Certifying Provenance of Scientific Datasets with Self-sovereign Identity and Verifiable Credentials

Iain Barclay, Alun Preece, Ian Taylor
Crime and Security Research Institute,
Cardiff University,
Cardiff, UK
Email: BarclayIS@cardiff.ac.uk

Swapna Radha, Jarek Nabrzyski
Center for Research Computing,
University of Notre Dame,
Notre Dame, IN, USA
Email: sradha@nd.edu

***Abstract*—In order to increase the value of scientific datasets and improve research outcomes, it is important that only trustworthy data is used. This paper introduces mechanisms by which scientists and the organisations they represent can certify the authenticity of characteristics and provenance of published datasets so that secondary users can inspect and gain confidence in the qualities of data sources. By drawing on data models and protocols emerging to provide self-sovereign ownership of identity and personal data to individuals, we conclude that providing self-sovereignty to digital assets offers a promising approach for institutions to certify qualities of their datasets in a cryptographically secure manner, and enables secondary data users to efficiently perform verification of the authenticity of such certifications. By building upon emerging standards for decentralized identification and cryptographically verifiable credentials, we envisage an infrastructure of interoperable tools being developed to foster improvements in the quality of information provided in support of shared data assets.**

I. Introduction

Owners and publishers of scientific datasets and digital assets have an opportunity to increase the scientific value [1] of their dataset or further the commercial opportunities [2] for other types of digital asset by demonstrating the faithfulness of claims about their data, or the authenticity of their assets, through mechanisms which can act as digital watermarks [3]. In order to have confidence in the quality and suitability of a dataset for their needs, potential users and other stakeholders need to have trust in claims made by the originators of the dataset. In practice, even in environments where funding organisations insist on researchers sharing data, there can be a resistance to reuse as Pisani et al. [4] found "lower-than-expected reuse of shared data may be because potential secondary users have few ways of checking the quality of those data". As such, a scheme which allows potential users to access dataset properties and attributes in the form of signed credentials, with an assurance that the credentials related to the data that they were inspecting, were issued by an authorised party and had not been revoked, or tampered with since issuance has the potential to increase the use of shared datasets.

This paper introduces a new method for providing signed certifications relating to properties or attributes of digital assets. The discussion considers a hypothetical scientific dataset, and reviews mechanisms by which parties with ownership or authority over the dataset can attach digital credentials to the dataset so that prospective users and other interested parties can access and verify the credentials to gain confidence in the dataset's qualities. The credential used to illustrate the discussion is an assertion from the originators of the dataset that the data it contains has been ethically sourced and is cleared for further use — a need identified by Scott et al. [5] when considering the implications of recorded voices in a published dataset being used in different scenarios.

The paper continues as follows: Section II introduces current mechanisms for describing properties and metadata about shared datasets, and provisions for evidencing attributes of datasets to provide provenance. Section III discusses established methods for secure information publishing using public and private key signatures, and identifies shortcomings in these methods when applied to sharing assertions about shared digital assets and datasets. The proposed approach builds upon Self-sovereign Identity (SSI) [6], and the enabling concepts of decentralised identifiers (DIDs) [7] and verifiable credentials (VCs) [8] which are introduced in Section IV. Section V explains how these SSI data models and protocols can be adapted to provide a mechanism that allows the publisher of a dataset to issue credentials attesting that their dataset has certain properties and qualities. Section VI extends the discussion to consider automation of credential exchange by treating data assets as self-sovereign entities, with software agents mediating the presentation and verification of credentials between publishers and potential dataset users. The paper concludes with a discussion of our ongoing work in this area as we seek to further integrate existing data sharing schemes with verifiable credentials so that they can be used together to provide improved confidence and trust in shared scientific datasets.

II. Sharing Datasets, Attributes and Provenance

New data assets are produced every day, from INTERMAGNET certified observatories[9], to Department of Health approved curation of social mobility datasets from birth certificate records[10]. However, due to a lack of proven methods to assess and gain trsut in the quality of data, secondary users can effectively utilize only a small subset of the collected data. For the data to be useful to its full extent it should be able to

be verified for its origin and trustworthiness [11], especially when it is applied in the field of scientific research.

Methods are available to add unique attributes to shared datasets, which add a certain level of confidence to the quality of data. Such methods are briefly explained in the following section. The Digital Object Identifier (DOI) System is widely used for identification and management of intellectual content and metadata, and to connect end-users with the requested content. It provides a technical and social framework to identify data by using persistent, interoperable identifiers that can be used across digital networks [12]. Within dataset records located by DOIs, different domains have their own specifications and requirements for metadata. Earth Science Information Partners, founded by NASA, recommends the use of the Provenance and Context Content Standard (PCCS) matrix to perform identification, capturing and tracking of all metadata that can be used to validate the data and to facilitate efficient scientific reproducibility [11]. NASA has use of the PCCS matrix as a requirement for all new Earth science missions, using it to capture and record metadata such as dataset product documentation, dataset product validation which includes the validation record and datasets, dataset calibration information such as calibration method, data, and software used. Additionally, the OpenGIS Sensor Model Language (SensorML) Encoding standard is used to demonstrate various important attributes of sensor and sensory systems such as geometric, dynamic and observational characteristics [13].

In the Biometrics field, Czajka et al. [14] introduce a digital watermarking method that can prove the trustworthiness of iris images without requiring to add supporting data to the iris image. This watermarking based method was designed to certify that biometric data had originated from a genuine sensor. Zhuang et al. [15] discuss the experimentation and evaluation of some of the existing feature extraction methods that are used to measure facial weaknesses. Since an open source annotated facial weakness images dataset was unavailable, experimentation involved first creating a facial weakness dataset using images and videos from public repositories like Google Images and YouTube. The generated "neurologist-certified dataset" was subjected to multiple reviews by experienced neurology trainees and also experts in the field of neurology. If the dataset created during this study is to be effectively shared and applied in future experiments, there would be value in providing demonstrable proof that it was reviewed and certified by neurology experts by providing a signed assertion from these parties.

Existing methods used to provide provenance and demonstrate the trustworthiness of datasets are designed to cater to the needs of specific use cases or domains. Although these methods are being successfully implemented in their respective fields, there is a lack of a common framework or verification mechanism that can be seamlessly applied across all areas of research and other fields. In order to foster the development of such a framework, and to facilitate the provision of high-quality toolsets, it is proposed to provide a mechanism by which datasets and core attributes of metadata about those datasets can be certified and verified with cryptographic assurance using open standards-based models and protocols.

III. Certification of Digital Assets

When evaluating a certification or attestation about an entity, it is critical that the certificate can be inextricably linked to the entity which it describes, otherwise the certificate is of limited value — a photographic identity card, for example, is only of use in attesting the identity of its holder if the inspector of the identity card believes that the photograph on the card provides a visual match to the holder. For digital assets, it is common to use a cryptographic hash value [16] derived from the asset itself to provide a unique digital fingerprint of the asset, and it is proposed to apply this technique to the identification of the datasets discussed in this study. A cryptographic hash is a unique fixed length value that can be generated for a dataset and which will change if the data in the dataset changes in any way, even by a single bit. If the hash value is recorded inside a certificate or credential document, then it can be assured that if the cryptographic hash of the dataset under inspection matches the stored cryptographic hash, then the credential refers to the same dataset. For example, if the certificate states *The dataset with the cryptographic hash 0x1122EEFF has been sourced ethically* then it will uniquely identify the dataset to which it refers. Note that cryptographic hash values are typically 256 or 512 bits in length [17], so the structure of the credential would not likely be readable in sentence form, but the cryptographic hash would be embedded in the credential in some way, as Figure 1 illustrates.

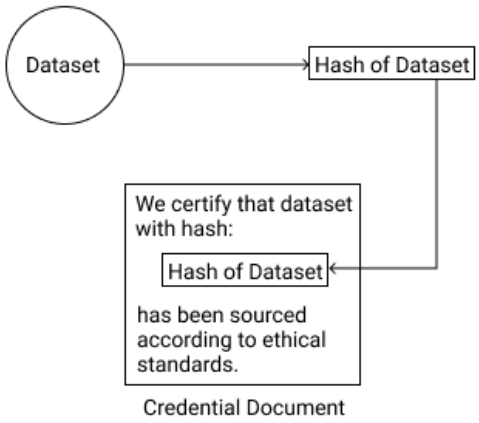

Fig. 1. A Credential Document containing a cryptographic hash of a dataset.

Consider the case in which the information in the dataset's credential is public information, and can be accessed freely by anyone. In this case, an authorised representative of the publisher of the dataset might create a credential document asserting that the dataset represented by the supplied cryptographic hash has been ethically sourced and make a digital signature of the document before finally storing the credential document, the signature and the corresponding public key [18] in a place where it can be inspected (for example, as part

of the supporting material for the shared dataset [19]), as illustrated in Figure 2. Anyone subsequently accessing the dataset's distribution archive can use the public key to verify that the credential document has been signed by the publisher. As discussed above, the credential document would need to include a statement along the lines of *The dataset with the cryptographic hash 0x1122EEFF has been sourced ethically* in order to provide an intrinsic link between the credential document and the dataset it ratifies.

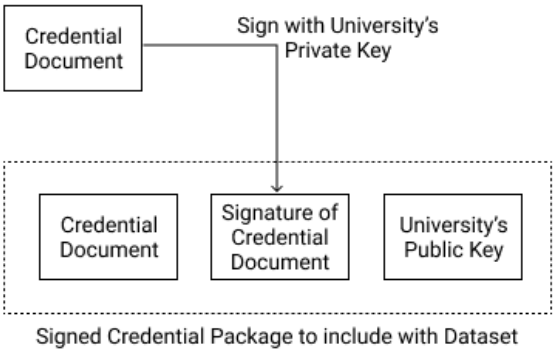

Fig. 2. Signed Credential Package for a dataset instance.

The primary advantage of such a system is that it provides assurance that the certification document is genuine and has not been tampered with, and the system relatively simple to implement as it makes use of well established cryptographic techniques. The publisher will have to establish a process for issuing and managing keys, and in particular for safeguarding access to the private key to ensure that only authorised representatives of the publisher can access and use it.

There are, however, shortcomings in this scheme: most notably, a signed credential issued alongside the dataset cannot be withdrawn or modified if a released dataset is subsequently found to have issues with its ethical status — previously-downloaded copies of the dataset will continue to be accompanied by the credential asserting that the data is ethically sourced, even if evidence subsequently disproves this. Furthermore, including the public key inside the distribution package doesn't guarantee that the public key belongs to the claimed signing party — a malicious party could simply build a new distribution, including a fake certificate signed with their own key, and include that key in the distribution package, giving the appearance that everything was in order whilst propagating false claims about the dataset. This could be mitigated to some degree by hosting the public key on a website or other location that was known to be under the control of the publishing party and then sharing the public key's location over a trusted channel. As such, the informal digital signature-based scheme presented here for verifying a certification document would benefit from having well-known mechanisms and structure for sharing public key locations, and for expressing semantic information about the dataset, such that processes for signing and metadata verification could be automated and be more readily interoperable with other systems and processes.

IV. Introducing Self-Sovereign Identity

The term Self-sovereign Identity (SSI) [20] is used to describe the ability of an individual to take ownership of their personal data and to control who has access to that data, without the need for centralized infrastructure. For instance, in a traditional siloed model identity system, a single or group of identity providers administer the entire system and usually maintains a centralized repository of user identities. This makes it challenging for users to have control over their own data and they depend on administrators for distribution and verification of data. Another disadvantage of a siloed model identity system is that it is not portable. The portability issue of the siloed model is solved by the federated model, which allows some degree of portability by allowing different services to share details about a single user; but the user still does not have complete ownership of their identity. Self-sovereign identity solves the issue of ownership by allowing users to maintain control over their identity. With the use of self-sovereign identity, it is possible to secure users' identity information from unauthorized disclosure by allowing the identity owner to selectively disclose data based on the requirement from the verifier. The data owner thus has the ability to decide how their identity and their personal data should be used and who has access to it.

The SSI community have developed data models and protocols [8] that provide cryptographically verifiable mechanisms for validating identities and issuing and presenting proofs of credentials. These mechanisms build on distributed ledger technologies and linked JSON-LD [21] data documents to provide immutable proof of control of an identifier, and to facilitate the secure exchange of credentials based upon these identifiers between consenting parties.

Whilst a significant focus of effort in the SSI community has been on personal identity and data privacy for individuals, the underlying computer science techniques can be applied to any type of entity, including digital assets and scientific datasets, where the attention of this paper focuses on demonstrating that SSI techniques can be used to provide assertions about the qualities and provenance of a dataset.

V. Representing Attributes as Verifiable Credentials

A core tenet of the decentralised identifier model is that parties claiming to be the controller of a DID can provide cryptographic proof that this is the case, facilitated by a protocol that the DID provides a route to a verification mechanism. This route is typically provided in the form of a JSON-LD document containing the public key of the DID, along with the methods by which a party can verify. By using the published verification mechanisms the holder of a document allegedly signed by the DID's controller can obtain

cryptographic proof that it was indeed signed by the DID controller, and furthermore can verify that the document has not been tampered with since it was signed.

This prescriptive mechanism for a party to prove that they have access to the private keys relating to a DID is utilised when issuing Verifiable Credentials (VC), which can be as simple as a document claiming a DID. If the claim document is signed by a reputable and trusted party, and the DID of that party is known, then the claims in the document can be taken to have been issued by the trusted party. In other words, if a university signs a document using the private key of a DID they control, then the claims in the document can be taken to be claims that the university is willing to endorse. The credential issuer will be the trust anchor in the system, such that anyone relying on credentials provided by the issuer will need to have trust in the issuer themselves to place value on the credentials [22]. Where the issuer is a university, NGO or other well-regarded organisation this trust may be inherent, in other cases the issuer may need to source credentials from bodies with a better established reputation in order to assert their own qualities as a trustworthy issuer of credentials.

Providing a mechanism to assure verifying parties that a DID belongs to a known and trusted authority is a governance challenge, requiring both policies and infrastructure to provide a white list or other trustworthy records that can be checked. In the short term, a DID scheme has been proposed which makes use of the fact that most reputable organisations run web sites, and typically have certificates proving the legitimacy of the identity of the web site. The *did:web* [23] scheme takes advantage of this by utilising the web site of an organisation to host the DID document, resolving *did:web* to a JSON-LD file located on the web site with a well known path [24] and relying on only authorised users being able to upload files to an organization's official web site.

```
{
    "@context": "https://w3id.org/did/v1
       ",
    "id": "did:web:uniofscience.com",
    "authentication": [{
        "id": "did:web:uniofscience.com",
        "type": "
           Ed25519VerificationKey2018",
        "controller": "did:web:
           uniofscience.com",
        "publicKeyBase58": "71ANMccQC..."
      }]
      ...
}
```

Listing 1. A fragment of the UniOfScience DID Document

An open source software package *vc-js* [25] can be used to generate VC documents based on DIDs using many schemes, including *did:web*. In order to produce a VC document for an illustrative dataset, domain names were registered for *UniOfScience*, a fictitious university, and to represent a web site for the dataset, at *DIDdoi.com*, and DID Documents were crafted for each, such that resolving the did:web address through the published route would reach the appropriate DID Document. Listing 1 shows part of the DID Document for *UniOfScience*.

The second required component is a Credential Schema [26], which defines the semantic vocabulary to be used to describe the attributes of the dataset and provides the format in which the claims about a particular subject will be made. To produce a VC document the *vc-js* library was integrated with the Node.js Express [27] framework to enable a simple web form to be served to allow a user — perhaps a scientist preparing to publish a dataset — to enter information about the dataset which was subsequently used to populate data fields in the credential schema, and *vc-js* invoked to encapsulate these values in a Verifiable Credential JSON-LD document containing a proof issued by the DID belonging to *UniOfScience*.

The inclusion of the DID of the issuer (and signatory) of the VC document enables other parties to verify its state, which is achieved by resolving the DID to locate the DID Document holding descriptors of the mechanisms for checking signatures, usually by provision of the public key. Verifiers can use methods in *vc-js* to receive cryptographic proof of the authenticity of the VC document, assuring them that it hasn't been tampered with since it was issued. As the payload of the VC document contains the DID of the dataset that it refers to, verifiers have cryptographic proof that the issuer has signed a document attesting to the properties of the DID of the subject. If the DID relates to a dataset, then the verifier can be assured that the VC document carries signed assurances about the properties of the dataset.

Presentation of a VC document provides systematic improvements over an ad hoc digitally signed document, through the publication of the location of the public key of the subjects and the use of a semantic vocabulary for expressing metadata claims, but still exhibits shortcomings. Among these is the possibility for anyone who holds a copy of the VC document to present it, even though it may no longer be relevant or valid. If this is done with ill-intent it could be considered a reply attack [28]. A mechanism to prevent such replay attacks is for the verifier of a credential claim to ask the holder to present a Verifiable Proof, which takes the form of a JSON-LD document signed by the VC holder containing a challenge, typically a nonce, issued by the verifier along with the credential and its proof. By inspection of the Verifiable Proof document through resolution of the DID of its issuer, and comparing this DID with the DID of the VC's subject, the verifier can determine that the challenge response is acceptable, and that the holder of the VC is still content to present it. Further verification of the credential contained in the Verifiable Proof can demonstrate that the credential has not been tampered with, and thereby provide assurance that a valid credential about the dataset has been issued by an authorised party, to a holder who still considers the credential appropriate to share.

VI. Self-sovereignty of Digital Assets

The process described thus far has implied that parties wishing to verify credentials make requests and human operators are on hand to manage private keys and to create and sign Verifiable Proof documents, which would quickly become impractical where there was high demand or a need for a timely response. By considering the scientific dataset as a self-sovereign entity in its own right, with control over its own credentials, we can begin to automate these processes using a construct, exemplified by the open source Hyperledger Aries project [29] and Evernym's commercial application libvcx[1], whereby the entity is represented by a software application (termed an agent) operating on behalf of the entity and mediating access to the entity's credentials. In the scenario presented, a cloud-based software agent (DSA) will represent the published dataset (DS) and will provide mechanisms to generate and store the private keys and issued VC documents securely in a digital wallet. Parties with interests in DS will interact with the representing agent, DSA, by addressing it through its publicly shared decentralised identifier (DID).

The dataset publisher will also have a software agent representing their role in the transactions, with the capability to issue credentials. The publisher's agent will be the trust anchor in the system, such that anyone relying on data credentials provided by the publisher's agent will need to have trust in the publisher themselves in order to place value on the credentials, as in the *did:web* scheme described above. For the discussion, the publisher is a university that employs the scientist who seeks to share a dataset with the research community.

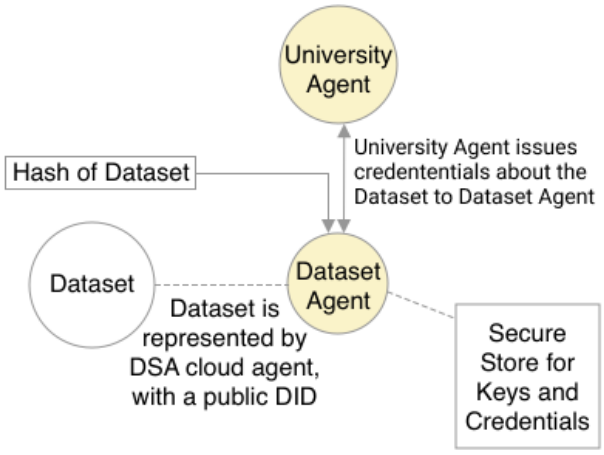

Fig. 3. A publisher agent (University Agent) interacts and issues credentials to a dataset agent.

In the first instance the university will configure and launch a new software process to act as the agent DSA and represent DS in future transactions. This software process will be part of the infrastructure provided by an implementation of the DID protocols, such as the ACA-Py [30] cloud agent component from the Hyperledger Aries project. As part of the commissioning procedure for the agent, a configuration script will be run to generate a digital wallet for the agent, which will hold its private keys and credentials, and generate a decentralised identifier (DID), by which DSA will be addressed.

The university will then establish a secure connection between their publisher agent and DSA, which will be manifested using DID protocols, and the resultant connection will be used to issue the appropriate credentials to DSA, as shown in Figure 3. Credentials are structured according to published JSON-formatted Credential Definition schema [31] and contain a set of key-value pairs which the issuing party asserts are true about the holding entity. For example, in Listing 2, *Hash of Data* is a credential holding the cryptographic hash of the dataset, which inextricably links the credential to the dataset it represents, and *Data Ethically Sourced* represents the ethical status of the dataset, as stated by the publisher. A practical scheme will hold other credentials, and could include a credential expiry date or other conditions for credential usage.

```
{
    "Hash of Data": "0xFFEE...AA1122",
    "Data Ethically Sourced": "YES"
}
```

Listing 2. A sample credential set

The DID for the dataset agent, DSA, is a public address, analogous to a website address, which should be shared in the downloadable package for the Dataset DS that it represents, such that users can use this address to request proof of the credentials of DS. Users of DS will themselves use software agents to communicate with its agent, DSA, these may be edge agents stored on a mobile device or cloud agents hosted by a web service. Any user with knowledge of the DID for DSA can seek to establish a connection with DSA and to request proof of the dataset's credentials via their own agent, and if the system policies permit, DSA will respond, presenting proof of the credentials it holds. These credentials will include the cryptographic hash of the dataset, to provide an inextricable link to the underlying dataset it represents, along with its ethical sourcing status, as written in the credential by the publisher. The public DID of the university will be included in the returned credential proof, and can be used as a trust anchor to assure that the credentials originated from the university. DID protocols ensure that the returned proof is cryptographically provable to have originated from the issuing university and to have not expired, been revoked or tampered with in any way. The architecture for such a scheme is shown in Figure 4.

VII. Conclusions and Future Work

This paper has introduced the use of emerging protocols and data models for decentralised identifiers and verifiable credentials, with a novel application towards provision of certification of qualities and attributes of digital assets. The discussion has been motivated through the use of an example

[1]https://www.evernym.com

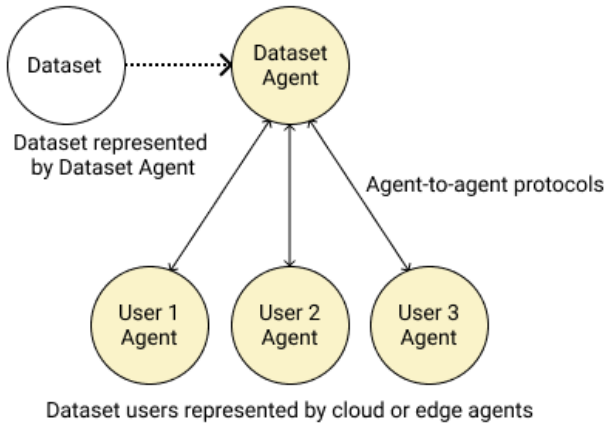

Fig. 4. Agents represent parties in the relationship.

ethical certification applied to a hypothetical scientific dataset, and has described how decentralised identifiers and verifiable credentials can be used to provide users of the dataset with access to a well-structured and semantically-rich digital attestation of qualities of the dataset.

By building upon concepts and data models from Self-sovereign Identity research, it is hoped that a standard's based approach to credential sharing and verification of properties of digital assets will emerge, such that users will be able to leverage a range of interoperable tools and platforms both to certify shared data and to verify claims made about third party data, and as a result increase the adoption of shared datasets. The strength of such a scheme relies upon trust among members of a research community, such that a signed certificate from a member of the community will be sufficient to demonstrate to other members that the assertions made are true. Building and supporting this trust network relies heavily on human relationships, which can be facilitated and supported by technology. If an effective trust-based data sharing ecosystem can be established, then significant added value is likely to be achieved through the emergence of decentralised platforms, underpinned by cryptographic assurances of claims made.

Continuation of the work presented here will provide implementation of the architecture discussed, building on open source software infrastructure and standards-based technologies where possible, in order to evaluate the benefits and practical utility of such a system with use cases from the multi-messenger astrophysics domain.

REFERENCES

[1] Y. Gil, C. H. David, I. Demir, B. T. Essawy, R. W. Fulweiler, J. L. Goodall, L. Karlstrom, H. Lee, H. J. Mills, J.-H. Oh *et al.*, "Toward the geoscience paper of the future: Best practices for documenting and sharing research from data to software to provenance," *Earth and Space Science*, vol. 3, no. 10, pp. 388–415, 2016.

[2] H. G. Miller and P. Mork, "From data to decisions: a value chain for big data," *It Professional*, no. 1, pp. 57–59, 2013.

[3] J. M. Acken, "How watermarking adds value to digital content: a digital watermark isn't just a tag or label for protecting content but an opportunity to increase the value of the content itself," *Communications of the ACM*, vol. 41, no. 7, pp. 75–79, 1998.

[4] E. Pisani, L. Merson, A. Ghataure, G. Castillo, A. Castillo, and Y. Moride, "Sharing health research data in low-resource settings: Supporting necessary infrastructure and building on good practices," 2018. [Online]. Available: https://wellcome.figshare.com/articles/Sharing_health_research_data_in_low-resource_settings_Supporting_necessary_infrastructure_and_building_on_good_practices/6042047

[5] K. M. Scott, S. Ashby, D. A. Braude, and M. P. Aylett, "Who owns your voice?: ethically sourced voices for non-commercial tts applications," in *Proceedings of the 1st International Conference on Conversational User Interfaces*. ACM, 2019, p. 17.

[6] A. Mühle, A. Grüner, T. Gayvoronskaya, and C. Meinel, "A survey on essential components of a self-sovereign identity," *Computer Science Review*, vol. 30, pp. 80–86, 2018.

[7] A. Hughes, M. Sporny, and D. Reed, "A primer for decentralized identifiers," *Draft Community Group Report, W3C*, 2019.

[8] M. Sporny, "A verifiable credentials primer," *Online, Available at: https://github.com/WebOfTrustInfo/rwot7-toronto/blob/master/topics-and-advance-readings/verifiable-credentials-primer.md*, 2018.

[9] R. Sidorov, A. Soloviev, R. Krasnoperov, D. Kudin, A. Grudnev, Y. Kopytenko, A. Kotikov, and P. Sergushin, "Saint petersburg magnetic observatory: from voeikovo subdivision to intermagnet certification," *Geoscientific Instrumentation, Methods and Data Systems*, vol. 6, no. 2, p. 473, 2017.

[10] N. L. Fleischer, C. Abshire, C. E. Margerison, D. Nitcheva, and M. G. Smith, "The south carolina multigenerational linked birth dataset: Developing social mobility measures across generations to understand racial/ethnic disparities in adverse birth outcomes in the us south," *Maternal and child health journal*, vol. 23, no. 6, pp. 787–801, 2019.

[11] D. Hills, R. Downs, R. Duerr, J. Goldstein, M. Parsons, and H. Ramapriyan, "The importance of data set provenance for science," *Eos Transactions American Geophysical Union*, vol. 96, 12 2015.

[12] N. Simons, "Implementing dois for research data," *D-Lib Magazine*, vol. 18, no. 5/6, p. 1, 2012.

[13] H. Ramapriyan and J. Moses, "Nasa earth science data preservation content specification," *Goddard Space Flight Center, Greenbelt, MD, USA, Tech. Rep*, 2011.

[14] A. Czajka, W. Kasprzak, and A. Wilkowski, "Verification of iris image authenticity using fragile watermarking," *Bulletin of the Polish Academy of Sciences Technical Sciences*, vol. 64, no. 4, pp. 807–819, 2016.

[15] Y. Zhuang, M. McDonald, O. Uribe, X. Yin, D. Parikh, A. M. Southerland, and G. Rohde, "Facial weakness analysis and quantification of static images," *IEEE Journal of Biomedical and Health Informatics*, 2020.

[16] B. Preneel, "Cryptographic hash functions," *European Transactions on Telecommunications*, vol. 5, no. 4, pp. 431–448, 1994.

[17] S. Gueron, S. Johnson, and J. Walker, "Sha-512/256," in *2011 Eighth International Conference on Information Technology: New Generations*. IEEE, 2011, pp. 354–358.

[18] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[19] A. Pepe, A. Goodman, A. Muench, M. Crosas, and C. Erdmann, "How do astronomers share data? reliability and persistence of datasets linked in aas publications and a qualitative study of data practices among us astronomers," *PLoS One*, vol. 9, no. 8, p. e104798, 2014.

[20] C. Allen, "The path to self-sovereign identity," *URL: http://www.lifewithalacrity. com/previous/2016/04/the-path-to-selfsovereign-identity.html*, 2016.

[21] M. Lanthaler and C. Gütl, "On using json-ld to create evolvable restful services," in *Proceedings of the Third International Workshop on RESTful Design*, 2012, pp. 25–32.

[22] J. Linn, "Trust models and management in public-key infrastructures," *RSA laboratories*, vol. 12, 2000.

[23] O. Terbu, D. Zagidulin, and A. Guy, "did:web decentralized identifier method specification," 2020. [Online]. Available: https://w3c-ccg.github.io/did-method-web/

[24] M. Nottingham and E. Hammer-Lahav, "Defining well-known uniform resource identifiers (uris) rfc 5785," April 2010. [Online]. Available: https://www.rfc-editor.org/info/rfc5785

[25] Digital Bazaar, "Verifiable credentials js library (vc-js)," 2020. [Online]. Available: https://github.com/digitalbazaar/vc-js/blob/master/README.md

[26] M. Sporny, G. Noble, D. Longley, D. C. Burnett, and B. Zundel, "Verifiable credentials data model," November 2019. [Online]. Available: https://www.w3.org/TR/vc-data-model/
[27] E. Hahn, *Express in Action: Writing, building, and testing Node. js applications*. Manning Publications,, 2016.
[28] P. A. Grassi, M. Garcia, and J. Fenton, "Nist special publication 800-63-3 digital identity guidelines," *National Institute of Standards and Technology, Los Altos, CA*, 2017. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-63-3
[29] "Hyperledger aries," https://www.hyperledger.org/projects/aries, 2019.
[30] "Hyperledger aries cloud agent - python," https://github.com/hyperledger/aries-cloudagent-python, 2019.
[31] Z. A. Lux, F. Beierle, S. Zickau, and S. Göndör, "Full-text search for verifiable credential metadata on distributed ledgers," *arXiv preprint arXiv:1909.02895*, 2019.