Convergence. Are we there yet?

Mark A. Parsons (@chutneyboy, parsom3@rpi.edu, https://orcid.org/0000-0002-7723-0950)

http://tw.rpi.edu

Rensselaer

TetherlessWorld

Alfred P. Sloan Foundation

Summary

Infrastructure is not truly designed. It evolves. Infrastructure matures and becomes more ubiquitous, accessible, reliable, and transparent as 'gateways' begin to connect myriad systems [1]. Often these gateways that facilitate interconnection are 'standards' — be they formally defined or tacitly accepted practice. There is a process of convergence where communities begin to agree not only to use a common standard, but to use it in a common way. This presentation examines two case-studies of how standards adoption for data interoperability has progressed over time: The development of an international Arctic data coordination network over the last 15 years and work establishing the Research Data Alliance.

Observations suggest that convergence on standards and their implementation is a dynamic process that works to negotiate and interpret multiple, sometimes conflicting, concerns. Standards can be adopted in different ways by different people and organizations and still enable interoperability. An inclusive, ecological methodology to system design is more resilient, adaptive, and responsive to interdisciplinary needs. Overall, maintenance of standards and their community is essential to their adoption and persistence.

Standards are

community policies

that add layers of abstraction.

Convergence is ongoing and ephemeral.

Observations from the Arctic

- Basic roadmap for distributed polar data discovery was agreed in 2006 as part of the International Polar Year [2]. Thirteen years later much of the vision is being realized but convergence is erratic in pace and degree.
- Agreed on ISO 19115 very early on, but adoption took many years. DIF still underpins some systems. Now most realize the need to provide metadata in multiple forms.
- The technical NSF Arctic data system was rebuilt multiple times. The current system is robust and discoverable, but it was not a straight path to get there. Politics of research funding severely complicated development and agreement. Technology and standards were never at issue.
- Top-down technical mandates repeatedly failed. They only worked when narrowly scoped., e.g. the Arctic Spatial Data Infrastructure decided to focus only on reference layers.
- A successful working prototype 'Arctic Data Explorer' enabled rich data discovery and access across many portals by implementing multiple (competing) standards with active maintenance and human engagement. It was not sustained.
- Technical agreements continually changed before convergence was achieved, e.g. whether to favor CSW, OAI-PMH, REST, or schema.org.
- **Vocabularies, vocabularies, vocabularies — Context, context, context.** Semantics are difficult and require a critical mass of agreement before they can be useful. Early convergence around GCMD keywords diverged because of lack of consistence maintenance of the terms, their accessibility, and the process for definition agreement.
- Network analysis is helpful in understanding a complex data ecosystem [3].

Observations from the Research Data Alliance

- Is RDA a standards organization? A continual debate.
- The vast majority of standards are developed by organizations that facilitate collaboration and cooperation among diverse technical experts [4].

RDA RESEARCH DATA ALLIANCE

> **"The basic irony of standards is the simple fact that there is no standard way to create a standard, nor is there even a standard definition of 'standard'." [7]**

- The emergence of FAIR as a discussion and outreach framework. An example of the periodic convergence on principles that allow a community to take a next step.
- Building into existing institutions and infrastructures can speed adoption of standards, but can also deepen path dependence. Scholix built on top of DataCite and CrossRef and was rapidly implemented making things easier for many but harder for some to implement [5]. So much depends on the players and use cases involved.
- Data Seal of Approval a rare example of convergence of standards, but infrastructure-wide adoption is daunting.
- Conceptual and logical models are easier to adopt than physical models because they are easier to adapt to existing systems, e.g. the Deep Carbon Observatory implemented the RDA Data Type Registry as an extension to its ontology [5]. Of course, this also adds a layer of mediation to enable full interoperability.
- **PIDs all over the place!** Huge acceptance in principle, but still very irregular implementation. We use persistent identifiers (registered, resolvable, unique identifiers) to do much more than identify. We continually try to bundle basic concerns under the resolution of a PID, including location, reference, resource type, authority, persistence, etc. Work continues on a core 'kernel' and definitions of '(FAIR) digital objects,' but it is early days.

A physical example

A noble and long lasting standard is dying because it cannot bundle more concerns. Gateways can become increasingly complicated until another layer is abstracted.

Conclusions

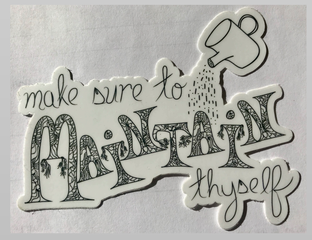
- Agreeing on a standard is is just a first date in the long-term interoperability relationship of classifying and categorizing.
- Standards can be viewed as policy in that they must continually be negotiated and interpreted.
- Standards often work to bundle multiple concerns. That's why they are difficult to negotiate and use consistently.

Created by corpus delicti from Noun Project

- We keep either bundling functions into a connector/gateway and that requires another layer of abstraction, or we precisely define functions which can be disaggregated for more utility, e.g. credit vs. access in citation.
- Even the best standards don't last forever, and sometimes the gateways get too complex, which is probably an indication for a need for another layer of abstraction.

> **"Standardization is dynamic, not static; it means not to stand still, but to move forward together."**
>
> **1920s motto for the Engineering Standards Committee (precursor to ANSI)**

- Functioning standards have an active and collaborative user community (not the same as the system end users). So we must include makers and users of standards in development and maintenance. It's all about the use cases (which change). Broad inclusiveness is essential.
- There is no homogenous view of convergence [8]. It is never complete or permanent because it is measured on many different dimensions. Convergence does not mean absolute adherence to any one system or practice. Convergence is a community activity not a directed activity, and it rarely lasts. It ebbs and flows. Don't give up.
- The mundane and ordinary require people to maintain. See themaintainers.org.



References:

[1] Edwards, P. N., S. J. Jackson, G. C. Bowker, and C. P. Knobel. 2007. "Understanding infrastructure: Dynamics, tensions, and design." http://hdl.handle.net/2027.42/49353.

[2] Parsons, M. A., et al. 2006. "International Polar Year data management workshop, 3-4 March 2006, Cambridge, UK." *Glaciological Data Series GD-33*. https://nsidc.org/sites/nsidc.org/files/files/Glaciological_Data_33.pdf.

[3] Pulsifer, P. L., L. Yarmey, Ø. Godøy, J. Friddell, M. A. Parsons, W. F. Vincent, T. de Bruin, W. Manley, A. Gaylord, A. Hayes, S. Nickels, C. Tweedie, J. R. Larsen, and J. Huck. 2014. "Towards an international polar data coordination network." Data Science Journal 13 PDA94–PDA102. https://dx.doi.org/10.2481/dsj.IFPDA-16.

[4] Biddle, B., A. White, and S. Woods. 2010. How Many Standards in a Laptop? (And Other Empirical Questions) (September 10, 2010). Available at SSRN: https://ssrn.com/abstract=1619440 or http://dx.doi.org/10.2139/ssrn.1619440.

[5] Parsons, M. A., and P. A. Fox. 2018. "Power and persistent identifiers." *International Data Week 2018* https://dx.doi.org/10.5281/zenodo.1495321.

[6] Ma, X., P. West, S. Zednik, J.s. Erickson, A. Eleish, Y. Chen, H. Wang, H. Zhong, and P. A. Fox. 2017. "Weaving a knowledge network for deep carbon science." *Frontiers in Earth Science*. http://dx.doi.org/10.3389/feart.2017.00036.

[7] Russel, A., and L. Vinsel. 2019-02-16. "The joy of standards." New York Times. https://www.nytimes.com/2019/02/16/opinion/sunday/standardization.html accessed 2019-12-07.

[8] Heichel, S., J. Pape, and T. Sommerer. 2005. "Is there convergence in convergence research? An overview of empirical studies on policy convergence." Journal of European Public Policy 12(5) (5): 817–40. https://dx.doi.org/10.1080/13501760500161431.

10 December 2019 — American Geophysical Union Fall Meeting, IN21D-0872