FAIRTool.org toward better Earth science data stewardship

Project Details

Name of project: FAIRTool.org toward better Earth science data stewardship

Project lead and contact details:

- Abdullah Alowairdhi, PhD student, Computer Science Dept., University of Idaho, Email: alow4530@vandals.uidaho.edu

Project partners and contact details:

- Prof. Xiaogang Ma, Computer Science Dept., University of Idaho, Email: max@uidaho.edu
- Dr. Luke Sheneman, Director, UI Northwest Knowledge Network (NKN), Email: sheneman@uidaho.edu

Proposed start and end date: start date: 02/01/2019 end date: 08/31/2019

Budget Requested: $7,000.00

Budget Summary:

Facilities: $3,500.00, for development infrastructure and expenses:

- $2,000.00, for Web service hosting server.
- $1,500.00 development machine.

Travel: $3,500.00, for attending ESIP 2019 Summer and 2020 Winter meetings to present research progress and output and to engage ESIP member institutions to collaborate on building and testing the FAIR Tool. Each travel will cost about $1,750.00.

Project Outline

Project description:

Good data stewardship is the key to knowledge discovery and scientific innovation. To generate value for a research community beyond the initial researchers, funding agencies are increasingly setting requirements for proper data stewardship of research output. Beyond proper collection, annotation, and archival, data stewardship include planning and long-term care, with the goal that data can be found, accessed, integrated, and reused in subsequent studies, either in or outside the original data collector's group [1]. To facilitate such better data stewardship, we propose a novel approach to build a FAIR tool following The FAIR Guiding Principles for scientific data management and stewardship. FAIR is an acronym of Findable, Accessible, Interoperable, and Reusable. Concisely, the FAIR Principles suggest that every data element should have a globally-unique identifier, and that this identifier should be

associated with contextual, searchable metadata ("Findable"); These identifiers should all resolve to data or metadata using an open, standard protocol ("Accessible"); The data and metadata should use a formal, broadly viable representation language, and apply open and widely-accepted domain-relevant vocabularies and ontologies ("Interoperable"); And finally, the data should be profusely described with an abundance of cross-references, and with a clearly-defined technique for accessing provenance and license information ("Reusable") [2].

There have been many recent discussions on applying FAIR Principles in research activities, including those in Earth science [3, 4, 5]. Nevertheless, the whole community is still in the early stage on technical implementation of FAIR. This project aims to contribute to the technical framework. In other words, the project is about how to embed FAIR Principles into an executable process to help researchers achieve better data stewardship. The solution will be built upon advanced semantic web technologies, through a combination of community standards and in house developed ontologies, following the FAIR principle guidance. The deliverable is a website (www.fairtool.org) that will incubate the Earth science community research outputs and will consist of five tabs. The first four tabs will take care of the four FAIR elements. The result will be displayed as metadata in the fifth web tab, which will also contain a FAIRness score and can be downloaded as an RDF format.

Project objectives, significance and impact:

The objective of this project is to automatically produce machine-readable metadata for the digital data outputs of the Earth science community. The metadata will be FAIR compliant as possible. The tool will include a metric system to measure the FAIRness components of digital data objects. The end result will be a tool that will promote the Earth science community research outputs toward better data stewardship. The work will strongly augment the active discussion on FAIR Principles by providing an executable pilot implementation. With small adaption, the tool can be reused by a large number of data repositories in Earth science. It can also be used by individual researchers or groups to evaluate and improve the FAIRness of datasets. In a broader perspective, FAIRness is not only limited to data objects. Based on the result of this project, we can also extend the implementation of FAIR Principles to other digital objects, such as open source software programs.

Description of key project steps and timeline:

Duration: 7 months: Feb 2019 to Aug 2019.

- Feb.-Mar., 2019 Project plan report; Architecture design; Ontologies collection, creation and integration.
- Apr.-Jun., 2019 Write program codes; Triple store setup; API development.

- Jul., 2019 Online testing, debugging and implementation of deliverables; Presentation at ESIP 2019 summer meeting.
- Aug., 2019 Project report writing.
- After the end of the project, we will continue a few use case tests by collaborating with interested ESIP members, and present the final output at ESIP 2020 winter meeting.

Outreach

What groups/audiences will be engaged in the project?

The project aims to engage researchers in the field of Earth science and Geoinformatics, including initially with ESIP, then with GSA, AGU, and more. In the future, it can be expanded to accommodate other scientific fields.

How will you judge that project has had impact?

If we ask "what are the available easy and engaging tools for Earth science scientific community to make their digital outputs FAIR complaint?" Unfortunately the answer is none, although there are tools supporting partial aspects of FAIR. Our aim is to develop an easy and engaging tool to enable researchers to make their research outputs FAIR. The developed FAIR Tool (www.fairtool.org) will be a solution-oriented project that will enhance the cyberinfrastructure for Earth science. It will empower the Earth science scientific community to improve the FAIRness of digital data and other objects. By making the FAIR Tool a public website, we will engage users from individual researchers and data managers of Earth science data facilities. The planned presentations and engagement at the two ESIP meetings will offer a nice opportunity for that purpose.

How will you share the knowledge generated by the project?

The FAIR Tool will be available online as a public website (www.fairtool.org). It will be open for use by the Earth science scientific community. The details of the project will be available in the ESIP GitHub including a permissive license. We will also share information about the research outputs through ESIP and AGU-ESSI email lists, and through presentations at ESIP meetings.

Project Partners

Description of project partners (individuals and/or organizations) and their involvement:

Currently the project has two collaborators, Prof. Xiaogang Ma and Dr. Luke Sheneman. Prof. Ma is the major advisor of Abdullah Alowairdhi at the University of Idaho. He has been an active ESIP member

since 2012, and has close collaborations with other efforts of implementing FAIR principles. Also, he will help engage use case tests from a number of data facilities in Earth science. Dr. Sheneman is the director of the Northwest Knowledge Network (NKN), which is a research data management support unit based in the office of research at the University of Idaho. NKN will host the website (www.fairtool.org) of FAIR Tool and provide needed technical support.

How will this project engage members of the ESIP community?

FAIR tool will be a useful framework to document the research outputs in a FAIR manner. With such tool, the ESIP community will find it an adequate service for better stewardship for their research outputs. The FAIR tool will be of great value to the ESIP community by providing several benefits to researchers and research community. In brief, the researcher benefits are 1) Increased Citations, 2) Career recognition, and 3) New collaborations; And the research community benefits are 1) Easier to find useful data, 2) Good data archived and preserved for the future, 3) More efficient research, and 4) New research will be made possible.

References

[1] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

[2] Rodríguez-Iglesias A, Rodríguez-González A, Irvine AG, Sesma A, Urban M, Hammond-Kosack KE and Wilkinson MD (2016) Publishing FAIR Data: An Exemplar Methodology Utilizing PHI-Base. Front. Plant Sci. 7:641. doi: 10.3389/fpls.2016.00641

[3] Stall, S., E. Robinson, L. Wyborn, L. R. Yarmey, M. A. Parsons, K. Lehnert, J. Cutcher-Gershenfeld, B. Nosek, and B. Hanson (2017), Enabling FAIR data across the Earth and space sciences, Eos, 98, https://doi.org/10.1029/2017EO088425. Published on 08 December 2017.

[4] Wilkinson, M. D., Verborgh, R., Bonino da Silva Santos, L. O., Clark, T., Swertz, M. A., Kelpin, F. D. L., Dumontier, M. (2017). Interoperability and FAIRness through a novel combination of Web technologies. PeerJ Computer Science, 3, e110. https://doi.org/10.7717/peerj-cs.110

[5] Mons, B., Neylon, C., Velterop, J., Dumontier, M., Olavo, L., Da, B., Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use, 37, 49–56. https://doi.org/10.3233/ISU-170824