

I'm here to talk to you about something exciting. One of the most exciting things that has happened in the data world in a long time— The Research Data Alliance

I was limited to three slides, but I wanted to show you 50; so I decided to show you none and just write it down.

I want to try and keep it simple, but it's hard. Data really seems to have come of age in the last couple years. Suddenly governments, universities, publishers, professional societies are all making statements and policies around data sharing. The recent memos from the white house, in particular, are driving a lot of activity in the US.

I mean, right? It's pretty cool. "Big data" as silly as it may be is the buzzword of the day. Data scientists are sexy. We need to take advantage of this moment.

And of course it is a global phenomenon. And that is where RDA comes in

RDA builds the social and technical bridges that enable open sharing of data.

That's our mission statement.

We build the social and technical bridges that enable data sharing. Bridges across technologies, scales, disciplines, cultures...

The social and technical bridges. It's an important metaphor and I'll use it a lot.

So I said I wanted to keep it simple. I'll try to describe RDA (and the moment) in three concepts.

- Implemented infrastructure
- People
- What we might call "glocality" a blend of the global and the local

So infrastructure, people, and the glocal

First infrastructure. Carole has shown the complexity of infrastructure and how it is not just the pipes and wires but more a body of relationships connecting machines and people and people and machines. It's about the connections, the interfaces, the relationships. The social and technical bridges if you will.

Also, If we look at how past infrastructures developed and consider the work of Star, Edwards, Bowker, and others it is clear that infrastructure evolves. It is not architected. It is more of an organic process.

We've seen time and again top down, build it and they will come systems not realize their potential or simply fail.

So RDA strives to be more bottom up. More organic. Anyone can join if they agree to our basic principles and they can work on whatever problem is important to them as long as they can demonstrate that it advances data sharing. That's key. We're not trying to solve *all* the data problems. We're focused on *implementing* data sharing solutions.

I mentioned our principles. These are the heart of RDA. They help guide us through the chaos of an organic bottom up approach.

- **openness** – Membership is open to all interested individuals who subscribe to the RDA's Guiding Principles. RDA community meetings and processes are open, and the deliverables of RDA Working Groups will be publicly disseminated.
- **Consensus** – The RDA moves forward by achieving consensus among its membership. RDA processes and procedures include appropriate mechanisms to resolve conflicts.
- **Balance** – The RDA seeks to promote balanced representation of its membership and stakeholder communities.
- **Harmonization** – The RDA works to achieve harmonization across data standards, policies, technologies, infrastructure, and communities.
- **Community-driven** – The RDA is a public, community-driven body constituted of volunteer members and organizations, supported by the RDA Secretariat.
- **Non-profit** - RDA does not promote, endorse, or sell commercial products, technologies, or services.

It's about balancing a grass roots approach with just enough guidance and process.

Implemented infrastructure. Organic and bottom up but managed towards adoption

Secondly, RDA is about people and the work they do.

RDA is just over a year old and we already have more than 1850 members. And they are generally active members participating in Working and interest groups and our twice-yearly plenary meetings.

Most of the people come from the US and Europe but we really have global reach. Our members come from more than 80 countries.

It is mostly academic folks, but we also have strong and growing representation from the government and private sector as well.

We also have around 30 or 40 organisational members including the research arms of tech companies like Microsoft research, university libraries, regional

efforts like Internet2 and the Australian National Data Service (ANDS), and other related organizations like CODATA, and WDS.

All these people come together in short-term, tiger-team-style working groups and broader more exploratory interest groups.

These Working and Interest groups are addressing all sorts of topics as determined important by the members themselves.

Some are quite technical--addressing issues such as PIDs, metadata, core data terminology and types, machine actionable rules and workflows, etc.

Some are more oriented to social issues such as legal interoperability or repository certification.

Some bridge the social and technical on issues like data citation or provenance or domain repositories.

Some are specific to certain disciplines or domains like oceanography, genomics, and history and ethnography. One of my personal favorites is the Wheat interoperability WG, which is involving agricultural organizations from around the world on what is clearly an important topic for humanity—feeding us all.

So all these people are working on lots of cool things, but it is not just abstract. A unique feature of RDA is its focus on implementation. Our WGs are short lived--only 18 months--. At the end of which they need to have actually implemented something—a particular specification or method or practice that improves data sharing.

Furthermore, adoption of the outputs is built in to the process. As part of the WG approval process, the group needs to demonstrate that it has members who actually plan to use what is developed. This helps focus the work and also ensures that it is relevant.

Many of the groups are co-sponsored by partner organizations like ORCID, DataCite, DSA, CODATA, WDS.

It's really quite exciting and it highlights the power of volunteer effort.
it's all about the volunteer

Finally, RDA is glocal. This is a somewhat contrived word although it is used some in the literature. The idea is that you need to act both globally and locally. It's more than think global act local, it's simultaneously playing at both levels.

We recognize that implementation is inherently a local activity, but it is most relevant and impactful if it is done in a global context.

So in addition to the global RDA there are local or regional RDAs. RDA/US in our context. This includes all the US players in RDA—about 1/3 of those 1850 people and about 1/3 of the leadership in our Council, Technical Advisory Board, WG chairs, etc.

We have very broad representation from many disciplines and about 45 of the 50 states (We're missing the Dakotas).

The point is to ensure that RDA is relevant to unique US needs—our needs to address the OSTP memos, for example, or the challenges of international collaboration.

Again RDA/US acts like a bridge or conduit between the national and global. It helps make international activity locally useful and helps local or national projects extend internationally. We also have specific initiatives like an early-career, training-the-workforce initiative. The idea is to help US lead and be more competitive in data science

So how does this relate to NDS?

Data is the hot topic right now. We need to capitalize on it. There's a lot of activity and new initiatives within NSF and elsewhere. We can't be overly competitive. Roles need to be defined. What are the gaps we need to fill? Maybe a better question is who are we missing? I really liked Margaret's comment on roles.

I think that is a good topic for this workshop. RDA is doing some things, but certainly not everything and I hope RDA and NDS can take advantage of the moment and develop a synergistic relationship that really makes data work.