

## *codataRDA Interest Group Charter Template*

***Name of Proposed Interest Group: Mapping of the Landscape of Research Data Activities***

***Introduction*** (A brief articulation of what issues the IG will address, how this IG is aligned with the RDA mission, and how this IG would be a value-added contribution to the RDA community):

The Internet now connects research data, computer resources and software from globally distributed resources in real time. Where on planet Earth these resources are geographically located is irrelevant, but to enable online access to them, there is a rising need for programmatic access to both data, and to software to find and process data across institutional, domain and national boundaries. This requires the development of standardized machine-to-machine interfaces that loosely couples data and software through agreed formats, interfaces, vocabularies and ontologies, preferably across multiple domains. The complexity of these online infrastructures require that they are built by much wider communities, through effective cooperation and governance, to enable new and innovative forms of interdisciplinary science from globally accessible data stores.

The time is ripe for identifying the key communities and partnerships within the major scientific domains that are developing infrastructures that enable sharing and processing of scientific data and ‘Mapping Of the Landscape’ (MOL) of these activities to further improve collaborations and partnerships, particularly those ‘umbrella’ alliances that are enabling interdisciplinary data sharing. The key advantage of a better Landscape Map is that researchers will know who is doing what where and hopefully avoid unintended duplication. Further, where duplicate or more activities are discovered, it is hoped that once groups are aware of equivalent activities, that MOL IG can help become a conduit where these groups can connect, share experiences and learned from each other to improve coordination and avoid any more duplication of effort.

At RDA Plenary 8 and Plenary 9 sixteen groups were identified undertaking “MoL” activities across a variety of data infrastructures and organisations. This not only reinforced that it was logical to attempt to coordinate all these MoL activities, but at the same time highlighted there was no agreed process on how to undertake a MoL activity so that outputs could be synthesised and leveraged

Key points identified at the P8 and P9 meetings were:

- 1) There was no agreed vocabulary or ontology to describe what research data infrastructures that each MoL is reviewing in a consistent way; and
- 2) That there was a diversity of infrastructures that each was trying to map (technology, data/information, computational systems, etc).

***User scenario(s) or use case(s) the IG wishes to address*** (what triggered the desire for this IG in the first place):

MOL activities identified so far are both within and across many scientific domains. These have similar goals and host parallel working groups that support the mission of advancing scientific research through

data interoperability. Several are looking for common ‘mapping’ methodologies so that ‘maps’ created by multiple groups can be interconnected and results shared.

**Objectives** (*A specific set of focus areas for discussion, including use cases that pointed to the need for the IG in the first place. Articulate how this group is different from other current activities inside or outside of RDA.*):

1. Develop a web page with a catalogue of MoL activities related to identifying research data infrastructures;
2. Develop a synthesis of existing MoL activities for research data infrastructure activities within and beyond RDA;
3. Investigate mapping practices including methodologies, tools, workflows, etc. and identifying whether any key pieces are missing; and
4. Discuss opportunities for collaborations on existing MoL exercises.

This group was partially informed by the RDA Atlas of Knowledge and TAB LOG mapping exercises, though in contrast to this activity, the proposed MoL IG will focus on activities eternally to RDA and at a higher organizational level.

**Participation** (*Address which communities will be involved, what skills or knowledge should they have, and how will you engage these communities. Also address how this group proposes to coordinate its activity with relevant related groups.*):

1. Arctic Data Committee Landscape Exercise (Peter Pulsifer, <http://arcticdc.org/products/data-ecosystem-map>)
2. EarthCube (<http://www.arcgis.com/home/item.html?id=9bde7150da474c828d61a5e67e98855d>, [http://www.goring.org/resources/neo4j\\_engagement.html](http://www.goring.org/resources/neo4j_engagement.html))
3. ESRI mapping tool (<http://dusk.geo.orst.edu/ec-story>) - developed by Dawn Wright of ESRI that was used to map the location of, and types of communities within EarthCube
4. Belmont Forum (Rowena Davis, [link])
5. Atlas of Knowledge (Simon Lambert, RDA/EU)
6. AuScope (Lesley Wyborn [link])
7. CODATA Task Group on Coordinating Data Standards amongst Scientific Unions (Marshall Ma, <http://www.codata.org/task-groups/coordinating-data-standards> )
8. TAB LOG (Steve Diggs, [link])
9. RDA Education IG connection? (Sophie Hou pointed to Amy Nernburger’s education landscape survey as a possible connection point at the AGU in-person meeting)
10. USGS Community for Data Integration (CDI) (Leslie Hsu, [CDI wiki](#)). Current working groups include Tech Stack, Semantic Web, Data Management, Citizen Science, Mobile App, and more. CDI Community can be engaged through Leslie Hsu, who coordinates communication to the 500+ members from within and outside of USGS. We have some initial coordination such as joint Tech Dive monthly calls with ESIP, and are interested in leveraging more opportunities, events, etc. to reduce redundancy and bring information to our members. Can serve as link to USGS data assets.
11. RISCAGE (European Research infrastructures in the international landscape) (Ari Asmi, [link])

12. ESIP (Earth Science Information Partners) (Erin Robinson, [link])

Related RDA groups

- TAB
- WG/IG Chairs
- [Education and Training on handling of research data IG](#)
- Brokering IG
- Data Foundations and Terminology IG

**Outcomes** (*Discuss what the IG intends to accomplish. Include examples of WG topics or supporting IG-level outputs that might lead to WGs later on.*):

Given that the landscapes of interest are eternally changing making a map or maps virtually impossible to keep current, this IG will instead focus on more manageable areas of alignment.

- Example WG topics:
  - Developing a vocabulary/ontology to describe components of research data infrastructures (that this does not exist has been a huge stumbling block for the MoL IG)
  - Mapping methodologies to document best practice for those wanting to undertake MoL's
  - Developing a portfolio of Landscape mapping tools and comparing/contrasting strengths and weaknesses of each
  - Example outcomes:
  - Recommendations for others working on 'mapping the landscape' activities to increase alignment and possible future integration.
  - Promotion of knowledge of existing exercises

**Mechanism** (*Describe how often your group will meet and how will you maintain momentum between Plenaries.*):

- ESIP meetings - ESIP runs two meetings each year in the US, one in Summer and one in Winter. Their off-Plenary schedule directly complements the RDA calendar and distance virtual meeting options are supported as part of the meetings.
- AGU meetings - The AGU Fall Meeting is a large international science meeting and attracted many interested MoL individuals this past year. We plan to continue taking advantage of this gathering as a feasible in-person discussion venue.
- telecons

**Timeline** (*Describe draft milestones and goals for the first 12 months*):

September 2017 - P10 session: Mini Summit and consolidation of work plan

December 2017 - AGU meeting and progress report

March 2018 - P11 session and revisit workplan

September 2018 - P12

*Potential Group Members (Include proposed chairs/initial leadership and all members who have expressed interest): Bold indicates co-chairs*

| FIRST NAME    | LAST NAME     | EMAIL                                 |
|---------------|---------------|---------------------------------------|
| <b>Lesley</b> | <b>Wyborn</b> | <b>lesley.wyborn@anu.edu.au</b>       |
| <b>Rowena</b> | <b>Davis</b>  | <b>rowenaidavis@email.arizona.edu</b> |
| <b>Ari</b>    | <b>Asmi</b>   | <b>ari.asmi@helsinki.fi</b>           |
| <b>Steve</b>  | <b>Diggs</b>  | <b>sdiggs@ucsd.edu</b>                |
| Helen         | Glaves        | hmg@bgs.ac.uk                         |
| Peter         | Pulsifer      | Peter.Pulsifer@colorado.edu           |
| Lindsay       | Powers        | lpowers@usgs.gov                      |
| Lynn          | Yarmey        | yarmel@rpi.edu                        |
| Colleen       | Strawhacker   | colleen.strawhacker@colorado.edu      |
| Dawn          | Wright        | DWright@esri.com                      |
| Jonathan      | Petters       | jpeters@vt.edu                        |
| Leslie        | Hsu           | lhsu@usgs.gov                         |
| Rebecca       | Koskela       | rkoskela@unm.edu                      |
| Ma            | Marshall      | max@uidaho.edu                        |
|               |               |                                       |
|               |               |                                       |
|               |               |                                       |
|               |               |                                       |