditionnel. En utilisant de manière innovante la plateforme de données Nakala (Huma-Num), il devient possible de générer des mini-URL et des codes QR afin de diriger les lecteurs vers ces schémas sur un appareil informatique, facilitant ainsi l'interaction entre la publication papier et l'édition numérique.

Mots-clefs : humanités numériques ; science ouverte ; données de la recherche ; optimisation des ressources numériques ; islamologie

Lien vers l'article : <https://technicotop.hypotheses.org/4457#more-4457>.

Sommaire

1. Problèmes techniques liés à la présentation et la diffusion des données visuelles de la recherche	2
1.1. Objectif du travail.....	2
1.2. Présentation de l'entrepôt en ligne Nakala.....	3
2. Se référer aux données déposées dans Nakala dans ses publications : mini-URL, codes QR et solutions hybrides.....	4
2.1. Intégrer une mini-URL grâce à l'API de Nakala	4
2.2. Générer un code QR associé à une URL d'intégration/de téléchargement	5
3. Quelques limites dans la présentation visuelle des données	6
Conclusions	7
Figures.....	8
Fig. 1 Exemple de diagramme complexe présenté dans la thèse	8
Fig. 2 Utilisation de l'Url d'intégration à partir de l'API de Nakala	9
Fig. 3 Présentation « hybride » avec code QR et mini-URL associés.....	10

1. Problèmes techniques liés à la présentation et la diffusion des données visuelles de la recherche

1.1. Objectif du travail

À l'image des discussions qui ont alimenté la mise au point de la méthode historico-critique pour étudier la réception des Écritures et la figure de Jésus dans le christianisme, les historiens des débuts de l'Islam débattent depuis plus d'un siècle de l'approche à adopter pour reconstituer la biographie du Prophète Muḥammad. Par ailleurs, depuis les premiers travaux des orientalistes, la recherche s'est focalisée sur l'étude du personnage historique et a délaissé l'examen des représentations variées du Prophète qui ont émergé à différents moments de l'histoire. L'objectif de mon travail de doctorat a constitué à établir une méthode qui permette de dater et de localiser la mise en circulation de récits associés ou attribués à Muḥammad dans la tradition islamique (*sīra-maḡāzī*, hadith et *tafāsīr*) entre le I^{er}/VII^e et le III^e/IX^e siècle au Proche-Orient (Arabie, Syrie-Palestine, Irak-Iran, Égypte), et de replacer ces discours dans leurs contextes historiques respectifs et au sein des débats qui animent les premiers savants musulmans.

Ma méthode de travail a principalement constitué à recenser un échantillon d'environ 3000 paroles ou hadiths attribués à Muḥammad, à ses Compagnons (*al-Ṣaḥāba*) ainsi qu'à la génération des Successeurs (*al-Tābi'ūn*) dans une base de données programmée en langage SQL, à l'aide de la version pour particulier du logiciel de gestion de base de données open source MySQL¹. Selon le sujet abordé dans le manuscrit de la thèse, j'ai ensuite pu programmer des requêtes en SQL afin d'extraire tous les hadiths sur un thème donné, ainsi que les autorités responsables de la transmission de ces paroles. Au total, plus de 1200 transmetteurs au Proche-Orient ont pu être identifiés grâce à cette méthode. Plus d'une centaine de diagrammes de transmission des hadiths ont ensuite été réalisés à l'aide du logiciel de cartographie conceptuelle CMap Tools². La présentation de ces diagrammes était absolument essentielle pour mon travail car toute la méthode de datation reposait sur l'identification d'un certain nombre de « cas cliniques » de diagrammes, c'est-à-dire de modèles récurrents dans la transmission du hadith. Or, ces diagrammes comportent régulièrement un nombre important de chaînes de transmissions, parfois plus de 30 ou 40, ce qui posait plusieurs problèmes pour leur intégration

¹ <https://www.mysql.com/fr/>.

² <https://cmap.ihmc.us/>.

Adrien de Jarmy,
ATER, GEO, Université de Strasbourg
Docteur de Sorbonne Université

en format papier (**fig. 1**). Une solution consistait à les placer seulement en annexes à la fin du manuscrit, mais cela aurait rendu la lecture particulièrement fastidieuse pour le jury, alors que le cœur de l’argumentation repose sur les diagrammes. Il fallait donc trouver un moyen de renvoyer le lecteur vers ces diagrammes au fil de la lecture.

1.2. Présentation de l’entrepôt en ligne Nakala

Nakala est un entrepôt conçu pour la gestion, la préservation et le partage des données numériques dans le domaine des sciences humaines et sociales. Il a été développé par Huma-Num, une très grande infrastructure de recherche (TGIR) mit en place par le Ministère de l’Enseignement supérieur en France pour soutenir la recherche scientifique en fournissant des services et des ressources numériques variés³. Nakala est avant tout un entrepôt numérique, dont l’objectif est de permettre l’archivage des données sur le long terme ainsi que d’organiser la gestion des métadonnées de toutes sortes. Contrairement à de nombreux entrepôts en ligne privés, Nakala a pour principal avantage de respecter les principes du FAIR, la description des données suivant le standard Dublin Core⁴, largement reconnu sur le web. Par ailleurs, les métadonnées sont rendues disponibles de manière interopérable grâce à l’API de Nakala et au protocole OAI-PMH⁵, facilitant ainsi leur exploitation ou leur référencement par des agrégateurs externes tels que le moteur de recherche spécialisé ISIDORE. En somme, les données déposées dans Nakala seront facilement retrouvables et utilisables des années plus tard après leur dépôt, tant que les serveurs Huma-Num continueront de fonctionner.

Précisons d’emblée que l’accès à Nakala est restreint aux chercheurs associés à un établissement de recherche français. Pour y accéder, il est nécessaire de prendre contact avec un référent Huma-Num afin d’obtenir un identifiant HumanID, qui permettra de se connecter à la plateforme⁶. En général, l’utilisation de Nakala intervient plutôt à la fin du projet de recherche, dans le cadre de la mise au point d’un plan de gestion des données (PGD). Toutefois, il est possible de s’en servir pour présenter les données visuelles de ses propres recherches dans le cadre d’un travail de doctorat ou même de la publication d’un article scientifique.

³ <https://www.nakala.fr/>.

⁴ Standard de description de données facilitant leur interopérabilité. Il s’agit du format obligatoire pour le protocole OAI-PMH.

⁵ Open Archives Initiative Protocol for Metadata Harvesting. Il s’agit d’un protocole d’échange de données.

⁶ <https://humanid.huma-num.fr/>.

2. Se référer aux données déposées dans Nakala dans ses publications : mini-URL, codes QR et solutions hybrides

2.1. Intégrer une mini-URL grâce à l'API de Nakala

Le procédé le plus simple que nous avons trouvé consiste à copier-coller dans son article l'URL d'intégration et/ou de téléchargement des données entreposées à partir de l'API⁷ de Nakala (**fig. 2**). L'URL d'intégration permet au lecteur d'accéder à la donnée et de la visualiser. L'URL de téléchargement permet au lecteur de télécharger le document sur son propre appareil. L'intégration d'une URL complète reste toutefois contraignante, notamment dans le cas d'une publication sur papier. Le lecteur doit en effet retranscrire une adresse particulièrement longue dans son navigateur. Dans le cas où le chercheur souhaiterait se référer à plusieurs documents, ce procédé peut devenir rapidement fastidieux et décourager le lecteur de les consulter. Nous avons trouvé là aussi, deux moyens de contourner ce problème :

La première solution que nous avons adoptée consiste à créer une mini-URL à partir de l'URL d'intégration ou de téléchargement. De nombreux outils gratuits en ligne existent pour créer une mini-URL, et je me suis en effet d'abord tourné vers le site vu.fr pour créer les miennes⁸. La durée de validité des mini-URL n'est toutefois pas connue, si le site vient à être hors-ligne, alors celles-ci ne seront plus valables. Fort heureusement, la plateforme Huma-Num fournit son propre outil appelé HNURL, qui procure aussi l'avantage de ne pas voir le nom d'un site tiers s'afficher dans l'URL. L'outil se charge de réduire sensiblement la taille de l'URL, ce qui permet de l'intégrer plus facilement en note de bas de page, ou même dans le corps du document, sans que celle-ci n'occupe plusieurs lignes de textes. En cliquant sur la mini-URL dans un fichier PDF, le lecteur ouvrira automatiquement une autre fenêtre et pourra consulter la donnée. Lors de la consultation de la donnée, le lecteur peut opérer des gros plans ou même faire pivoter l'image. En revanche, sur un format papier, le lecteur devra tout de même retaper la mini-URL dans la barre d'adresse du navigateur.

⁷ Interface de programmation de l'application. « Interface logicielle qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service afin d'échanger des données et des fonctionnalités », voir [https://www.cnil.fr/fr/definition/interface-de-programmation-dapplication-api#:~:text=Une%20API%20\(application%20programming%20interface,des%20donn%C3%A9es%20et%20des%20fonctionnalit%C3%A9s.](https://www.cnil.fr/fr/definition/interface-de-programmation-dapplication-api#:~:text=Une%20API%20(application%20programming%20interface,des%20donn%C3%A9es%20et%20des%20fonctionnalit%C3%A9s.)

⁸ <https://vu.fr/>.

2.2. Générer un code QR associé à une URL d'intégration/de téléchargement

La seconde solution consiste à créer un code QR (ou code à réponse rapide) associé à l'URL d'intégration ou de téléchargement fourni par Nakala ; voire à la mini-URL elle-même. La création d'un code QR rend la consultation des données encore plus simple : le lecteur n'a qu'à « flasher » le code d'un fichier PDF ou d'un document imprimé sur papier avec une tablette ou même un smartphone. La plateforme Huma-Num ne fournissant pas d'outil pour générer des codes de ce type, il faut se tourner vers des outils en ligne tels que QR Code Generator⁹ ou QRCode Monkey¹⁰. Ces deux sites proposent de réaliser gratuitement des codes QR personnalisables. Ce système est particulièrement efficace pour diffuser des données difficilement intégrables dans un fichier PDF ou un document imprimé, comme des diagrammes complexes, des cartes détaillées en couleur ou tout type d'iconographie.

Cette solution a également l'avantage de proposer une véritable interaction entre les supports papier et numérique. Le lecteur peut profiter du support papier pour lire la démonstration dans de bonnes conditions et surligner le texte manuellement, tout en gardant un œil sur les diagrammes grâce à un ordinateur, une tablette, ou même un smartphone si l'écran de celui-ci est assez grand. Il est aussi possible de guider le lecteur dans le corps du texte, en lui proposant de réaliser un agrandissement sur une partie plus spécifique d'un diagramme ou d'une carte. En somme, en plus de permettre l'archivage de ses données et de les mettre à la disposition du plus grand nombre grâce à Nakala, cette méthode ouvre des vraies potentialités en termes de présentation et même d'argumentation scientifique.

Un autre avantage du code QR réside dans son fonctionnement même. Contrairement à la mini-URL qui renvoie l'utilisateur vers un serveur tiers qui a généré cette adresse, le code QR contient lui-même directement les informations. L'adresse URL originale est ainsi directement encodée, ce qui rend théoriquement les codes QR utilisables à vie, tant que le serveur qui héberge les données reste accessible. À ce sujet, on pourrait envisager que dans le futur, la plateforme Huma-Num se dote de son propre générateur de codes QR pour éviter aux chercheurs d'avoir à se reposer sur des sites tiers.

Il est aussi possible d'envisager des solutions hydriques. Dans le manuscrit de la thèse, nous avons ainsi fait le choix d'intégrer dans le corps du texte une mini-URL ainsi qu'un code QR pour faciliter l'accès aux données à partir d'un fichier PDF ou du support papier (**fig. 3**).

⁹ <https://fr.qr-code-generator.com>.

¹⁰ <https://www.qrcode-monkey.com/fr/>.

Adrien de Jarmy,
ATER, GEO, Université de Strasbourg
Docteur de Sorbonne Université

Les diagrammes ont également été placés à la fin de chaque chapitre, afin que le lecteur puisse tous les consulter d'une traite sans avoir à « reflasher » chacun des codes précédents.

3. Quelques limites dans la présentation visuelle des données

Malgré ces possibilités, Nakala reste avant tout un entrepôt numérique, et l'emploi que nous en faisons pour la diffusion de la recherche relève en quelque sorte d'un usage détourné de la fonction originelle de cet outil. Ainsi, la première limitation concerne la présentation relativement austère de la plateforme, réservée à un public de chercheurs et d'habitues des outils numériques. Pour présenter et vulgariser l'ensemble de ses données au public, Nakala dispose d'une extension nommée « Nakala_Press », censée permettre la création d'un site web à partir de données regroupées en collection¹¹.

Si cet outil est à première vue une très bonne idée, celui-ci offre relativement peu de choix dans la présentation des données et n'est de fait, pas très attrayant. Un module de rédaction en HTML-CSS, à l'image d'un WordPress permettrait de combler en partie ce problème. De fait, les chercheurs français préfèrent toujours exposer leurs projets par l'intermédiaire d'un carnet Hypothèse, dont la forme ressemble de fait plutôt à un blog. À l'étranger, les institutions utilisent notamment Omeka, un outil bien plus flexible, mais qui demande des connaissances plus étendues en informatique. L'utilisateur doit apprendre à créer une base de données relationnelles en MySQL, laquelle servira à stocker les données présentées sur le site. En outre, l'utilisateur doit soit payer un abonnement à un tiers pour pouvoir héberger son site et le mettre en ligne, soit se coordonner à son institution de recherche pour l'héberger, ce qui pose là aussi des questions concernant le référencement du site internet.

Un second inconvénient concerne le statut des données une fois celles-ci déposées dans Nakala. En effet, leur modification n'est plus possible une fois les données rendues publiques. Or, s'il est possible de faire des erreurs, on peut aussi envisager que le chercheur souhaite les actualiser. Ce problème s'est posé à nous lorsque, après la thèse, il a fallu enrichir certains diagrammes pour la publication d'articles dans des revues scientifiques. La seule solution consiste alors à déposer les données dans Nakala, qui les considérera comme de nouvelles données individuelles des précédentes.

¹¹ Pour plus d'informations, voir le tutoriel Huma-Num. (2021, 30 septembre). *Module 5 Diffuser ses données avec NAKALA_PRESS - Présentation*, Laurent Capelli, in *Gérer ses données en SHS avec les services et outils proposés par la TGIR Huma-Num*. [Vidéo]. Canal-U. <https://www.canal-u.tv/107976>. (Consultée le 14 janvier 2024).

Conclusions

L'évolution des humanités numériques engendre de vraies transformations dans la façon de diffuser les données de la recherche. Dans des disciplines telles que l'islamologie classique, caractérisées par la nécessité de visualiser de manière approfondie les données, l'exploitation novatrice de plateformes telles que Nakala (Huma-Num) revêt une importance stratégique. L'adoption de stratégies telles que l'utilisation de mini-URL et de codes QR permet de faciliter l'accès aux schémas et aux données visuelles complexes, facilitant ainsi l'interaction entre les formats papier et numérique. Cependant, malgré ces avancées notables, des défis subsistent, notamment en ce qui concerne l'esthétique de la présentation et la flexibilité des outils disponibles. Nakala, bien qu'étant un entrepôt numérique robuste, nécessite des améliorations supplémentaires pour répondre pleinement aux exigences diversifiées des chercheurs. La méthode que je propose résulte en effet d'un usage détourné de la plateforme.

Néanmoins, cette étude de cas en islamologie classique illustre ainsi de manière le potentiel considérable des technologies numériques pour la promotion de la science ouverte, tout en soulignant la nécessité continue d'innovation et d'adaptation pour exploiter pleinement ces possibilités dans un contexte académique évolutif. En fin de compte, l'intégration harmonieuse de l'archivage et de la diffusion des données de recherche via des plateformes comme Nakala représente une étape vers le développement d'outils plus appropriés pour faciliter l'interaction entre les formats papier et numérique.

Figures

Fig. 1 Exemple de diagramme complexe présenté dans la thèse

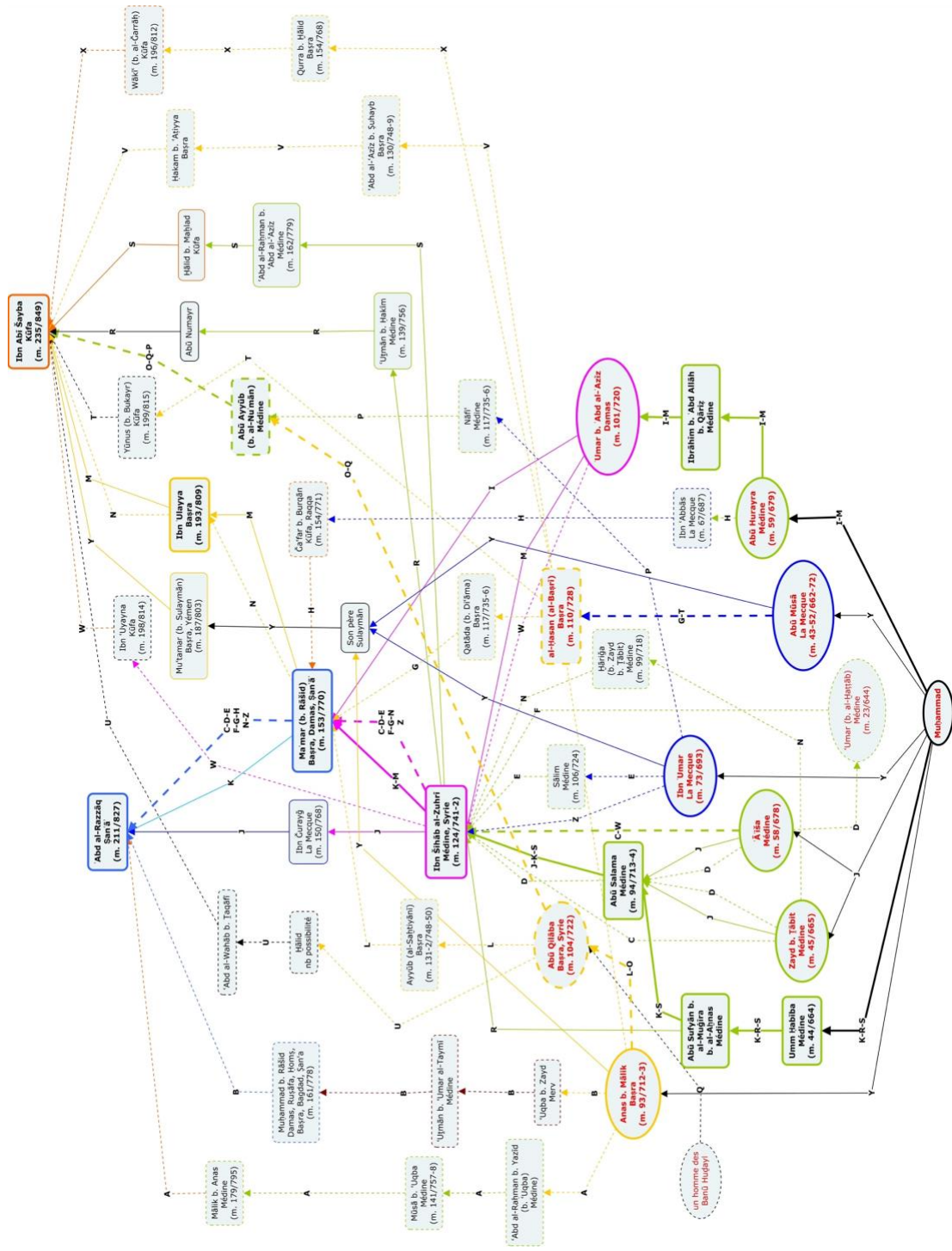


Fig. 2 Utilisation de l'Url d'intégration à partir de l'API de Nakala

DOI : 10.34847/nki.f034e3k3

Public

Gérer les collections de cette donnée

Fichiers

- 10.1. premiere revelation 7...
- 10.2. Iqra premiere revelati...
- 10.3. Recit Iqra Hira.jpg
- 10.4. Isra.jpg
- 10.5. ascension Miraj.jpg
- 10.6. Prière Miraj.jpg
- 10.7. Lumiere Amina.jpg
- 10.8. Coeur purifie.jpg

Visualisation

Copier l'ID

Copier l'url d'intégration

Copier l'url de téléchargement

Fig. 3 Présentation « hybride » avec code QR et mini-URL associés

Dans une variante de ce récit, le plat est constitué de dattes (*balah*). Muḥammad demande à ce qu'un torchon soit placé sur le plat, et de la même façon, les dattes réapparaissent à chaque fois que les Compagnons soulèvent le morceau de tissu. Dans cette variante, le plat de dattes est préparé par 'Umra bt. Rawāḥa ou Umm 'Āmir b. Ašhaliyya¹¹⁶.

Diag.11.8 La multiplication des dattes



<https://vu.fr/YQkLO>

L'analyse des diagrammes 13 et 14 est relativement aisée. On peut remarquer que les deux diagrammes ont en commun la chaîne A attribuée à 'Umra bt. Rawāḥa. La chaîne C du diagramme 14 remonte également à cette femme. De son côté, al-Wāqidī fournit de nouveau des chaînes et des versions parallèles de ces récits. Dans la première variante, dit le miracle de l'agneau, al-Wāqidī rapporte une chaîne rompue (*mursal*) qui remonte directement à Ğābir b. 'Abd Allāh. Dans la seconde variante qui relate la multiplication des dattes, il se fonde sur un récit transmis par me médinois al-Qāsim b. 'Abd -Raḥman, et propose une variante attribuée à Umm 'Āmir b. Ašhaliyya et s'appuie sur une chaîne de transmetteurs quasi inconnus dans la tradition islamique. On peut alors proposer la chronologie suivante :

L'absence de nœud dans les deux diagrammes et le manque de transmetteurs connus dans les débuts de chaînes nous conduit à dater l'émergence de ces récits au milieu du II^e/VIII^e siècle, au moment où Ibn Ishāq (m. 150/767) rédige son *Kitāb al-mağāzī*. À cette époque, plusieurs variantes du récit gravitent autour du personnage de 'Umra bt. Rawāḥa. Dans le premier motif,