RDA Third Plenary Meeting

The Data Sharing Community: Playing YOUR part

26 - 28 March 2014, Dublin, Ireland

"We are taking our work beyond Europe's borders, to reach global scale. To make the scientific resources of the world work together, interoperating and open to discovery. For example we are working with partners like the US and Australia in the Research Data Alliance to make scientific progress broader, deeper and more workable". Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda - Open Access to science and data = cash and economic bonanza, 19 November 2013

We live in the data generation. All of society's grand challenges—be it addressing rapid climate change, curing cancer and other disease, providing food and water for more than seven billion people, understanding the origins of the universe or the mind—all of them require diverse and sometimes very large data to be shared and integrated across communities, scales, and technologies. The Research Data Alliance (RDA) is building the social and technical bridges that enable open exchange of data on a global level.

Why is this important? Scientists and researchers around the world are conducting millions of experiments every day and according to statistics more than 80% of the data generated by them sits on their computers never to be shared, exchanged or preserved for future use. Many times this is not tied to a lack of willingness to share research data but due to technical challenges in converting, accessing and storing it in the right place, right format so that while the originator of the data still owns it he / she can offer it for use by others now and in the long term future to create new experiments and find new answers and solutions to the societal challenges faced globally.

'The Data Sharing Community: Playing YOUR Part'

Over 1600 RDA members, experts from all over the world in various different fields, are working together to solve many of the data challenges that exist today. Ireland co-hosted the third biannual meeting of these experts and from 26 to 28 March 2014. 500 members gathered in Croke Park to work intensively on problems ranging from how to address

scientific community needs of utilizing big volumes of data, to issues related to data importance for the development of global agriculture, to promoting & sustaining wheat data sharing, reusability and operability to the development of a common global framework for the management of marine data.

Photo credits: Johnny Bambury/Photographer

The Research Data Alliance enjoys the privilege of performing inter-disciplinary and multi-disciplinary research on data by some of the best professionals from all over the world. Having a trend to work in working groups and interest groups, RDA members are trying to address big issues by fragmenting them.

An example of this is the PID Information Types WG (PIT WG), which at the 3rd RDA Plenary, focused on in-depth conceptual and technical discussions on the scope and functionality of its main output, an Application Programming Interface (API) for interaction with typed information closely associated with Persistent Identifiers (PIDs). While the Structural Biology interest group was filled with ideas, interest and curiosity. From a “data generator” and user point of view, it would appear that there are several road blocks which need attention beginning from the lower levels of the data life cycle, such as the varied use of techniques within structural biology, availability of wet laboratory protocols, accessing primary unpublished data, data management, and maintaining and developing SB data and meta data. From the discussion at the end of the seminar, it was evident that these issues need to be tackled from the level of the researcher.

The feedback and engagement of the Data Citation WG participants was remarkable. Not only have the existing concepts immediately been challenged by insightful questions, but the amount of pilots that have been proposed was impressive, a total of nine additional use cases from equally many disciplines. The domains that have been suggested ranged from nuclear physics via oceanography towards the humanities and included a highly diverse set of data formats and technology stacks.

The Community Capability Model Interest Group (CCM IG) highlighted areas of particular attention in the development and employment of the profile tool, in particular: legal, ethical and commercial issues; gaining informed consent for reuse and repurposing; appraisal and quality control; trustworthiness; scale and complexity of data; publication and sharing; citation attribution and accreditation in scholarly communications. All these issues are, of course, topics at the heart of many discussions within RDA as a whole. Consequently, by including all these areas in their work, the CCM IG members are concerned with a whole range of pressing issues that are of interest for RDA

members. And indeed, it is the core purpose of the IG to do this work for the benefit of the whole Alliance.

Social media awareness amongst the participants was incredible with the Twitter walls at the meeting were constantly updated with information about the sessions, thoughts and ideas to move on. This way people could interact, and somewhat follow the overlapping sessions – and check the Tweets later on. One of my favourite Twitter quotes at the meeting was: "You have to take care of your data, otherwise it will be digital dust in a few years." (Rob Baxter, EPCC/ EUDAT)

As a first-timer on the RDA Plenary, you are not sure exactly what to expect. A lot of people interested in the field of research data? Sure. Many presentations on current trends and challenges? Of course. But what was most striking most during these days in Dublin were the constant, on-going intense discussions which took place literally everywhere. People gathering in small groups in the corridors, outside the conference venue, even in the rest rooms, not just to catch up but to actually talk about issues related to research data.

Rather than being a niche gathering, it was ***a place for people who "care about how the sharing of research data can progress to discoveries that have the potential to be of benefit to all,"*** said Dr Ruth Adler, the Australian ambassador to Ireland. ***Research data is costly enough and difficult enough to generate in the first place; having it sitting forgotten on a hard drive somewhere, never to be shared, is not only careless but also not in the spirit of scientific endeavour***.

Croke Park, 3rd Plenary meeting venue, is a historical land mark in Ireland and represents the strength and resilience of the Irish people during times of turbulence. In unison with this, research is continuously evolving and changing with data being generated exponentially so implementation of the policies and solutions proposed by the RDA to accommodate this will be challenging. As challenging as this may be, the eagerness and attentiveness of the people present at the meeting was undeniable suggesting that the RDA will succeed in overcoming these issues undoubtedly.

The next RDA Plenary meeting takes place in Amsterdam 22-24 September 2014, you can't afford to miss it. https://www.rd-alliance.org/rda-fourth-plenary-meeting.html

Join this global endeavour, register to the on-line community and become a Member of RDA it's open to all and free of charge https://www.rd-alliance.org/user/register

RDA Third Plenary in a Nutshell

The 3rd Plenary focused on the theme 'The Data Sharing Community: Playing YOUR Part' and about exploiting RDA's work to date to its full potential. The program offered a series of keynotes, panels, networking, Working and Interest Groups as well as 'Birds of a Feather' sessions on disciplines ranging from agriculture to particle physics, and from humanities to bioinformatics, with a cross-disciplinary approach to key data challenges including those linked to sharing, access, preservation, re-use and citation.

All parts of the data lifecycle are being addressed, from foundational data terminology to data publication and re-use. Furthermore taking part in the Plenary allowed participants to be directly involved in the shaping of RDA and to get up-to-date on key pertinent issues.

The plenary meeting was opened by the Irish Government and the Government of Australia, and daily keynote addresses were given by Prof. Ian Chubb, Australia's Chief Scientist, Dr Tony Hey, Vice President of Microsoft Research Connections and Prof. Milena Žic Fuchs, Chair of the Standing Committee for the Humanities, European Science Foundation.

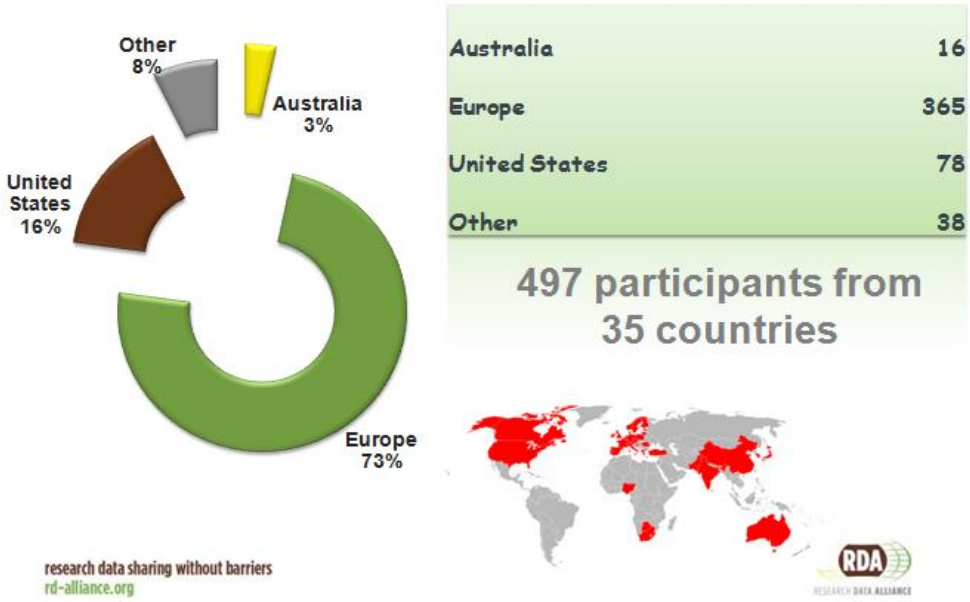
The RDA Third Plenary Meeting was co-organised by RDA-Europe, Australian National Data Service, the Digital Repository of Ireland and Insight with the support of the Science Foundation Ireland, the Irish Research Council, Failte Ireland, meetinireland.com, and Microsoft Ireland.

> Dr Sandra Collins, Director of the Digital Repository of Ireland, added *"We are very proud to have brought the RDA conference to Ireland – our core mission is to preserve, discover and share data for the Humanities and Social Sciences, and RDA is central to achieving this on a global scale. Ireland can show great leadership in this area."*
>
> Prof Alan Smeaton, Director of Insight Centre for Data Analytics, DCU, said *"There is enormous potential for benefits to be gained by adopting a data-driven approach to solving problems throughout society. Bringing the RDA meeting to Ireland allows more Irish researchers to become directly involved in fashioning how data, in all its forms, can be used to best effect."*

Who attended?

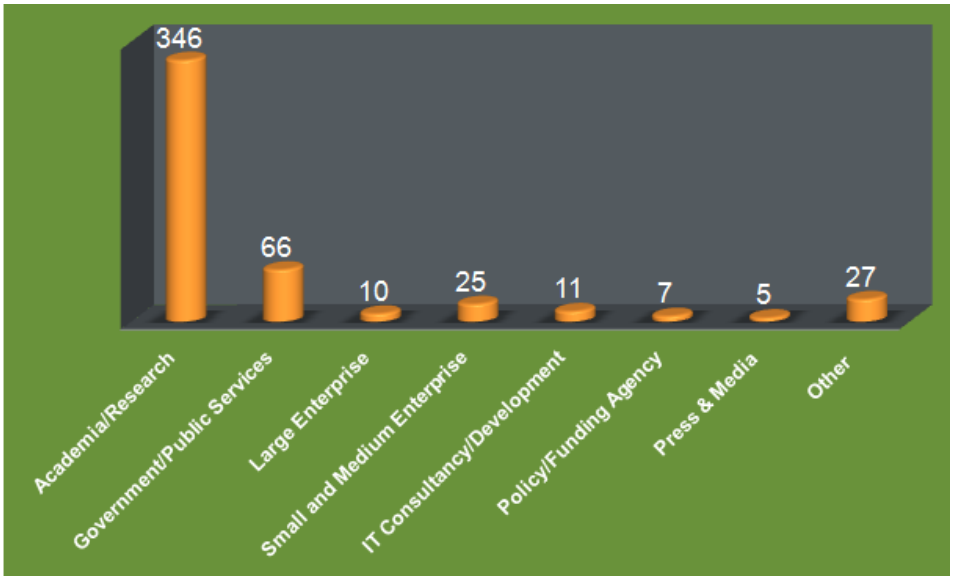
500 RDA members met in Dublin and a total of 35 countries were represented. The largest regional grouping came from Europe with 73%, much to be expected given the location of the meeting in Europe. The United States represented 18% of the participants with the Australian contingent significantly higher than previous events. The list of

registered participants is available at https://www.rd-alliance.org/rda-third-plenary-participants.html.



Plenary3 participants break down by type country

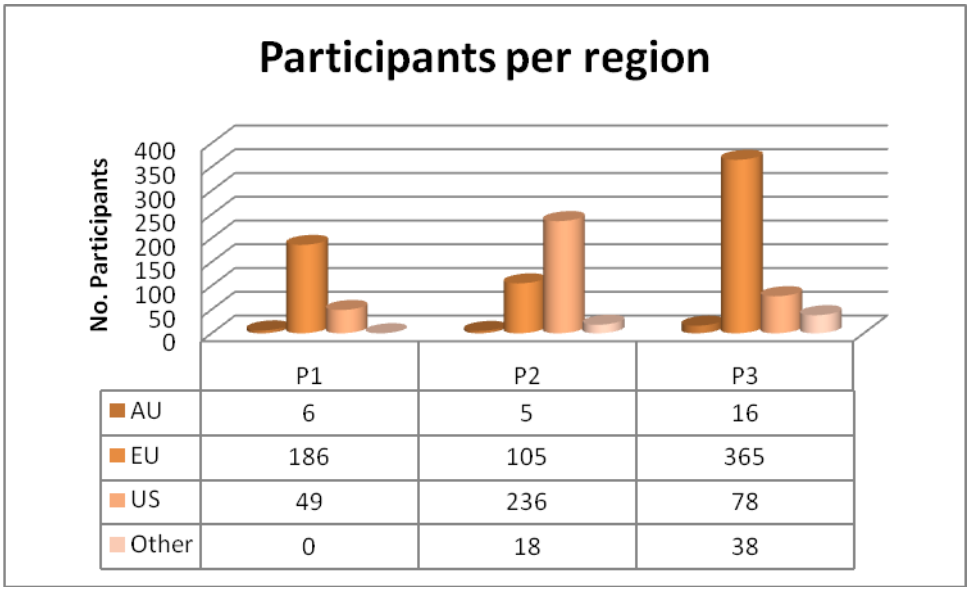
Representation from academia and research is still the highest at 70% which very much reflects the main RDA target audience.



Plenary3 participants break down by type of organisation

Plenary3 participants break down by professional title

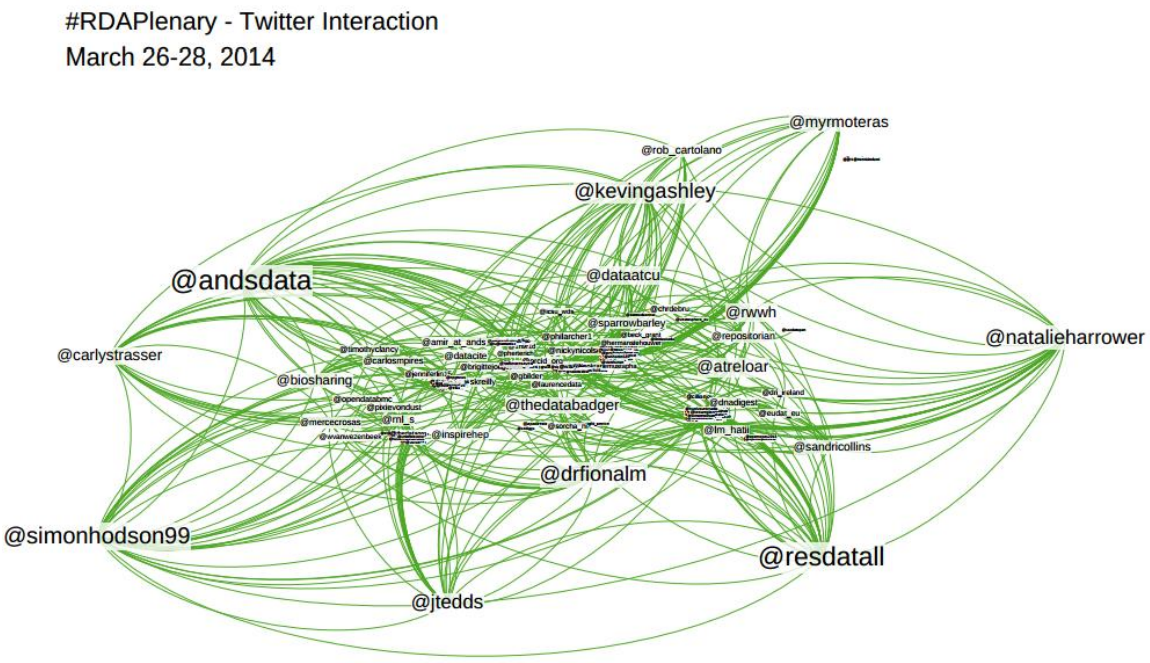
Growth of Participation by region



	P1	P2	P3
AU	6	5	16
EU	186	105	365
US	49	236	78
Other	0	18	38
Total	**241**	**364**	**497**

Communication & Social Media

During the 3 day event, close to 3000 tweets have been made using the #rdaplenary
Not only that, there was wide collaboration and interaction between participants using Twitter as illustrated by the interaction graph below



Plenary 3 Twitter interaction graph #rdaplenary

An extended analysis of the social media coverage of the event is available at https://www.rd-alliance.org/plenary-meetings/third-plenary/plenary-3-communication-social-media.html

Programme:

The detailed programme and presentations are available at: https://www.rd-alliance.org/plenary3/programme.html

Post-event Survey

A post event survey was set up to be completed by the participants. The survey respondants have, among other rated the Working groups & Interest group meetings as the most beneficial aspect of the RDA Third Plenary meeting , followed by meeting colleagues and content & speakers. Working with colleagues was the main reason for attending, the interest in content and professional development.

RDA Plenary 3 Meeting Co-located Events

A number of co-located events took place on Tuesday 25th March including the EUDAT Training: Fundamentals of data Infrastructures, MUMIA Mtg on Verifyable Results in Multi-lingual/Multi-faceted Search: Challenges in Sharing Data, Tools and Results Workshop, SIM4RDM: Building Collaborations to address research data management workshop and APARSEN – EUDAT – SCIDIP-ES Workshop on Data Preservation and Reuse.

The co-located event details have been published at https://www.rd-alliance.org/rda-plenary-3-meeting-co-located-events.html

Irish Research Council Poster Session

The Irish Research Council (IRC) has funded a poster session and reception on the first day of the RDA Third Plenary Meeting - 26th March 2014, during which some of the multi-disciplinary IRC-funded PhD and Post Doc research that demonstrates open access and sharing of publicly funded research data, or uses (exploits) openly accessible publicly funded research data was showcased.

Using a crowd-sourced approach, delegates at the meeting voted for the "best" poster. Eavan O'Dochartaigh from NUI Galway was the author of the winning poster "Defrosting the Data: Exploring the Art and Literature of the Franklin Expeditions to the Arctic" that has received a research networking and travel grant of €1,000. The award was presented by Dr. Eucharia Meehan, Director of the Irish Research Program during the RDA plenary 3 evening reception on Wednesday 26th March at Croke Park Conference Centre.

European Early Career Researchers & Scientists working with Data

RDA Europe has supported 22 European Early Career Researchers & Scientists working with Data to attend the Third Plenary meeting. The aim of this programme was to introduce European early career researchers & scientists to RDA, highlighting what data scientists / practitioners are doing and leveraging on the Early Career Researchers & Scientists knowledge to support the Plenary 3 activities and support the Working & Interest group activities.

Furthermore an Early Career Researchers & Scientists Poster Session was scheduled for Thursday 27th in the morning, on Day 2 of the Plenary.

18 blog entries, including articles from the RDA Europe Early Career Researchers & scientist Grant Winners, covering the Third Plenary meeting have been published on the website https://www.rd-alliance.org/blog.

This report includes excerpts of these and other overviews of the RDA Third Plenary.

Short report from the RDA 3rd Plenary Meeting – Day 1

Written by Vassilis Protonotarios, Agro-Know Greece[1] -

The 3rd Plenary Meeting of the Research Data Alliance (RDA) took place between 26-28/3/2014 in Dublin Ireland. It was my 2nd participation in a row to an RDA plenary meeting and I managed to organize my trip (even at the last moment) and be there.

...

Day 1 Plenary Session The plenary session started with keynote speeches and greetings; I personally enjoyed more the Keynote Address by Prof. Ian Chubb AC, Australia's Chief Scientist. Some of his key-points, as extracted by various tweets were the following:

We need national, international and inter-disciplinary collaboration in research and innovation;

We need research and data to be able to feed 9 billion people, producing carbohydrates and fiber while climate moves;

You can also watch his presentation recording http://media.heanet.ie/page/829f40520371455bb81e96de81c4f1bd.

Additional presentations/speeches took place afterwards, mostly highlighting the fact that data exist and it is up to the users to find a meaningful way to use them and that infrastructure is already here, waiting for useful applications. The highlight was a cartoon presented by Dr. Ross Wilkinson, Executive Director, Australian National Data Service, showing a donkey, a cart and a carrot; an image really familiar to many of us. This led to nice and funny discussions as well as an explosion of related tweets! In general, the topics were

[1] https://www.rd-alliance.org/blogs/short-report-rda-3rd-plenary-meeting-%E2%80%93-day-1.html

focused on the importance of open data, the role of the e-infrastructures and policies as well as ways to open up existing data.

Day 1 WG/IG sessions

I opted to attend the "BoF Education and skills development on Data Intensive Science" organized by Yuri Demchenko and Wouter Los , which aimed to identify opportunities for the new field of data scientists. Miguel-Angel Sicilia from the University of Alcala was also there, proposing his approach on the subject and discussing the possibility of an interest group, a proposal of which has already been submitted. Discussions were interesting and focused on the existing curricula all over the world. Next was the meeting of the Agricultural Data Interoperability IG, which was chaired by agINFRA colleagues Johannes Keizer (FAO) and Devika Madalli (Indian Statistical Institute, Bangalore). There I made a presentation titled "Global RDF Descriptors for Germplasm Data", describing the work done in the context of the agINFRA project and the RDA WG towards the exposure and publication of germplasm data as linked data (always based on the work already done by other experts in this field). It was followed by another presentation by Esther Dzale from INRA, about the Wheat Data Interoperability Working Group and then there was a discussion on various topics affecting the group.

...

RDA Plenary 3: Data Type Registries WG

Written by Lisa Donatini, Università di Pisa, Italy[2]

In Dublin it was my first time at a RDA Plenary Meeting.

I was not sure what to expect, I went to the conference with curiousity about a subject that I've dealt with only marginally in my work, but that I think it's becoming more and more crucial for many aspects of research, in many different fields. I have to say that a thing that really hit me, right from the very beginning of this conference, was the passion and partecipation. Everybody speaking seemed to firmly believe in this project, and the multidisciplinarity of the place was just amazing. At your typical conference, you don't see altogether people from humanistic and scientific backgrounds, engineers and biologists, computer scientists and law students, and so on.

Seeing such a large and various crowd was very inspiring to me. I believe in the importance of keeping an eye on what happens outside of your usual field, because it enriches you as a person and thus as a professional too, and it can give you new perspectives, exposing you to new approaches. And it's always good to be aware of what's going on outside of your lab.

[2] https://www.rd-alliance.org/blogs/rda-plenary-3-data-type-registries-wg.html

Beside, the organization of the conference as all was just great.

I was assigned to work with the Data Type Registries WG, and it was a really nice experience, especially because in this group, that has been very active, you could already see many ideas taking the shape of results.

Dr. Larry Lannom and Dr. Daan Broeder were the co-chairs of the meeting. Dr. Lannom made a presentation summing up the work done until now and what to do next. The audience was very active too, showing propositions and interest towards the subject, talking about the use cases and related efforts. We had a contribution from Dr. David Giarretta from Alliance for Permanent Access, presenting some slides about the work his company is doing.

This WG has already been doing some tangible work. The idea revolves around the concept of type of data, that is the characterization of data structure at multiple levels of granularity, from individual data points up to and including large data sets. The aim is to build a common data model and expression to describe types: if these types are standardized, this will make it easy to add them to registries, in order to have standard ways to discover and treat data; so we could guarantee interoperability and, as additional step, offer a common API for machine consumption.

For example, we can imagine users having some datasets, and we can imagine some tools that will easily discover the standardized types associated to the dataset, so that the user can know how to treat the data. We can also imagine to have some services that will process the data directly according to their type.

These types can also find application in certification and access control (that is: for this type, there are these rules), or for data acquisition and experiments. The possibilties offered by such an approach arise also some related sets of problems that will need to be further discussed (for example about metadata, about replication of information), while keeping in mind use cases and the needs of those who work with data.

During the meeting, a prototype for data type registries was shown: each type should have an ID, human descritpion, provenance, properties, etc. There can be some "primitive types" that can be used to define new types.

Now the challenge is to go from the prototype to the real usage and to evolve the data model.

I'm looking forward to the Fourth Plenary Meeting, I really hope that I will be able to take part and I am curious to see how the many ideas I've seen in these days will develop in the next months (and years).

Let's keep up the good work!

Early Career Researcher and Scientist support programme…A newcomers perspective to the RDA!

Written by Nicola Kavanagh, Royal College of Surgeons Ireland[3]

...

A Newcomers session was held in the Royal Irish Academy in Dublin city centre prior to the conference, which I felt was very beneficial. Being able to put faces to the names involved in the organisation and establishment of the RDA was, in my opinion, very valuable. It made the entire experience quite personal and helped me realise that my opinion counted.

The plenary session was hosted in the Croke Park Stadium, which was very apt for the event. Croke Park is a historical land mark in Ireland and represents the strength and resilience of the Irish people during times of turbulence. In unison with this, research is continuously evolving and changing with data being generated exponentially so implementation of the policies and solutions proposed by the RDA to accommodate this will be challenging. As challenging as this may be, the eagerness and attentiveness of the people present at the meeting was undeniable. This suggests to me that the RDA will succeed in overcoming these issues undoubtedly.

The Structural Biology interest group meeting was filled with ideas, interest and curiosity. From a "data generator" and user point of view, it would appear that there are several road blocks which need attention beginning from the lower levels of the data life cycle, such as the varied use of techniques within structural biology, availability of wet laboratory protocols, accessing primary unpublished data, data management, and maintaining and developing SB data and meta data. From the discussion at the end of the seminar, it was evident that these issues need to be tackled from the level of the researcher.

.....

Experiences from the RDA 3rd Plenary

Written by Anni Jakobsson, CSC, Finland [4]

[3] https://www.rd-alliance.org/blogs/early-career-researcher-and-scientist-support-programmea-newcomers-perspective-rda.html

[4] https://www.rd-alliance.org/blogs/experiences-rda-3rd-plenary.html

...

It was just a bit more than a year ago when the Research Data Alliance was launched – and the organizers were happy to host almost 500 participants from around the world. The Plenary started with introductory talks, followed by more hands-on Interest Group and Working Group meetings and BoF sessions (Birds of a Feather).

I took part to the introductory session for Data Publishing Interest Group. Ambitions were high - the session was so popular it ran out of seats. And this wasn't the only time during the Plenary the rooms were full – you could really feel the excitement in the air!

I met some of the Early Career Scientists which RDA gave support to attend the meeting. They represented various disciplines, multiple countries and both genders, which I was pleased to notice.

...

Whilst some of the Plenary sessions were more traditional talks with a Q&A sessions, the Interest and Working groups were more varied, dialogic and concentrated on solving a certain problem in the world of global research data. Some of the groups were working on documents – deliverables – to be published after the next RDA Plenary in Amsterdam on 22-24 September 2014.

It was also nice to notice how social media aware the people at the Plenary were. Twitter walls at the meeting were constantly updated with information about the sessions, thoughts and ideas to move on. This way people could interact, and somewhat follow the overlapping sessions – and check the Tweets later on. One of my favourite Twitter quotes at the meeting was: "You have to take care of your data, otherwise it will be digital dust in a few years." (Rob Baxter, EPCC/ EUDAT)

...

RDA Plenary 3 in Dublin 26-28 March – experiences of a first-timer

Written by Marie Sandberg, CSC, Finland[5]

As a first-timer on the RDA Plenary, you are not sure exactly what to expect. A lot of people interested in the field of research data? Sure. Many presentations on current trends and challenges? Of course. But what struck me the most during these days in Dublin were the constant, on-going intense discussions which took place literally everywhere. People gathering in small groups in the corridors, outside the conference venue, even in the rest rooms, not just to catch up but to actually talk about issues related to research data.

When asking a person I used to work with what he was expecting from the Plenary he simply raised his eyebrows like it was a redundant question and said "you know, everybody who is anybody in this field come to the RDA Plenaries, you simply have to go if you want to meet them all in the same place." But it is not just about the "anybodies" - the people who are already well-known for their achievements. Many young people also attended the Dublin Plenary, much thanks to the "Early Career Researcher and Scientist support programme" which allowed for 22 persons from all over Europe to take part in the conference. Another contributing factor was the Irish Research Council's Poster Session, where part of the multi-disciplinary IRC-funded PhD and Post Doc research was showcased. When speaking to some of the researchers showcasing their posters, all of them concluded that the crowd showed a real interest in their research and additionally asked "unusually clever questions".

The high level of ambition could also be noticed in the various Working Groups, Interest Groups and BoFs – where the genuine interest of actually wanting to solve issues was dominating the atmosphere. Not much time was spent on lengthy introductions in these sessions, which may be a bit challenging for a newcomer, but is a must to be able to reach some real progress in a limited amount of time. I spent much time in these sessions trying to scribble down the essence of interesting discussions like "if there is a transferable cost model for data curation" or "how to make researchers aware of workflows for data publication". The overall feeling when leaving Dublin was that the RDA Plenary is not just another conference – it's a place to get things done.

[5] https://www.rd-alliance.org/blogs/rda-plenary-3-dublin-26-28-march-%E2%80%93-experiences-first-timer.html

The WG Data Citation Session

Written by Stefan Proell, SBA Research, Austria[6]

...

Although just one year has passed between the initial event and the third meeting, the results and the engagement that the RDA community has produced so far are incredible. At the first meeting, the lively atmosphere in and between breakout sessions was already an indicator for the dedication for data sciences that characterises the newly formed RDA community. The genesis of working and interest groups was particularly interesting to watch, especially as I had the opportunity of being involved in the WG Data Citation (WGDC) from the very beginning. Dynamic data citation is my primary research area that I am interested in the scope of my PhD thesis. Therefore, I was particularly happy that I could be a part of this group and learn from experts on that field from the very first day.

The initial birds of a feather session that was held during the first plenary meeting was highly important to define the scope of what eventually became the WGDC. After almost one year of iterative improvement of the initial ideas and core concepts WGDC was endorsed and could be presented to a broader audience as an official WG within the umbrella of the RDA. Watching this process was very exciting and informative. My personal highlight of the third plenary meeting in Dublin was participating in the "Making Dynamic Data Citeable Session", which was chaired by Andreas Rauber, Ari Asmi and Dieter van Uytvanck.

The session was opened by an introduction and an overview of the achievements that have been made by the WGDC. The presentation was followed by a retrospection of the core principles and the presentation of the six pilot use cases that have already been submitted during the formation phase of the working group. The main goal of the session was to encourage experts in a broad and diverse range of domains to submit their use cases, based on which the core principles developed within the WGDC will be tested. Furthermore, the use cases should serve as a source of challenges that require new concepts in the area of data citation that need to be tackled.

The feedback and engagement of the participants of the WGDC session was remarkable. More than 30 experts from 25 organization participated in the discussions***. Not only have the existing concepts immediately been challenged by insightful questions, but the amount of pilots that have been proposed was impressive. At the end of the session we counted nine additional use cases from equally many disciplines. The domains that have been suggested ranged from nuclear physics via oceanography towards the humanities and included a highly diverse set of data formats and technology stacks.*** The session was closed with the solicitation of posting details about the use cases to RDA wiki and engaging in discussions on the group mailing list.

The RDA meeting effectively demonstrated the importance of scientific data for an incredibly diverse set of communities across all continents. The engagement of the RDA members and their exchange

[6] https://www.rd-alliance.org/blogs/wg-data-citation-session.html

of ideas show that we all can benefit from exchanging and sharing data. What we need is agreements on technology, on policies and practices, which be achieved best by meeting face to face and discuss the questions that we need answers for.

How domain champions can aid the acceleration of data-intensive research in their disciplines - the Community Capability Model Interest Group's state of affairs at the time of the 3rd RDA Plenary

Written by Tanja Friedrich, Researcher, GESIS Leibniz Institute for the Social Sciences, Germany[7]

At the Third Plenary for the Research Data Alliance, held from March 26 to 28 in Dublin, the Community Capability Model Interest Group presented its development and progress since the last plenary, described current challenges, and investigated the next steps to take. With the objective of assessing capability to do data-intensive research in a broad range of disciplines, the group will continue to engage communities of researchers to complete the Community Capability Model Framework profile. "Domain champions" from within RDA can give valuable support for this endeavour by helping to adapt the tool to their disciplines and by acting as a link to their communities.

...

In the current context, a community is a group of principal investigators from one discipline. By identifying areas for change and investment, the tool contributes to the assessment and enhancement of a community's capabilities. Its general applicability makes the model an aid for decision making and planning for different stakeholders. Furthermore, in the light of the RDA's mission of building bridges, it provides deep insights in the social, technical and organisational structures of research data management in each investigated discipline. These insights are intended to form a productive resource for all present and future Working and Interest Groups of the Alliance.

At the 3rd RDA Plenary the Community Capability Model Interest Group (CCM IG) presented their work done since the 2nd plenary and discussed further development. In the beginning the chairs highlighted areas of particular attention in the development and employment of the profile tool, in particular: legal, ethical and commercial issues; gaining informed consent for reuse and repurposing; appraisal and quality control; trustworthiness; scale and complexity of data; publication and sharing; citation attribution and accreditation in scholarly communications. All these issues are, of course, topics at the heart of many discussions within RDA as a whole. Consequently, by including all these areas in their work, the CCM IG members are concerned with a whole range of pressing issues that are of interest for RDA members. And indeed, it is the core purpose of the IG to do this work for the benefit of the whole Alliance.

---

[7] https://www.rd-alliance.org/blogs/how-domain-champions-can-aid-acceleration-data-intensive-research-their-disciplines-community

Since its first Meeting at the 2nd RDA Plenary the CCM IG has improved the profile tool, held two workshops at international conferences and completed two case studies. The CCMF profile tool is now implemented as an MS Excel spread sheet, containing separate worksheets for each of the eight CCMF factors. The community-specific characteristics of these factors can be assessed with a scorecard tool.

...

Accomplishing Bit Preservation! Representation Information!! Provenance and Context Content Standards!!!

Written by Naresh Kumar, CNR- ISTI, Italy[8]

…

The RDA organisation is enjoying the privilege of performing inter-disciplinary and multi-disciplinary research on data by some of the best professionals from all over the world. Having a trend to work in working groups and interest groups, RDA members are trying to address big issues by fragmenting them.

One such group is the Preservation e-Infrastructure Interest Group that took two sessions to discuss different issues. This group is dealing with Issues related to the infrastructure setup for preservation of digital objects. Members of the interest group (IG) believe that the Open Archival Information System (OAIS) reference model is a good base for any infrastructure for preservation. Now IG works for further build up on the OAIS reference model to set need oriented appropriate infrastructures for preservation.

...

RDA Plenary 3: Research Data Provenance IG

by Simone Roma, University of Pisa, Dept. Information Engineering, Italy[9]

...

Now I understand the needs for data sharing and communication between researchers that belong to different communities: free and direct data access allows to improve your research, speed up your tasks, and validate your results. For this reason, it is absolutely mandatory to regulate the manner in which the data is available to the scientific community.

---

[8] https://www.rd-alliance.org/blogs/accomplishing-bit-preservation-representation-information-provenance-and-context-content

[9] https://www.rd-alliance.org/blogs/rda-plenary-3-research-data-provenance-ig.html

I was assigned to the Data Provenance Research group. In particular, this group attempts to solve the following problem: heterogeneous digital data that have been produced by different communities with varying practices and assumptions, and that are organized according to different representation schemas, encoding, and file formats, present substantial obstacles to efficient integration, analysis, and preservation. This is a particular impediment to data reuse and interdisciplinary science. They present two complementary conceptual models for data representation, the Basic Representation Model and the Systematic Assertion Model.

At the begin, the discussion focused on these conceptual models. Then, we attempted to provide some use cases. Having some concrete examples could help either to set the goals of this group, either to develop a framework for identifying provenance models to apply to specific use cases. In addition, WG/IG overlap must be evaluated: we may need to have a multi-pronged approach.

...

RDA P3: PID Information Types WG session summary

Written by Tobias Weigel , German Climate Computing Center (DKRZ), Germany[10]

The working session of the PID Information Types WG (PIT WG) at the 3rd RDA Plenary focused on in-depth conceptual and technical discussions on the scope and functionality of its main deliverable, an Application Programming Interface (API) for interaction with typed information closely associated with Persistent Identifiers (PIDs). The session also addressed critical issues of typing for PID information in a larger context and the implementation and finalization plans for the remaining WG lifetime. Towards the end of the session, some motivating type examples were gathered which will illustrate the intended use of the PIT mechanisms.

...

The second session included a productive and creative brainstorming on possible type examples that should be included in the final WG deliverables to better illustrate the use of the PIT mechanisms. The examples gathered during the session show that the essential mechanics have been understood after the first session cleared up important terminological ambiguities. The gathered examples clearly show that a core set of value types can be determined, which particularly includes many uses of temporal and non-atomic (tuple) information. It was pointed out that the issue of encodings should be held back for the moment, transferring responsibility for serialization and deserialization to clients.

The session concluded with a short discussion of next steps. Implementation of the Java prototype will be done over summer, and the design and precise documentation of the RESTful API take a high priority since it represents a protocol specification highly desired by future adopters.

---

[10] https://www.rd-alliance.org/blogs/rda-p3-pid-information-types-wg-session-summary.html

RDA Plenary 3: Data In Context Interest Group

Written by Alessia Bardi, ISTI-CNR, Italy[11]

The goal of the group is investigating current practices to contextualise research data and possibly define contextual data profiles that may be used in the curation lifecycle to describe the context of research datasets.A data profile defines the metadata that must be collected to describe the object in the research domain. Typically, a data profile puts together elements from different standard metadata schemas, vocabularies and ontologies. The elements included in a data profile depend on the use case the profile is addressing. A number of existing life cycle models for the management and curation of research data must be taken into consideration, as well as the different stakeholders involved.

...

The chairs also underlined the importance of collaborating with other RDA interest and working groups and introduced delegates from the related RDA groups.

...

About 25 people took part in each session. Attendees actively participated to both sessions with questions and comments highlighting the importance of collaboration with other RDA and W3C groups. Questions from the floor underlined the need of a more detailed explanation of the relationship between the use case template and the data profiles. It was argued that data profiles need to be different based on the domain and should include metadata collected by users in addition to metadata added by data creators/curators. In order to address both issues, the group should first perform the analysis of existing life cycles and understand how they are implemented in different use cases.

...

Cloud Computing and Developing World Research

Written by Yuri Malitsky, University College Cork, Ireland[12]

The Cloud Computing and Developing World Research Interest Group postulates that providing the researchers of the developing world an infrastructure for cloud computing will allow for an even playing field with the rest of the world. Simply put, such an infrastructure is paramount in helping to provide solutions to the critical problems these researchers may encounter. After all, cloud computing offers a number of immediately obvious benefits,

[11] https://www.rd-alliance.org/blogs/rda-plenary-3-data-context-interest-group.html

[12] https://www.rd-alliance.org/blogs/rda-plenary-3-cloud-computing-and-developing-world-research.html

amongst which is the ability to gather data from geographically disparate locations and then performing computationally expensive analysis on remote machines. All while circumnavigating any infrastructural difficulties that might be encountered in any specific lab.

While the ultimate objective and benefits of the intrest group are clear, the path there is more opaque. As an example, consider the findings of European Commission funded ei4Africa project. After recently compiling an extensive survey of research labs in Africa, it was found that among the top most necessary resources for establishing e-Infrastructures in Africa are fast internet, computer hardware, and education. While all three are worthwhile problems to tackle, during the 3rd RDA plenary meeting Dr. Hugh Shanahan, co-chair of the interest group, advocated the need to focus on the education aspect.

To start the discussion, Dr. Hugh Shanahan emphasized the long term nature of the desired process, and citing that in the past there have been too many well-intentioned projects that dwindled away once either funding or interest dwindled. It was therefore the underlying current of all subsequent discussions that what was necessary was the establishment of a perpetually continuing solution. This thus led to the idea of setting up a summer school and some sort of accreditation for Data Analysts.

The ideas exchanged about the formation of a summer school revolved around the success of University of Michigan, which for the last 12 years has been running a Statistical Analysis workshop in Cape Town. It was brought up that the success of this program stemmed largely from the formation of partnerships with local communities. In this system, the volunteer instructors came to first teach the initial round of students, and in subsequent visits sought old participants to help and later guide the courses themselves. A teaching the teachers paradigm. Furthermore, while in the beginning the workshop was funded by a foundation, it was noted that it was not overly expensive, with most costs eventually covered by small grants. The proposed strategy was therefore to price the summer school to pay for itself and then to go find scholarships from places like Microsoft and Google.

Yet while setting up a summer school was most likely the best course to pursue, would such a curriculum be helpful to anyone? A person that finished the curriculum would likely need to later prove to future employers their knowledge of the subject. To help with this, the idea of an accreditation system, much like the standard exams used in actuarial science, was brought forth. The meeting, however, was split in the adoption of such a broad and lofty goal. Especially since the field of Data Analytics itself is still so nebulous. So the group proposed to spin the issue into an alternate and separate RDA interest group.

After a session packed with content and discussion, the group agreed that the next step forward would be to establish a more concrete proposal for a summer school prior to the 4th DRA plenary meeting in Amsterdam.

Summary of the Long Tail of Research Data Interest Group

Written by Artemis Lavasa – CERN & ATEITH, Greece[13]

The Long Tail of Research Data Interest Group session, which was chaired by Kathleen Shearer (COAR) and Wolfram Horstmann (The Bodleian Libraries) was divided into two parts on Friday, both of which attracted a rather significant amount of people.

Session 1 was dedicated to scoping the landscape and delving into the current situation in the area of long tail data. The purpose of the session was to explore how long tail data is being managed through several examples. The examples were separated into external services, institutional services and research solutions. The topic of long tail research data, generally characterized as small and/ or multidisciplinary data sets that fall outside the scope of the big data repositories, is very current. It could also be said that it is generating a lot of interest as reflected by the over-whelming response to the call to contributions to this session.

Within 90 minutes, 13 examples were presented, namely: Dryad, Scientific Data, F1000 Research, Ubiquity Press and Zenodo in the external services category, the California Digital Library/ UC3, Oxford, Columbia, the Notre Dame /Northwestern/ Indiana/ Cincinnati/ UVa collaboration and the University of Leicester in the institutional services category and finally the Strasbourg Astronomical Data Center, SiDORA (Smithsonian) and Scratchpads in the research solutions category. The presentations, even though brief due to time restrictions, were very informative and to the point and succeeded in displaying the main features of each of the services. They are available on the Long Tail of Research Data webpage: https://rd-alliance.org/internal-groups/long-tail-research-data-ig.html

Session 2 began with a presentation of the results of a survey of current practices for discovery of research data in repositories by Kathleen. The survey targeted long tail repositories and received 60 responses, 30 of which were complete. It was noted that the number of responses to the survey are not a representative sample of data repositories, but rather an indication of which way the wind is blowing. The survey found that Dublin Core and DataCite metadata were the most common schemas used in the data repositories and less than half of the respondents were using DOIs. In terms of discovery, most respondents indicated that the metadata was sufficient for users to find the datasets in the context of searching within the repository, however, the metadata may not support widespread

[13] https://www.rd-alliance.org/blogs/summary-long-tail-research-data-interest-group-session.html

discovery via search engines or directories. The Interest Group discussed the tension between content recruitment, whereby the aim is to make the deposit process as easy and quick as possible versus the need for data documentation and metadata if the datasets are to be found and re-used. The group discussed some strategies to improve the discovery of datasets, which included assigning a DOI, connecting the data to the journal article, adding greater descriptive information about the data and attaching data management plans to datasets.

As a final discussion point the group was asked to provide suggestions for specific areas that could be pursued through the long tail IG. One concern expressed was that further research is needed in order to have a more spherical understanding about certain issues; what are the tools, support and environments needed that will facilitate research engagement and good practice.

Other suggestions concerned collecting evidence to incentivise researchers to deposit, creating environments to make it easier for researchers to deposit their data, sharing practices about discovery, and ways to achieve interoperability across repositories, as well as preservation planning. Taking all the suggestions into consideration, in the immediate future, the group will start building on some of the ideas, which resulted from this session.

Summary of BoF Geospatial Information meeting at RDA 3rd Plenary meeting

Written by Suchith Anand, University of Nottingham, United Kingdom[14]

Around 30 participants (mainly from Europe, North America) participated in the BoF session on Geospatial Information on 26th March 2014 at RDF3 . The BoF Session was chaired by Simon Cox (CSIRO). The aim of the meeting was to bring together those who were interested in this theme to discuss ideas for moving forward. All participants briefly introduced themselves and their interest in the theme.

The Geospatial Interest Group proposal was presented by Suchith Anand (University of Nottingham) at the BoF. Geospatial Interest Group is a domain oriented interest group to work on all issues related to geospatial data and we will work with the wider geospatial community to finetune the proposal (including those who were not able to attend RDA3). Presentations from Simon Cox (CSIRO), Phil Archer (W3C) and Andrex Perego (European Commission) at the meeting helped give the bigger picture of various activities happening in the geo domain.

[14] https://www.rd-alliance.org/blogs/bof-geospatial-information-meeting-rda3.html

Though there has been lot of developments and advances in geo data collection through mobile communications, sensor platforms, spatial search, and pervasive computing but still, the inter-disciplinary research needed to transform raw data into useful intelligence, to improve the planet's environmental, economic and societal well being, remains constrained by a range of barriers - disciplinary, organisational, historical, and a non-existent or non-rigorous approach to quantifying uncertainty in collected datasets. The Geospatial interest group aims to bring together all major stakeholders producing, managing, aggregating, sharing and consuming data for geospatial research and innovation for building synergies and accelerating future developments. Efforts will be made to get an active participation of the major international institutions . These include the various commissions of the ICA, OSGeo, OGC , the joint W3C/OGC WG, a number of W3C community groups (Cartography, Geometry API, Geospatial Semantic Web, Places, SVG mapping) working in the geo field, which all work on different aspects of geospatial research and innovation. It will take stock of existing problems and experiences and will pave the way for a number of domain specific working groups to make precise proposals for solutions in specific areas. This interest group will help to promote good practices in our research domain : data sharing policies, data management plan, data interoperability, quantifying uncertainty in datasets , how to make geo data re-usable across domains, and to work towards cross-domain interoperability of location information.

RDA Plenary III. BoF WG Big Data Infrastructure session

Written by Rebecca Reffell, University of Essex, United Kindgom[15]

By the time the session on Big Data Infrastructure (BDI) began at 15:30, every seat in the room was filled. As time went on, more and more delegates joined the session to hear Wo Chang and learn about the working group's activities since the last plenary.

Wo explained that the working group started in the US and are now interested to hear the European view on the subject. He therefore welcomed questions and suggestions throughout.

Firstly, the reasons for creating the working group were explained. It is important for those implementing big data applications to have a simple, effective and cheap infrastructure. The platform should enable breakthroughs and allow for changes in technology. Users don't want to deal with the technical parts of big data analysis, a user friendly application enables them to concentrate on the analytics. However, there are many problems associated with creating such a platform.

[15] https://www.rd-alliance.org/blogs/rda-plenary-iii-bof-wg-big-data-infrastructure-session-26032014.html

One main issue raised was that there is no 'one size fits all' solution when it comes to big data infrastructure. There are many federated platforms catering for different needs within different disciplines which can cause problems when scientists do not note which tools or operating system they have used. Computations cannot always be verified and calculations cannot be reproduced when so many different platforms are being used. Therefore, the main dilemma for BoF WG BDI is how to create a standardised, generic platform for data scientists.

A second concern was the cost of storing, moving and analysing data. Often, data is acquired far more rapidly than it can be processed and a lot of time is spent on cleaning up data. The costs involved in this can be very high. Everyone wants a platform that is cheap, fast, effective and trustworthy meaning that BoF WG BDI should create an infrastructure that will keep costs to a minimum.

Wo explained that the group had collected 51 use cases (which can be found at http://bigdatawg.nist.gov/usecases.php). It was argued, however, that the number of use cases collected does not make a difference, it may be more useful to proceed with a single use case at first so that a specific problem can be identified and addressed.

It was suggested that BoF WG BDI should collaborate with IG Big Data Analytics (BDA) allowing them to share use cases and work together on solving these problems. On Day Three of the plenary, the two groups got together and had a joint discussion on how they may be able to help each other. It was decided that people from different disciplines with similar problems should be brought together because from a big data infrastructure point of view 'a data set is a data set' no matter what the data relates to. BoF WG BDI plan to provide platforms for IG BDA allowing them to run algorithms and IG BDA will provide the analytics. The two groups will exchange technical details and come to a conclusion between them. The end goal is to create five to six unique applications relating to different use cases.

So, the next step for BoF WG BDI is to concentrate on a small number of use cases. They will work together with IG BDA to solve problems in analytics and attempt to capture unique applications and identify any patterns or interactions between different domain specific algorithms. The problem of how big data infrastructure can fit for everyone is undefined but the working group will continue to work towards their goal of establishing best practice implementation guidelines for how to deploy and manage big data applications.

IG Legal Interoperability - meeting report

Written by Pawel Kamocki, IDS Mannheim, Germany[16]

The RDA's Interest Group on Legal Interoperability met for three sessions on Thursday, March 27, during the RDA's 3rd Plenary in Dublin. All the sessions - attended by over 30 people - were chaired by Paul Uhlir from the National Academy of Science in Washington and Yours Truly had the honor of being the meeting's rapporteur. The goal of the meeting - to bring together lawyers, managers and researchers from different fields - was fulfilled.
The Interest Group's aim is to include several case studies in its final report. Initially the number of such case studies was settled at four, but due to big interest expressed by different members of the group, it was finally extended. The Dublin meeting was devoted to presentations of the projects to be included as case studies for the final report. Each presentation was meant to contain information concerning five areas (determined by the group's co-chairs: Bob Chen, Enrique Alonso Garcia and Paul Uhlir):

The legal frameworks and specific policies (or lack of them) governing different types of research data important to a specific scientific domain or problem area important to each case study.

The perceived barriers to data sharing or interoperability and perceived needs for increased interoperability that spurred interest and investment in new legal interoperability approaches.

A description of any effective legal interoperability processes, techniques and institutions that have been developed or adopted to overcome the barriers that have been identified.

The stakeholders involved in developing, testing, and implementing legal interoperability approaches and their roles, level of engagement and investment, and impact.

Progress to date in implementing legal interoperability approaches, including identification of criteria or metrics used to assess success or impact, use of technology or other mechanisms to promote adoption, and estimates of funding and other resources provided to support implementation.

Finally, a total number of 8 projects was presented during the Dublin meeting:
#1 CLARIN ERIC (language resources and technology)
#2 Plazi (text mining of data for taxonomy)
#3 CReATIVE-B (biodiversity data)
#4 iMarine (biophysical ocean data)
#5 All-Island Research Observatory (spatial data for evidence and planning in Ireland)

[16] https://www.rd-alliance.org/blogs/ig-legal-interoperability-meeting-report.html#overlay-context=blog%3Fpage%3D1

#6 Digital Repository of Ireland (interactive repository for contemporary and historical data in the humanities and social sciences)
#7 Cybercartographic Atlases (the legal and policy aspects of digital cartography)
#8 The Polar Research Commons (data from the International Polar Year)
The presentations revealed that legal issues concerning data in different domains are in fact similar; it also confirmed that some legal systems (e.g. the Swiss one) are more favorable to the creation of data infrastructures than others due to more robust copyright exceptions and less strict (or nonexistent) personal data protection laws.
During the final debate, a possibility of adding another questions to the case studies was discussed, such as the level of legal information available in the community or possible public-private partnerships. The case studies will be presented and further discussed during the Amsterdam plenary in September.

After publishing the report, the next aim of the Group is to be transformed into a Working Group which would then be able to adopt best practice guidelines for dealing with legal interoperability issues.

RDA Plenary 3: Big Data Analytics Interest Group

Written by ParinazAmeri, KIT (Karlsruhe Institute of Technology), Germany[17]

Chairs: Morris Riedel, Rahul Ramachandran, Peter Baumann
The Big Data Analytics IG took place at the second day of the third RDA Plenary on Thursday 27th of March 2014 in Croke Park conference centre.
The session started as the chairs around the room's big round table were all full and there were more people coming to the room and made the second and third rows of the audiences. Regarding the importance of Big Data analytics and since this is a hot topic, it seems natural to see a lot of people being interested in this session.
Morris Riedel, co-chair of the session, started with introducing the goals of the session for this time as:

1. checking to see if any of the participants has a good use case for the IG
2. checking if there is a possibility for collaboration and paper publication

In addition, the agenda of the session has been reviewed. But before going any further with the agenda, all of the participants have been asked to introduce themselves for more introductions and say what their personal interest in analytics is. Some of the participant's interests were as following:

[17] https://www.rd-alliance.org/blogs/rda-plenary-3-big-data-analytics-interest-group.html

- use analytics in real cases
- multimedia analysis
- finding biases
- IT architecture
- dirty data, and corrupted data
- problem and performance of data analysis
- efficient big data infrastructure

Morris has then explained about discussions in previous plenary meetings and which material they have already provided in the wiki for group members to read. Some highlights from the previous talks and wiki materials are:

- There is a need to have smart analytics.
- Some terms are around for long time, so what is the difference today concerning big data?
- Differences of analytics & analysis
- Data analysis supports the search for 'causality'
- Big data analytics is focused on 'correlation'
- Big Data analytic (clustering, classification, …) is what scientific computing and big data have in common

There were two use-cases chosen to show this difference:

- First use-case: Event tracking analytics: data sets from satellites(events with changing geolocations)

- Second use-case: Automatic outlier detection in big data (PANGAEA), open for one month in B2SHARE

Then, the first speaker, Guiseppe, presented three use cases from solid earth analytics (seismic analytics) and pointed out the characteristic of the data set and difficulties of analyzing the data for each case. Cases were differ from near real-time analysis of continuous streams of data to check for events like earthquake, to offline analysis on gathered data for pattern recognition and "synthetic" data for event predictions, in which very large data are simulated.

Next speaker was Stephan Decker who gave a talk on the Insight Centre for Data Analytics. He presented their experiences of working with industry and some of the works that are done in Ireland regarding Big Data Analytics.

The third speaker, Peter Baumann, had presented some use cases regarding multi-dimensional arrays and stated that different communities have data with different dimensions. As an example, in climate data modeling, there exists cube data since satellite data are dense. There were also some of the databases mentioned, which has already

implemented the multidimensional arrays, like SciDB, Mone, PostGIS Raster Oracle, and array model on top of Hadoop.

As the last planned speaker, Wo Chang, talked about how to capture a workflow? He mentioned that they want to identify different use-cases to study, generalize and ease the way people learn from data sets to from an infrastructure.

Afterward, Phil Archer, a volunteer speaker, from W3C talked shortly to present what is his purpose of being there, and what W3C can offer to help for this IG.

At the end the, Idea of having a by invitation hands-on workshop in RDA US workshop was presented, and concluded that it might be better to have this session in Amsterdam.

I think the session could successfully absorb and meet the needs of its targeted audiences. At the end of the session, Morris asked if the people found the session interesting and want to follow up its activities? In return, he got a lot of positive feedbacks and more than 10 new people wanted to subscribe to the group.

Notes on Marine Data Harmonisation IG at Plenary 3

Written by Erdal Tokat, Institute of Marine Science and Technology, Turkey[18]

Engagement with other Interest Groups (IG) and with other Working Groups (WG) is needed.

WGs create deliverables and Marine Data Harmonization IG can contribute other WGs deliverables.

Prioritization of relevant IG or WG: High, medium and low priority

High priority IG or WG:

Metadata Standards Directory WG

Data Description Registry Interoperability WG

Data Citation WG

[18] https://www.rd-alliance.org/blogs/notes-marine-data-harmonisation-ig-plenary-3.html

Brokering IG

Preservation e-Infrastructure IG

A deluge of data in Dublin

ISGTW FEATURE | BY ANDREW PURCELL[19]

497 people attended the event in Dublin, Ireland. Image courtesy Johnny Bambury.

The data deluge is coming. In fact, in many research disciplines it is already upon us. The era of 'big data' poses enormous challenges to researchers across almost all fields of endeavor, from natural scientists to humanities researchers and from citizens to policy makers. However, big data also presents a wealth of opportunities, especially in today's global, interconnected world.

With scientific data output alone growing at a staggering 30% per year, it is vital that researchers come together to build the social and technical bridges required to enable open sharing of data. The organization charged with achieving this is the Research Data Alliance (RDA), which is supported by funding bodies from Australia, Europe, and the US. Having only recently celebrated its first anniversary, the RDA has already grown to include over 1,500 members. "The growth has been precipitous," says Francine Berman, co-chair of the RDA Council. "Our community is expanding in both scope and numbers and our organization is evolving."

[19] https://www.rd-alliance.org/isgtw-feature-deluge-data-dublin.html

“I think we’re on a very exiting cusp of a change in how research is done,” says John Wood, Berman’s fellow co-chair. “It’s not just about data, but the democratization of science.” Wood explains that the RDA’s vision is to enable researchers from across the globe to openly share data across technologies and disciplines, so as to tackle the grand challenges of the 21st century, such as disease, malnutrition, and climate change.

Berman and Wood were speaking at the third plenary meeting of the RDA, which was held in Dublin, Ireland, last month. The 497 attendees at the event engaged in discussion on a wide variety of related topics, ranging from the role of publishers and persistent identifiers to heritage data and legal interoperability. Data applications discussed also included geospatial information, marine observation, food production, and urban quality of life indicators.

“Research practices have changed substantially over the last five-to-ten years,” says Australia’s chief scientist, Ian Chubb, who spoke during the opening plenary session of the event. “Today, things are far more global.” During his speech, he emphasized the role that data sharing has to play in addressing global public health issues and drew particular attention to a working group within the RDA focusing on the interoperability of data relating to wheat crops. He argues that the growing global population and shifting rainfall patterns due to climate change make this a vital area of work. “Global challenges can only be solved by global research endeavor,” adds Chubb.

Tony Hey, vice president of Microsoft Research, also gave a keynote address on the second day of the event. He cited work by researchers from the Harvard-Smithsonian Center for Astrophysics, Massachusetts, US, that shows that 44% of URL links embedded in papers published by the American Astronomical Society in 2001 were broken just a decade later. However, things are improving, notes Hey: “The emphasis on data is long overdue,” he says. “Data is becoming a first class citizen, it is no longer something that you don’t look after.”

...

Carlos Morais-Pires, who coordinates the area of scientific data e-infrastructures at the European Commission Directorate General for Communications, Networks, Content and Technology, was also a member of the panel. He echoed calls for greater involvement from both research communities and industry in forming data policy. Morais-Pires also highlighted the Open Data Charter, a document signed by world leaders during the UK presidency of the G8 last year. In this document, five principles for open data are established: open data by default, quality and quantity, useable by all, releasing data for improved governance, and releasing data for innovation.

...

it is the research communities, those working at the coal face of the big data challenge, who will bring about the sea change necessary in how data is handled. The modern-day agora that is the RDA, with its community-led focus and fuelled by passionate argument, is

undoubtedly the organization best placed to tackle these challenges posed by the data deluge. Through this work, it may be ensured that we, as a society, are able to equip ourselves with the tools of the big data era in our fight against the grand challenges of our century.

When it comes to scientific data, sharing is caring

Feature written by Marie Boran, published on Mon, Mar 31, 2014, THE IRISH TIMES[20]

Data sharing is crucial to scientific progress, and can ensure that valuable research isn't lost

There is no doubt that funding scientific research is costly. The Hubble Space Telescope cost US taxpayers a staggering $1.5 billion to build, making it one of the most expensive pieces of scientific equipment ever made. Of course, the fact that Hubble has added to our knowledge of the universe is payment enough; since its launch in 1990 Hubble has helped scientists calculate the age of our universe, advanced our understanding of black holes, and sent us breathtaking pictures of the birth of distant stars.

According to Dr Ross Wilkinson, executive director of the Australian National Data Service, Hubble is also one of the best examples of a healthy return on investment in scientific research. his is because raw data from the telescope has been available for the past 20 years to whoever wants it, resulting in an archive that generates $1 million (€728,000) per annum in revenue. Having an open data policy has, in other words, doubled Hubble's return on investment. This example from Wilkinson perhaps best illustrates the reason why 497 academics and policymakers from around the globe gathered in Dublin last week for the Research Data Alliance's Third Plenary Meeting. Rather than being a niche gathering, it was ***a place for people who "care about how the sharing of research data can progress to discoveries that have the potential to be of benefit to all,"*** said Dr Ruth Adler, the Australian ambassador to Ireland. ***Research data is costly enough and difficult enough to generate in the first place; having it sitting forgotten on a hard drive somewhere, never to be shared, is not only careless but also not in the spirit of scientific endeavour***.

...

---

[20] https://www.rd-alliance.org/irish-times-when-it-comes-scientific-data-sharing-caring.html