



**Les recommandations *Data-Citation*
et *Scholix*
au sein du *Virtual Atomic and
Molecular Data Centre***

C.M. Zwölf, N. Moreau, Y-A. Ba, M.-L. Dubernet

research data sharing without barriers

rd-alliance.org

Le Virtual Atomic and Molecular Data Centre (VAMDC)

2

VAMDC est une e-infrastructure diffusant des données de physique atomique et moléculaire :

- Fédère de façon interopérable ~30 bases de données hétérogènes
 - **Interopérable** → mêmes langage/protocole d'interrogation + même format standard (XSAMS) pour les données restituées.
- Architecture mondialement distribuée, sans point de contrôle centralisé.
- Les Bases de données fédérées sont répertoriées dans des registres (type IVOA)
- Des clients ad-hoc permettent d'interroger simultanément toutes les ressources
- Depuis le début de VAMDC (2009), un de objectifs a été d'augmenter le facteur de citation des producteurs de données.
- → En 2014 (RDA 3, Dublin) nous avons rejoint la RDA pour travailler autour de la citation de données au niveau d'une plus ample communauté

Data Citation WG



i Group details

Status: Recognised & Endorsed

Chair(s): Andreas Rauber, Ari Asmi, Dieter van Uytvanck

Case Statement: [Download](#)

The RDA Working Group on Data Citation (WG-DC) aims to bring together a group of experts to discuss the issues, requirements, advantages and shortcomings of existing approaches for efficiently citing subsets of data. The WG-DC focuses on a narrow field where we can contribute significantly and provide prototypes and reference implementations.

Goals of this WG are to create identification mechanisms that:

- allows us to identify and cite arbitrary views of data, from a single record to an entire data set in a precise, machine-actionable manner
- allows us to cite and retrieve that data as it existed at a certain point in time, whether the database is static or highly dynamic
- is stable across different technologies and technological changes

Solution: The WG recommends solving this challenge by:

- ensuring that data is stored in a versioned and timestamped manner.
- identifying data sets by storing and assigning persistent identifiers (PIDs) to timestamped queries that can be re-executed against the timestamped data store.

- Depuis 2014 VAMDC est devenu un *Pilote* pour la recommandation en cours de définition:
 - La recommandation est-elle conceptuellement applicable à notre architecture distribuée? Ou faut-il corriger certains aspects? → Pas de points bloquants
- Au cours de 2017 (financement RDA-EU3) nous avons implémenté la recommandation
 - Développement & Déploiement du **Query-Store VAMDC**
 - Nous avons relevé un défis technique majeur (par la caractère distribué).
- Documentations & références :
 - <https://github.com/VAMDC/QueryStore> (code source)
 - <https://youtu.be/rfHfnPvH1r4> (webinar RDA avec détails techniques)
 - *New model for datasets citation and extraction reproducibility in VAMDC*, [10.1016/j.jms.2016.04.009](https://doi.org/10.1016/j.jms.2016.04.009)
 - *The VAMDC Portal as a Major Enabler of Atomic and Molecular Data Citation*, [10.3390/galaxies6040105](https://doi.org/10.3390/galaxies6040105)
 - *Implementing in the VAMDC the new paradigms for data citation from the Research Data Alliance*, submitted to *Codata Data Science Journal*

L'implémentation, et après?

5

- Le *Query-Store* (<https://cite.vamdc.eu>) recense :
 - toutes les requêtes reçues par l'intégralité de l'infrastructure
 - les données que ces requêtes ont produites
 - leur contexte de production (provenance + réf. bibliographiques)
- Le *Query-Store* s'est révélé être une plateforme idéale d'intégration d'outils et méthodes supplémentaires qui augmentent considérablement le caractère FAIR des données exposées via VAMDC:
 - Intégration avec ZENODO
 - Certification des données par Blockchain

Query-Store et ZENODO – Implémentation de SCHOLIX

6

- Pour chaque enregistrement du QS (requête+donnée produite) l'utilisateur peut déclencher l'attribution d'un DOI via ZENODO.
- Les données et métadonnées sont copiées vers Zenodo qui leur attribue un DOI (durée de vie minimale de 20 ans).



Query-Store et ZENODO – Implémentation de SCHOLIX

- Pour chaque enregistrement du QS (requête+donnée produite) l'utilisateur peut déclencher l'attribution d'un DOI via ZENODO.

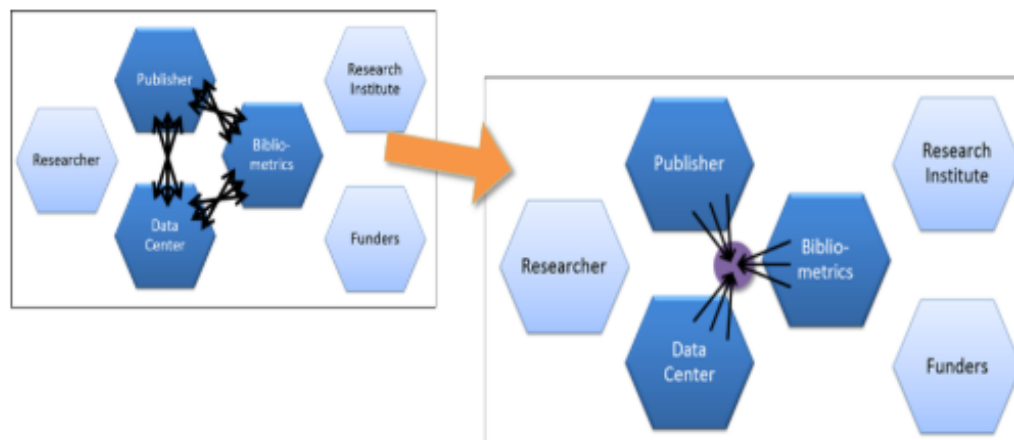


- Les données et métadonnées sont copiées vers Zenodo qui leur attribue un DOI (durée de vie minimale de 20 ans).
- Zenodo est indexé dans OpenAire qui implémente Scholix.



Summary of outputs:

Building on pre-existing components and **international initiatives**, this WG was focusing on a one-for-all cross-referencing service for the links between data and publications. The challenge for the working group has been *"How to move from a plethora of (mostly) bilateral arrangements to a one-for-all service model infrastructure for the research data publication landscape?"* For more details on the working group see our [2-page overview](#).



Query-Store et ZENODO – Implémentation de SCHOLIX

8

- Pour chaque enregistrement du QS (requête+donnée produite) l'utilisateur peut déclencher l'attribution d'un DOI via ZENODO.



- Les données et métadonnées sont copiées vers Zenodo qui leur attribue un DOI (durée de vie minimale de 20 ans).
- Zenodo est indexé dans OpenAire qui implemente Scholix.
- Grâce à Scholix et au schéma de métadonnées derrière DOI v4, quand on cite un DOI on donne du crédit bibliographique aux producteurs de données, aux curateurs des bases des données et personnes en charge des services de diffusion.



Query-Store et ZENODO – Implémentation de SCHOLIX

9

- Pour chaque enregistrement du QS (requête+donnée produite) l'utilisateur peut déclencher l'attribution d'un DOI via ZENODO.



- Les données et métadonnées sont copiées vers Zenodo qui leur attribue un DOI (durée de vie minimale de 20 ans).
- Zenodo est indexé dans OpenAire qui implémente Scholix.
- Grâce à Scholix et au schéma de métadonnées derrière DOI v4, quand on cite un DOI on donne du crédit bibliographique aux producteurs de données, aux curateurs des bases des données et personnes en charge des services de diffusion.

Contributors optional ▾

Contributors Contact person ▾ ×

Optional.

[+ Add another contributor](#)

References optional ▶

- ✓ Contact person
- Data collector
- Data curator
- Data manager
- Editor
- Researcher
- Rights holder
- Sponsor
- Other



ethereum

- Toutes les données recensées dans le Query Store sont certifiées dans une blockchain de type Ethereum (écriture par proof of authority) à l'aide de fireblock.io
- Chaque fichier de données est horodaté et son contenu signé avec une clef privée de VAMDC. Le certificat numérique obtenu est écrit dans la blockchain.
- Il est possible à tout moment de vérifier en un click l'intégrité d'un fichier extrait, de s'assurer du fait qu'il a été bien produit par VAMDC et de sa conformité par rapport à la requête qui l'a généré.
- La pérennité de la blockchain est assurée par son caractère distribué, sans organe de contrôle central. Tout le monde peut avoir sa copie locale et y participer.
- **Fireblock.io** est open source et libre pour les projets open source

- Démonstration de l'écosystème de citation : <https://youtu.be/js4d-b9U0II>
- L'implémentation des recommandations s'inscrit dans le plan «Turning FAIR into reality» du European Commission expert group on FAIR data:
 - *La réutilisabilité s'appuie sur des informations de provenance*
 - ✓ Provenance fine dans les métadonnées de chaque requête dans le QS
 - *Recommandation 26: Soutenir la citation des données et les métriques de nouvelle génération*
 - ✓ Scholix et crédits à toute la chaîne des personnes impliquées dans le service de diffusion des données.
 - *La confiance est essentielle pour la réutilisation et l'authenticité des données est un enjeu*
 - ✓ Certification numérique automatique par blockchain (Fireblock.io)