



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



RDA 25th Plenary Meeting [part of IDW 2025]

Collaborative Notes

Group(s) name(s) / session title	BoF, Brian Ballsun-Stanton - <i>Documenting LLM Interactions throughout the Research Data lifecycle</i>
Session link for more information	https://www.rd-alliance.org/members/denubis/plenary-participation
Session programme time	Monday, 13 October 06:30 – 08:00 UTC or 16:30-18:00 AEST, Breakout 2
Presentation slides	https://docs.google.com/presentation/d/18f2mMrN6lsCiQFqiP_gKpQPjWK7GI-2B/edit?slide=id.p3#slide=id.p3



Session summary (to be filled by group chairs):

We will use the content in the table below to highlight your work to the RDA community in a report by the Technical Advisory Board ([RDA P24 report example](#)). **Please complete ALL fields below by Friday, 31 October, to be included in the report** and other communication. We also ask you to update your group webpage with the key takeaways from your post-plenary session.

1. Describe what your session is about in three sentences maximum:	- As researchers integrate LLMs into their workflows through prompting, fine-tuning, and other interactions, there are no established standards for documenting this work to ensure transparency and reproducibility. - Participants broke into groups to discuss challenges around standards, publisher requirements, ethics, and technical implementation. - Discussions covered both researcher-facing documentation practices and technical infrastructure needs around engaging with and producing methodological process documentation.
2. Highlight a maximum of five key outcomes/actions/takeaways of your session:	- We identified people who wanted to meet during the conference and establish the terms of reference for an interest group. - Established initial frameworks including the three P's (Prompt, Process, Provenance) and treating LLM interactions as methods documentation - Surfaced cultural differences across disciplines about what constitutes legitimate AI use and requires documentation - Identified tension between comprehensive documentation needs and researcher burden when AI use is iterative (hundreds/thousands of prompts per paper)



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



	5. Recognized that AI is now embedded in many systems (spell-checkers to research tools), making it challenging to define documentation thresholds across disciplines
3.	What collaborations or synergies did you discover between your group and other RDA group(s) or external (outside of the RDA) organisations? RDA Groups: • Global Open Research Commons IG • Artificial Intelligence and Data Visitation WG • Research Data Policy IG External Organizations: • Cambridge University Press • Australian Research Data Commons (ARDC) • CSIRO
	Disclosure/logs https://claude.ai/share/2ac28754-1215-4677-94b2-05a5bd90df59 to produce the above notes, and Gemini of the live recorded session transcribed and diarised the session. The system prompt to Gemini 2.5 pro was: "Please transcribe the presentation of the primary speaker, take care to be accurate, and make sure that pauses between slides and paragraphs are respected." Data training opt-out was engaged as per Google's and Anthropic's terms of service.



Attendee Check-in

New to the RDA? Check out on where to start [here](#)

Get involved in the [RDA Community](#)

This meeting will take place in accordance with the [RDA Code of Conduct](#)

Please complete this table to indicate your attendance (add rows as needed):

Full Name	Affiliation	Location	Email
Joonas Kesäniemi	CSC	Helsinki	joonas.kesaniemi@csc.fi
steven de costa	Link Digital/CKAN	Canberra	chairman@linkdigital.com.au
Olga Churakova	University of Bern	Switzerland	olga.churakova@unibe.ch
Amy Nurnberger	Massachusetts	USA	nurnberg@mit.edu



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



	Institute of Technology		
Chris Seal	University of Auckland	Aotearoa, New Zealand	c.seal@auckland.ac.nz
Christian Pagé	U Toulouse CERFACS/CECI	Toulouse, France	christian.page@cerfacs.fr
Kiera McNeice	Cambridge University Press	UK	kiera.mcneice@cambridge.org
Vanessa McBride	International Science Council	France	Vanessa.mcbride@council.science
Louis Mapatagane	Walter Sisulu University	South Africa	lmapatagane@wsu.ac.za
Jens Klump	CSIRO	Australia	jens.klump@csiro.au
Andrew Treloar	ARDC	Australia	andrew.treloar@ardc.edu.au
Charlie Lembit	Murray-Darling Basin Authority	Australia	charlie.lembit@mdba.gov.au
Rebecca Cowley	CSIRO	Australia	rebecca.cowley@csiro.au
Katie Mills	University of Western Australia	Perth, Australia	katie.mills@uwa.edu.au
Mikiko Tanifuji	National Institute of Informatics, NII	Tokyo, Japan	tanifuji@nii.ac.jp
Christina Drummond	UNT / OAEBUDT	USA	Christina.Drummond@oabookusage.org
Thomas Jejkal	Karlsruher Institute of Technology	Germany	thomas.jejkal@kit.edu

Collaborative Session Notes *(To be used by participants and chairs during the session)*

Slides link?



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



https://docs.google.com/presentation/d/18f2mMrN6lsCiQFqiP_qKpQPjWK7Gl-2B/edit?slide=id.p3#slide=id.p3

Background:

This Birds of a Feather session aims to establish frameworks for documenting, sharing, and standardizing Large Language Model (LLM) interactions within the research data lifecycle. As LLMs become integral to research processes across disciplines, the prompts and interactions themselves constitute valuable data artifacts that warrant standardization and preservation. Our objectives are to:

1. Frame LLM prompting part of documentation of research methodologies, thereby requiring documentation standards comparable to code, protocols, and other methodology components, enabling transparency and reproducibility
2. Address the current lack of norms, formats, and tools for documenting AI use in research by establishing foundational principles for an “LLM interaction standard” that aligns with FAIR data principles
3. Explore methods for using AI to create, improve, and annotate existing datasets in reliable and reproducible ways.
4. Develop preliminary specifications for formats and tools (such as bookmarklets or browser extensions) that could capture LLM interactions as structured, shareable research objects
5. Explore opportunities to extend existing frameworks (like RO-Crate) to incorporate LLM interaction documentation, and identify potential repository solutions for a community-maintained “grimoire” of research prompts
6. Initiate cross-disciplinary discussions that will form the basis for potential working groups focused on technical standards, ethical norms, and disciplinary best practices for LLM use throughout the research data lifecycle

This BoF distinguishes itself from existing RDA groups by specifically addressing the emerging need to treat researcher-LLM interactions as data artifacts worthy of standardization, documentation, and sharing within the open science ecosystem. We would be delighted to merge with any other BoFs on related AI topics.

Goal: Instigate an interest group on disclosing AI use in research



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



Idea: we should treat prompts/LLM logs like we treat code or data (i.e. keep records of prompts, tools, etc. so that these can be assessed if needed as part of peer review, ethics investigations, etc.)

- Standards for disclosure are inconsistent
 - Disclosure alone is insufficient. It is not documentation or reproducibility or transparency
 - This is heading towards a Kuhn-ian paradigm shift
 - Models are “vibe consistent”
 - Just like data LLM interactions should be published and include: Standards, Formats, Tools, Norms
- Prompting as methodology documentation: how do we show our work
 - How do we enable our peers to judge our judgement (about use of LLMs)?
 - Question: Reproducibility has the assumption that knowledge is constant/consistent. LLMs are not.
 - Response: an LLM operates in read only mode.
 - Question: Same prompt results in different answers. How is this reproducible?
 - Words will be different, but the “vibe” will overwhelmingly be consistent. Ethan Mollick 2023, [“now is the time for grimoires.”](#)
 - Tacit knowledge is made explicit when rendered in a prompt
 - (perfect) Reproducibility is not a new challenge in science
 - Perhaps we should be using ‘transparency’ as a term rather than reproducibility
 - Question: Please articulate what you’re trying to identify
 - Each chat interaction engaged in as part of research methodology should be part of transparency
 - How do you document AI agents
 - Excellent question. There should be standards
 - Comment by Andrew Treloar: LLM versions are not backwards available. What good is it to know that I used prompt “xxxxxx” on LLM version yy, if I can’t get back to that LLM version?
 - Response: there are ways to interact with older versions through API, if someone is sufficiently determined. The idea of assessing/judging & re-iterating is more interesting than repeatability
 - Andrew comments in doc: Yes, and this is somewhat how we think about research software. I might not re-use it directly, but it has value as a way of encoding knowledge for me to learn from and build on. This idea of research software as knowledge encoding is something I



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



learned from Tom Honeyman when he was writing a [National Research Software Agenda for Australia](#).

- Some question (refer to recording)
 - Response
- The cyclical process: Research workflow with documented LLM use. You benefit most from your own documentation
 - We are developing “taste” with respect to AI outputs - subjective judgements on good/effective use of AI and whether it gives useful results
 - As with data, *you* benefit most from your own documentation!
- There is a fragmented landscape
 - Many researchers don’t like writing papers. Fine if you have “good” prompts & “good” models (along w shared infrastructure, like we have for code)
- Research grimoires
 - Tested prompts, documented workflows, reusable patterns, tooling for extractions of interactions, community knowledge
 -

Section 2: Current challenges

1. Missing standards
 - a. Prompt documentation format
 - b. Response/interaction capture
 - c. Metadata requirements
 - d. Version control
 - e. Attribution
 - f. Enter others at will
2. Publisher requirements: What should journals actually require?
3. Technical barriers: capturing interactions is hard
 - a. Volume,
 - b. privacy p
 - c. Proprietary platforms,
 - d. integration
 - e. Etc.
4. Ethical considerations
 - a. Sensitive data in prompts
 - b. PII: Personally Identifiable Information
 - c. Proprietary information
 - d. Failed attempts vs successful ones
 - i. Quite a lot of AI use doesn’t lead to useful results - do we need to document every failed attempt?
 - e. Context & interpretation



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



Section 3: Answer the following about group chosen topic

- What elements of LLM interactions need standardization?
- Which existing standards might be extended? (FAIR, RO-Crate, Software Citations?)
- What tools would facilitate documentation?
- There are prompts provided: (Use Claude if you can)

You are facilitating a research discussion on [TOPIC].

Your role is to ask ONE specific and critical question at a time to help the group develop concrete recommendations.

Start by asking: What is the most important immediate challenge this group should address regarding [TOPIC]?

After each response, synthesize what you heard and ask the next most critical question. Continue until the group has actionable recommendations.

Missing standards & metadata conversation

Metadata:

Chat log

Motivation for research topic; motivation for choice of using AI; motivation for model used?

Model used; when used (timestamp); who?; how to access the model (download/api/ui)?

data used for training (maybe impossible?)

System prompt vs. User prompt

API use

A tool to extract the metadata around interactions would be useful (TypingMind may do something like this?)

Access type (webui / api)

Metadata related to the “package” conversation / conversation set

Standards

- Cf. Tom Honeyman. Even if software isn't perfect reproducibility, it is useful information
- Also similar to documenting mental workflow
 - Needs methods for capture



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



- What should standards cover?
 - Version question should cover both model and inputs
 - Provenance
 - What about a model this is further specifically trained on local data
- Need to know model, date, infrastructure, version, further training data, model temperature settings. Not capturing for perfect reproduction. Capturing for the sense of process that you followed. For assessment, for transparent
- Think about use of LLM as methods and how they are captured. Minimum information needed for methods section (MIAMI)
 - Think about establishing the Dublin Core as a starting point that disciplines can take further
 - The minimums in different disciplines may be interesting
- Can you make the model spit out a record of the interaction in some agreed upon format/structure
- What could be extended
 - FAIR for research data software
 - CODEmeta
 - A workflow tool to map out the paths
 - E.g., workflow hub
 - In a workflow language (there are so many)
 - protocols.io?
- Should use Methods section as a paradigm

Ethics

- Governance of AI and protocols
- What use of LLM is allowed and appropriate?
- Needs to be specified to context
- Running the LLM localised. Potentially version controlled by the institute
- Having trust indicators with levels
- There could be sensitivities in prompts which could potentially not be disclosed publicly e.g. Indigenous cultural knowledge, health/related data, personal data
- How to know it is trustworthy: **Prompt - Process - Provenance**
- Recording the LLM use in catalogues and metadata as a determiner role
- obtaining ethics approvals to put sensitive data into LLM and be used as prompts. Would require plain language and clear process documentation
 - (Brian note: Microsoft presidio library for redaction)
- Disclose the process but not the actual sensitive data
- Good LLM can be expensive, potentially prohibitive
- Open and commercial AI - something in between so there is collective ownership



What should publishers require? Discussion notes:

- [STM working group output](#) – agreed *not* to require or even recommend publisher policy because community view vary so much
 - Differing opinions even within disciplines – one editor-in-chief said they'd immediately reject an article if the author acknowledged using Grammarly (!)
- Several stages of use – what are people using AI *for*? Some uses are benign, some are contentious and need more careful documentation
- Should be some limits in what needs to be documented?
 - If you work through thousands of prompts when refining your article, do you really need to share all of these?
- If the purpose is to justify the quality of the science, it makes sense to try to document LLM use – but it may not be practical for scientists to keep records without a machine-actionable data management system (with metadata, AI version, etc.)
 - Important to know who used what information – citation of results of finding is still really important. If ten authors using several AIs with hundreds of prompts, whose contribution is the most important in the findings?
Challenging in current system
 - Journals not capable of doing this – publishers rely on authors to disclose who does what
- Depends how you use the AI – do you use it just like a GPS and blindly follow what it says? Or use it as a helper, then use that information to help you think about what you want to do?
 - Nobody takes the response to a prompt and pastes it into a document exactly as it is! It is an iterative process
 - On the other hand, “publish or perish” is a perverse incentive to pump out as many articles as possible, using AI as shortcuts – journals need to be more discerning. But APCs as a business model don't exactly encourage this...
- Recent article about use of AI in weapons systems – which sounds terrible, but the point is that AI is a tool, we have to own the responsibility for using AI in our decision-making processes
 - There's fair agreement on the fact that AI can't be listed as a journal author for exactly this reason – but much more varied opinions on using AI during the research/writing process
- Maybe there *is* a cultural difference – in science it's all about the data, but humanities may be different. Germany now has a “sport” of taking down politicians based on their PhD theses...
 - What's acceptable differs a lot between disciplines – fields where writing is seen as a fundamental part of the researcher's contribution are a lot less favourable to using AI to generate text
 - Are these fields aware that even Microsoft Office's in-built spell checking is an “AI service”?
- If an LLM generates a piece of text that is objectively better than your original writing, why wouldn't you use it? You are ultimately still responsible for what you submit to journals.



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



- AI can genuinely do good science, in some cases – e.g. solving specific physics problems.
- Would be good to establish standards for metadata to describe LLMs – which model, used at which time, etc.
- People are not honest in their reporting for many reasons. There also needs to be a culture change in research to accept that AI use isn't a badge of shame.
 - Kiera note in doc: [This recent article](#) interested me in the way it frames *disclosure* as eroding trust - surely it's the AI use itself that people think is less trustworthy? How can we shift the conversation? (See notes below on culture change too...)
- AI is supposed to save us time... now we're going to spend all our time on extra documentation? Nobody will want to do that!
 - But **reporting standards are part of good research – and should be here too.**
 - **Need to think of AI as a tool/implement, not a “collaborator”.**
- **Journals can work on the messaging problem** – call AI an instrument, and refer to experimenting with it as a tool in research
 - Definitely try to keep away from things like “I asked ChatGPT to...”
- How many authors are acknowledging it? Because 100% of authors are using it!
 - Difficult to tell – **no standards for how/where it's declared in articles**
- How will genuine science survive if workflows become more automated, just generating huge amounts of data?
- Should AI use for code be declared separately to AI use for text generation?
 - [STM classification](#) defines these as separate uses – but it's up to publishers to decide what needs to be declared, what uses are permitted, etc.
 - Some learned societies and research communities have VERY strong opinions about AI use – and they own the journals
- FAIR² is already essentially generating articles wholesale from data
 - (Brian comment, I'll show you [data-to-paper](#). It's vile)
 - We talked about this sort of things 15 years ago in the context of overlay journals – everyone got very excited about metadata and using it for this sort of purpose
- Scopus and other databases are all creating their own LLMs - and not permitting people to use the full text of their own content to be used by other LLMs
 - Same problem as with text and data mining - very fragmented, have to go to multiple sources and do rights clearance to gather a corpus for analysis
 - Google Gemini doesn't allow you to query non-public data in Google Scholar, which limits its usage for e.g. systematic reviews - could Google help facilitate this sort of licensing/access issue?
- What about use of AI/LLMs in peer review?
 - (Brian comment, look at the black spatula project)
 - Uploading a manuscript to a third-party LLM tool is a breach of peer review confidentiality policies
 - But if you have a totally enclosed environment, this would be okay to use...?



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



Interest in the Interest Group

Drop your email here if you want to signal to the RDA that this is a viable interest group. (We will contact you)

- brian.ballsun-stanton@mq.edu.au
- shawn@fieldnote.au
- kiera.mcneice@cambridge.org
- Imapatagane@wsu.ac.za
- nurnberg@mit.edu
- c.seal@auckland.ac.nz
- Christina.Drummond@unt.edu
- jens.klump@csiro.au
- liz.stokes@ardc.edu.au

Live topics for the interest group:

- (Some text highlighted in the publisher policy & standards notes for this)

Gemini 2.5 Pro Transcripts

System Prompt: "Please transcribe and diarise this conference presentation and discussion. Slides and notes are provided for context and language."

Meeting 1 - BoF session

****Transcribed and diarised conference presentation and discussion****

****Session Title:**** Documenting LLM Interactions throughout the Research Data lifecycle
(Birds of a Feather)

****Presenter:**** Brian Ballsun-Stanton (Brian)

****Conference:**** RDA 25th Plenary Meeting

****Date:**** 13 October 2025

****Brian**** (00:02): So, before folks get too comfortable, there will be elements, I'm so sorry, of group work.

****Brian**** (00:11): Yes. So maybe don't sit in the nether realms where you cannot collaborate with anyone. Um,



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



****Brian**** (00:22): or move, you know, during the session. Both are valid.

****Brian**** (00:53): And because this topic has come up once already, we will not be covering the use of AI in teaching or grading during the session. That is what alcohol after the session is for. I have been bouncing around Europe lecturing on that for the last 110 days. So I have opinions, but that's for after, and I will just be stomping on anyone who brings up the teaching side of things, 'cause if we start down that path, forever it shall dominate our destiny. Um,

****Brian**** (01:27): so, just letting that be out there.

****Brian**** (01:40): So my current goal is to join the online session on my laptop so that I can see chat. Um,

****Brian**** (01:53): which is a series of choices.

****Brian**** (02:04): And then I get to see me, all questions. An hour ago. Fascinating.

****Brian**** (02:19): Oh, this is for the Chat break for the LLM session because, of course, I cannot clear chat.

****Brian**** (02:37): Um, okay, so let's just there we are.

****Brian**** (02:45): Hey folks, if you will give me a few moments so that I can put the appropriate links to the shared doc in chat. Um, while I get that set up, I will ask you all to join the shared notes document because this is going to be interactive in the sense that we are going to be generating that document.

****Brian**** (03:08): Um,

****Brian**** (03:16): excellent.

****Brian**** (03:18): Okay. No, too many attempts. Retry after 60 seconds.

****Brian**** (03:24): Sure.

****Brian**** (03:28): Welcome. I'll I'll leave this up for a few moments more. Unfortunately, I cannot show both the rough agenda and this at the same time as the slides. So we'll have a few moments.

****Brian**** (03:45): In a rough sense, what my goal is today is to establish the basis of a Birds of a Feather but the basis of an interest group. I think that there need to be standards for how we engage with and publish our interactions with large language models. I'm just going to say AI. I know AI is 70 years old. I've stopped trying to fight that fight, so I'm just going to say AI, and I apologize for my sin.



****Brian**** (04:24): Um, our interactions with AI in terms of our prompts and in terms of our engagements, in terms of our outputs and in terms of our models, because quite frankly, I think that the disclosure, here's what I used is utterly pointless. And that here at RDA, we we are in a position to start having conversations about creating those standards and about creating those expectations. So, does everyone who is interested have a link to the shared notes document? If not, wave vigorously. Also, because this is hybrid, first off, anyone who is hybrid, please give me a thumbs up or say hi in chat just so that I can see that we have folks online. Um, but because we're hybrid, any time that you're speaking to the room, I will unfortunately have to pass you the microphone. Um, so let's just keep that in mind.

****Brian**** (05:25): Okay. It doesn't look like anyone else is loading the shared notes document, so let us proceed.

****Brian**** (05:34): The plan for today is that I will give roughly a 20-minute introduction to where my thinking is around these, why is this a problem? And then we will break into groups and have discussions about the specific challenges thereby.

****Brian**** (06:00): The best Birds of a Feather are where everyone gets to actually talk and think. This is not a lecture, this is not a keynote, this is not a presentation. This is a conversation. In that vein, if I say something strange or you have a question during the first part, ask me, because I guarantee you that other people have that thought.

****Brian**** (06:25): I built this uh, Birds of a Feather proposal, this working towards an interest group with my colleague Shawn Ross, who is unfortunately not able to make it today. Um, so we just have my name on things because that's what he asked me to do.

****Brian**** (06:42): Oh, and these slides are creative commons, so please enjoy. Um, the fact that the license isn't clearly showing is deeply frustrating to me, but I used the templates. It's creative commons license. Use them, have fun. I have links to my prompts at the end so you can see how I worked with AI to help think through them because it's really pointless if I don't follow my own norms.

****Brian**** (07:08): I would like to begin with an acknowledgement of country. Uh, we acknowledge and celebrate the First Australians on whose traditional lands we meet, and we pay our respect to their elders past and present.

****Brian**** (07:25): The guiding principles here are really, really important because it is quite likely that in this session we will disagree.

****Brian**** (07:37): My objective here is for us to build consensus around work towards an interest group.

****Brian**** (07:47): My objective is not to prescribe how people should use AI. I have deep opinions, but that waits for beer.



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



****Brian**** (07:57): Right? And so in so far following these things, following the code of conduct because we're going to have very different opinions on the quality and nature of AI is deeply important.

****Brian**** (08:15): Any questions so far? Should I swap over to the QR code for the late comers? Or can you, can you no, no, no, no, no, no, join, join, join.

****Brian**** (08:24): Um, so, how do we disclose AI use in research? This comes from my research. Um, so I'm Brian Ballsun-Stanton, Macquarie University. I have a paper out with Queens uh QUT on frameworks for making research policy uh about AI. How do you make a research policy for your university? Here is a very complicated diagram and a bunch of steps that you will ignore.

****Brian**** (08:52): I'm honest, sorry. And for the vast majority of journals and I and I'm going to note this, for the vast majority of journals, they require disclosure statements. So they require scientists, okay, I cannot stand behind that lectern, that, so they require scientists to step up and pin a red badge of, I've used AI on their papers.

****Brian**** (09:22): I have used AI to make this presentation.

****Brian**** (09:28): Okay. The vast majority of journals have that as their basis, that disclosure or acknowledgement is there. Now, happily we've moved away from the benighted times of 2023 where ChatGPT is my co-author. And most journals have pushed against the idea that you should cite personal communication, chat GPT. So most journals are uh require a disclosure. Uh, some medical journals, especially, require discussion in methodology if you use large language models as part of your research process, but very few journals require actually showing your work.

****Brian**** (10:25): My fundamental thesis here is that we should treat our prompts and the output of AI just like we treat our code and just like we treat our data so that people can look at our prompts and judge if that is appropriate. The overwhelming majority will not, but if something goes weird and someone wants to know the provenance of some aspect, those logs should be available. Usually that's going to be as part of an ethics investigation, but still, that's one of the reasons why we publish our data, so that you can figure out if you didn't drag all the way down the column in Excel and if half your measurements are dates.

****Brian**** (11:11): That's my thesis. And so because we have folks, I'm going to swap back to the word document. I'm going to swap back to the word document so that people can join the shared notes document as they wish.

****Brian**** (11:37): Any questions so far while we wait for folks to to join the shared notes? Also, by the way, if people want to be annotating and having discussions in the shared notes document, even in this first part, I will be very happy and very excited. I would be delighted if we actually start building our shared notes while there's things to take notes on.



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



****Brian**** (12:01): Okay, so I'll swap back. Uh, maybe grab this from a neighbor if there's still challenges.

****Brian**** (12:08): So, the requirements are fundamentally inconsistent. Our norms are inconsistent. Where should a large language model be used during the paper writing process? Opinions vary.

****Brian**** (12:35): Quick question. How many philosophers in the room? So if I talk, start talking about Kuhn, am I about to be crucified? Cool. Um, I would actually posit that it's starting to look a lot like a paradigm shift, not in the happy, happy, like technology by our product sense, but in the, we have a bunch of facts that are difficult to explain sense, and the conflict between scientific research paradigms is starting to show. I am not saying which one is better. I am saying that they are different in a way that is noticeable. Our standards for what and how to disclose are fundamentally inconsistent by journal, by discipline, by faculty, by time of day, which means that we as researchers have no expectations on what is allowed, what is licit but we don't talk about, what is illicit but we still do, and what is actually problematic.

****Brian**** (13:42): And I think that we at the RDA have the ability to have to start that conversation because disclosure of use or acknowledgement of use, it's not acknowledgement, it's not a person, you're not thanking them. It's not a citation, they're not an authority, you're not citing them.

****Brian**** (14:02): Disclosure is not documentation, but even more to the point, disclosure isn't reproducibility. I think that we can borrow the ideas of FAIR data and reproducible research for AI. Sure, we can we can call them spicy auto-complete all we want. Sometimes it definitely feels that way, but they are going to be vibe consistent for a given model, which means that there is a fundamental vibe reproducibility there that we can engage with in the same intuitions that we do about data and that we do about code.

****Brian**** (14:48): Therefore, what I would love to gather a group together, an interest group together is for publishing large language model interactions. And I had a lovely conversation just as we were starting with someone, um, about uh, fine-tuning models. And I would say that interactions is can be both valid for prompting, which is my original thesis, and fine-tuning. So any interactions that we as researchers have with models after the point that models are published, we treat like code, with the exact same standards. Note, I am not just saying the outputs. The outputs are pointless. I'm saying our inputs, because it's our inputs that lead to the outputs and it's our inputs upon which we can be judged. Therefore, we have to talk about what our standards are, what the formats are. Because I have a bunch of OSF repos with PDFs that I've produced of my interactions because I follow my own norms, right? But what are our standards? What are our formats? Can we build tools to extract all useful metadata with our interactions from AI?

****Audience Member**** (16:17): (cheering)



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



****Brian**** (16:18): All right. Thanks crowd. Right? Can we do that? And more, more to the point, can we as an RDA interest group, articulate norms, the shoulds, the should nots, the musts, the must nots that help us to actually keep research useful and to a degree trusted, whatever that word means this week?

****Brian**** (16:51): Because I would assert that prompting, actually, before we do that, that was a section break, pause. What questions do you have?

****Brian**** (17:00): Chat has none.

****Christina Drummond**** (17:03): Ah, excellent. I am going.

****Christina Drummond**** (17:06): Do you need me in the mic?

****Brian**** (17:07): Unfortunately, we we are recording this so.

****Christina Drummond**** (17:11): Unfortunately. On the record.

****Brian**** (17:12): Yes. Oh yeah, no, this is this is me on record. It's great. Also, please, sorry, please state your name and institution. I got a grumpy email saying that I had to remind people that.

****Christina Drummond**** (17:22): No worries. Christina Drummond, University of North Texas and Executive Director for the Open Access Book Usage Data Trust. So we're working with a data space and the data space protocol. Um, and I bring that up here because I I want to point out something key between the difference between publishing data, which is a one-time transaction to make it accessible, and what you're proposing here, which is a transaction over time. This is something where if you're interacting, that's not just one point, it's you're now managing those transactions in an ongoing dynamic way, and I I just want to posit for the room that that transforms how we have to think about it. We can't just drop it in a repository and say good enough.

****Brian**** (18:05): I completely agree. Um, right now, the way I do that is I have little folders in Claude where I have a little projects and try to organize things, and then for a given paper, I try to put all of my things for that paper. Obviously, there is much more that we can uh, wish for, and that if we think about our use of large language models as something that is something we do on our way to work and have a conversation about some of the points in a paper, how do we capture that? So yes. There's quite a large space that we can play in. The place where I want to start is how do we show our work? That is, that is my fundamental question because I say that the disclosure statement is insufficient. And I also say that forbidding it doesn't work. We cannot reliably detect good AI-generated text. We can, however, do all of the problems of English as a second language discrimination. We can do neuro-diverse discrimination. We should be absolutely savage at poor judgment when it comes to AI, and again, talk to me in the pub about how that relates to grading. I have opinions. Um, they will not make the budgets of our universities happy. But yes, how do we



show our work? How do we allow our peers to judge our judgment in the use of AI as it relates to our research process?

****Brian**** (19:50): Okay. No other questions on chat. We'll keep going. Oh, yes, wonderful. Thank you.

****Louis Mapatagane**** (19:58): you're ignoring.

****Louis Mapatagane**** (20:02): Um, this is uh Ntsundeni Louis Mapatagane from um Walter Sisulu University in South Africa. I actually agree with uh your thesis, uh first of all. But help me understand this. Now, when we talk reproducibility, right? Um I'm a scientist, so I know I'll do an experiment, I'll document it, and somebody else can run it and get the same results. Then or, yes. Now, this assumes, well in the use of AI and LLMs, this assumes that that information and knowledge is constant.

****Brian**** (20:41): Is our audio coming through or am I just projecting? I'm an academic. Okay.

****Louis Mapatagane**** (20:47): So, in my head, the assumption is that then that knowledge is constant for it to be reproduced. Now, we know that LLMs are not. They the knowledge grows with time, depending on the different interactions that are taking place and context. Now, help me unpack that. Thank you.

****Brian**** (21:11): Great. Um, a large language model operates in read-only mode. A large language model infers the next token purely from the context of your current conversation star. The star is there because OpenAI is doing their best to disguise that fact, but a given model, when it produces output, produces output in a read-only mode. It produces output on the series of interactions that you have had within that chat only. Anything beyond that is a function of it interacting with traditional databases or searching the web in some way. New models are released, which is why having documentation of which model we use and which service we use is essential. But they do not grow or change or learn for a given model release. They just look like they do. Did I answer your question? I'm answered. Great.

****Robert Shen**** (22:22): Robert Shen from RMT University. Can I follow up on this one? Because in term of the same model, you use the same prompt engineering for 10 times, there is a high chance the answer will be slightly different. So how, what's if this will be the case, what is the purpose for reproducibility?

****Brian**** (22:43): The answer will be different in terms of the words, but it will be mostly consistent in terms of the vibe. The sense of the message will be consistent. The punctuation it uses and the way it goes, the overwhelming majority of the time will be consistent. Obviously, if we are us doing research with large language models on a large data set, we set the temperature to zero, and then it is consistent for much greater values of consistency. But reproducibility here is, what did I prompt? What was the output as a series of interactions within that specific context window? And do pe can people sit in judgment of that? Because in terms of code, sure, it is deterministic. In terms of prompts, it is remarkably



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



consistent over time. Um, if you want to take a look, uh, Ethan Mollick, 2023 wrote, "Now is the time for Grimoires," where literally you can share prompts as a way of making knowledge explicit. And so, in many ways, what I'm saying here is how we should document our methodology.

****Kiera McNeice**** (24:11): Um, hi, Kiera McNeice, uh research transparency manager at Cambridge University Press. Um, I liked at the intro how we talked about um, treating LLMs as just like another tool that we need to document the same way we talk document data. I also take the point that they're not 100% consistent. I mean, especially when you look at things like assisted information retrieval, you can get different results. But I mean, this is not a new or unique problem in research. There's a lot of kinds of research where you're never going to get 100% the exact same results. There's a lot of, there's things like observational studies, there's whole areas of humanities and social sciences which hate the phrase reproducibility because their work is never necessarily strictly reproducible. So there are a lot of fields that have looked at this similar problems, I think. And just because we're not going to have perfect reproducibility of the exact results doesn't mean that it's not valuable to document as much as we can, as with any other field where you don't have exact results every time.

****Brian**** (25:00): Thanks.

****Brian**** (25:01): And I would definitely say that there is vibe reproducibility, which is a statement.

****James Wilson**** (25:07): James Wilson, UCL. Yeah, you're quite right here in that in the humanities and stuff, we don't talk about reproducibility, we talk about transparency, typically. Should we be using transparency as the terminology here or is there a reason to use reproducibility and that slight, slightly stricter connotations that it has?

****Brian**** (25:27): That's a really good question.

****Kiera McNeice**** (25:31): Okay. Uh, as I'm flying. Sorry, I'll yeah. Um, I mean I would say what I've found since joining Cambridge University Press, initially my job title was research data manager, and two thirds of our journals are HSS journals, they wouldn't listen to me because they didn't want to talk about data. But talk about transparency, and a lot of those people will start listening and discussing what transparency means in their field as opposed to data or reproducibility.

****Brian**** (25:56): The only reason why, sorry Jens, are you waving or just cool. Um, the only reason why I would not completely, completely agree is I think that there is a methodological reproducibility here. Uh okay, so digre oh.

****Mikiko Tanifuji**** (26:14): Just talk by, thank you. I I'd like to clarify your uh your uh issue identified here is to based on the idea that one LLM for one research topic, or?

****Brian**** (26:28): Who are you and?



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



****Mikiko Tanifuji** (26:29):** Ah, sorry. It's Mikiko Tanifuji from NII Tokyo, uh, providing National Research Data Infrastructure. Now, What was my question? Um, So one LLM for one of the research processes, is that what you're trying to identify?

****Brian** (26:51):** So I am when you engage with Claude or Gemini or at worst, chat GPT, please don't use the free version, you will be sad. Um, every single chat interaction that we do as part of our research process, understood broadly, is part of our methodology and is subject to other people going, "Hey, do I like your judgment?"

****Mikiko Tanifuji** (27:22):** Well, in such case, how do you how do you uh document AI agents, which is normally part of entire a research project?

****Brian** (27:36):** Um, that's a really good question that we need to set standards for. Um, a lot of the movement of large language model industry is moving towards less and less transparency. And I think that we have a position to say, "No, if you're going to publish with this, we've got to see the gears, we've got to see the guts and inspect them." To get back to the prior point, uh, Mollick's Grimores, the fundamental thesis there is that the tacit knowledge in our head is rendered explicit when we turn it into a prompt that works. And just like a spell book, we can pass it to someone else and they can take advantage of the explicit knowledge that we have turned into text. And it's this process of making the tacit explicit where I think that there is flavors of reproducibility. I don't think it's reproducibility in the classical sense, but I think it's more than transparency because there's very much that reuse question. Behind you and then you.

****Andrew Treloar** (28:46):** Andrew Treloar, Australian Research Data Commons. So, I like the spell analogy. The problem is that the cauldrons keep changing. Oh yeah. And the version, and the versions of the cauldrons keep changing. And, and worse, it's often not possible to go back to an earlier version of the cauldron. So, knowing that you use spell 35 on cauldron version 1.0 is not going to do me a lot of good if I can't get access to that particular cauldron.

****Louis Mapatagane** (29:21):** different topic. It's I was just adding on what what was been said. Uh, just an idea that came up because we usually in the re well, science general science research, you would when you do your statistics, then document your variance. Is it possible with LLMs to then when you are with the thesis that you're proposing, to also then document what would be the variance of your output, then perhaps then you could more or less control for the variability of what would come out in future if we are trying to do reproducibility.

****Brian** (30:05):** So let me respond to this in a stack. Um, yes, and if you're using AI as part of your research workflow, i.e., to code things, sorry, I use code in the social science sense. This is a good message. This is a bad message. We 100% do multiple runs to ensure consistency. In terms of using AI as part of our research, I I I genuinely think that there is an argument from pragmatism there. In terms of cauldron one, spell 35, love that analogy. Um, A, there are ways of interacting with older models through the API that don't exist through



uh, the happy happy interfaces. Um, so if someone is sufficiently determined, they can rerun it. I would say that for a lot of what we're doing, the idea of rerunning is a lot less interesting than the idea of looking at and judging and iterating. Like, and again, this is why I think that there is a legitimate tension between transparency and reproducibility, but I think that that iterating, 35A that I can iterate on because I saw your 35 is meaningful in a way that isn't just transparency. And I think playing with that tension is going to be a very important part of our breakout. I've already been talking for half an hour. Time is my enemy. Uh, so I would say that a rough thing that I'm talking about in terms of AI use is you build your prompts, you interact with them, you capture everything, you review, you refine, and you iterate, and as part of that, you are publishing. This is also how we build judgment in our outputs. The metacognitive looking at us looking at the output and judging if it's appropriate is a really important component of this whole thing because we are effectively, and I know that this is going to cause everyone who's not in the social sciences to look at me as if I'm weird because I am, um we're developing taste. We are saying yes, this output from AI accomplishes my goals. There is an there is an aesthetic subjective sense of utility there that is weird to everyone who says that they are just a hard line scientist.

****Brian** (32:50):** I would say that this process of using AI, especially the refine part, is central to the effective use of AI by us as professional researchers with professional trained judgment.

****Brian** (33:12):** Also, let's be clear, the best reason to publish your prompts and do all of this is you're building your own grimoire. Because no one else cares. Like, let's be clear. The person who benefits from all of this work for by myself is me because I can just go back to my old prompts and copy and paste. The second order people who benefit are peer reviewers who if they really care, can go look at my judgment. But I'm sorry, all of us have peer reviewed. We don't have time for that. So we, like, this is also a way of introducing better judgment of AI use into the process, just like better code documentation exists in that wise.

****Brian** (34:01):** We've covered the journal policies. I would say that there's deep inconsistent practices across disciplines. We can talk about that about attraction to technology, we can talk about that as do I as discipline X like writing papers? Because I certainly hear from my colleagues in uh science and engineering that they'll do all the research quite happily up until the paper and then they will drop it into an AI because they don't want to write the paper. Okay. That's fine if your prompts are good and if you're using a good model, please don't use bad models. Don't use free models for that. You're going to be sad. Right? There's also no shared research infrastructure. What we need literally, as you were talking about, cauldron one, spell 35 is something that we have built dedicated research infrastructure for for code. There is no reason why we cannot build that same research infrastructure given money, given time, given the long view of time for AI use where we genuinely can go, yes, this is reproducible as a chat, the here are the 10 variations of my chat. This is a robust line of thinking. The world's our oyster, theoretically.

****Brian** (35:39):** But we have no standards. We have no standards for use. We have no standards for, here are the places where if you use AI, you're just going to be sad this week



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



because of course the models change every six to eight weeks. But there is certainly a set of places where you will be sad if you use AI in that specific part of the research process. They change.

****Brian**** (36:08): So having some standards for that. This, by the way, is the paper I did with QUT, so if you're interested, that link should resolve for you all. Um, there are no standards. There are no norms.

****Brian**** (36:22): So, grimoires. One of the useful outputs of this besides raw prompt logs are patterns of interaction, spell 35, of tested prompts. But beyond tested prompts, documented workflows in the open data sense. Here is the process by which I take data of such and such nature and do this. Here is the process by which I do gap analysis. Here is the process by which I X. I work on spell 35A1, because what I'm really talking about is establishing norms for creating reusable patterns from prompts as a way of sharing our expertise, and tooling and if I'm very, very, very, very lucky, building community knowledge so that we can pass these repositories of useful techniques around, just like we used to and currently pass grad students around. The tour, anyone? Just me? Cool. Laboratory life, Bruno Latour. um and uh Peter Gallison talks about the tacit knowledge involved in grad students moving from lab to lab. and things don't work until the right grad student shows up. So, so like there is precedent for this in science.

****Kiera McNeice**** (37:52): The title of that slide is wrong.

****Brian**** (37:54): Grimoires. Yeah. No, it's correct. Okay. And that's how you know I wrote that. Yeah. Ah, I blame me.

****Brian**** (38:10): Okay. Um, we've been having some wonderful interaction. Anyone on this side? Thoughts? Comments? This part of the class is good. Awkward silence? Oh, all right.

****Brian**** (38:32): So, I'm going to outline a bunch of challenges that I thought of. I'm delighted to say that our discussion has made more. This is great. Um, what I'm going to want us to do, and yes, this is going to require thought and words and movement on your part, is cluster together for people who feel strongly about these challenges and come up with some discussion around that challenge. So here are a few, do not be limited to this. Here are a few. So one, missing standards. What needs are there for standardization? So what is a valid prompt documentation format? How do we capture interactions? What metadata is there on the shape of the cauldron? Can we possibly use version control? And how do we solve the attribution problem? Because that's going to be awful.

****Brian**** (39:32): Publisher requirements. What are your understandings of the various disciplines and the various publisher requirements? And what I want you to answer in this group is a normative question of what should journals and publishers actually require. I'm not sure if I should forbid you or encourage you to join that group.



****Brian**** (39:59): Technical barriers. What are the technical barriers we already know that are problems? Volume is going to be a fun one, privacy is going to be awful, proprietary is going to be awful. like literally it goes on, it's fractal.

****Brian**** (40:18): Ethics, because yeah, I'm also a philosopher and thinking about the ethics of this is critical as part of establishing an interest group, because we need to have opinions. How do we deal with sensitive data? How do you deal with a researcher who's going along talking about their research, digresses into problems with their home life, and then goes back and dealing with their research? Not speaking from experience there. How do we deal with PII? How do we deal with proprietary information? Here's a really important question. Quite a few of my interactions with AI as I'm working towards a paper or a methodology don't work. This is the whole null publication problem. How do you like, do you want to document every attempt or do you only want to document the successful ones? There's a real tension there. And then I think a real big one is context. What amount of context is appropriate and ethical to document here?

****Brian**** (41:29): So, it's 5:09. This session ends at 6:00. I think a 20-minute discussion at the end is appropriate because once groups start going, they will get going. Um, the online is quiet so I'm going to engage with them myself if they have questions, otherwise we'll go on. I would like you all to group up and each answer one of those questions or, even better, a question of your own devising. Um I would say if people have deep opinions about one of those things to shout loudly right now to gather people who are interested with them. What I've done, if you want to be fun, is I have given you a prompt that you can drop into AI. Like, this is one of my fundamental techniques as a way of build of operationalizing and building consensus. If you want to be spicy, drop this into Claude, I would say, is probably going to be better. Um, Gemini will "well, actually" you to death, but one of the two. Um, play with that.

****Brian**** (42:42): So, anyone who feels strongly, start saying what your topic is, otherwise group up and drop your discussion into the shared notes. I will be orbiting. We will have at least 30 minutes and I might push it if we're still talking.

****Kiera McNeice**** (42:58): I'd like to talk about journal requirements and publisher policies.

****Brian**** (43:00): Excellent. Journal requirements and policies, person.

****Speaker X**** (43:04): Can you go back to the question?

****Brian**** (43:05): Absolutely.

****Group Discussion**** (43:10): [Background chatter as groups form and begin discussions]

****Brian**** (45:06): Sorry, I have to do this on this computer, not that computer. Um, one moment and I'll... delightful. Um, it's not complicated enough now. This is simple. It's simple stuff. Sure. How hard can it be? Okay. Why... Cut, paste. Great. There you are.



****Jens Klump**** (45:36): Hi, I'm going to just be back in the chat. Yeah yeah. Because that's the kind of tragic individual I am. The complication is you're assuming that people are just using AI as black boxes. You know, you give it an incantation, it spits something out. I first approximation. The complication is what happens once can be into agent?

****Brian**** (45:57): Well so you see how neatly I dodged that that question?

****Jens Klump**** (46:01): No, I wasn't here when you did the dodging.

****Brian**** (46:03): Uh, yes. Basically basically uh, we need to be pushing for transparent agentic runs. Good luck with that. Well, yes, but also we need to have standards on how to document the agentic runs.

****Jens Klump**** (46:19): The problem is that once you've got black boxes that are invoking other black boxes, I just think it's all.

****Brian**** (46:23): It's turtles all the way down.

****Jens Klump**** (46:24): Well, not it's not just turtles all the way down, it's going to become exponentially worse at each step.

****Brian**** (46:29): I yes. I still think that there is virtue in what we are doing rather than abdicating all responsibility because you can look at the run and see if the run is sane. And there is a norm setting responsibility that we have.

****Group Discussion**** (46:48): [Group discussion continues]

****Brian**** (56:59): you're trying to do.

****Group Discussion**** (57:00): [Group discussion continues]

****Brian**** (59:02): your check.

****Group Discussion**** (59:03): [Group discussion continues]

****Brian**** (59:22): they're the best analogy I've seen.

****Group Discussion**** (59:23): [Group discussion continues]

****Brian**** (59:33): to make research.

****Group Discussion**** (59:34): [Group discussion continues]

****Brian**** (59:40): power.

****Group Discussion**** (59:41): [Group discussion continues]



****Brian**** (59:51): does it try to please you?

****Group Discussion**** (59:52): [Group discussion continues]

****Brian**** (59:56): I'll box.

****Group Discussion**** (59:57): [Group discussion continues]

****Brian**** (01:00:11): That's always, that's always.

****Group Discussion**** (01:00:13): [Group discussion continues]

****Brian**** (01:00:20): it's something that we're talking about.

****Group Discussion**** (01:00:22): [Group discussion continues]

****Brian**** (01:00:30): That's why we want to.

****Group Discussion**** (01:00:31): [Group discussion continues]

****Brian**** (01:00:48): to be a bit careful about what we're going to.

****Group Discussion**** (01:00:50): [Group discussion continues]

****Brian**** (01:01:28): just get the sense to get to do the right thing for what we want.

****Group Discussion**** (01:01:31): [Group discussion continues]

****Brian**** (01:01:41): to.

****Group Discussion**** (01:01:42): [Group discussion continues]

****Brian**** (01:01:51): I know nothing, I'm just looking at the picture.

****Group Discussion**** (01:01:52): [Group discussion continues]

****Brian**** (01:02:43): to.

****Group Discussion**** (01:02:44): [Group discussion continues]

****Brian**** (01:03:04): No, that absolutely not.

****Group Discussion**** (01:03:05): [Group discussion continues]

****Brian**** (01:03:51): it's like we're trying what we can and obviously we're.



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



****Group Discussion**** (01:03:55): [Group discussion continues]

****Brian**** (01:04:33): Okay.

****Brian**** (01:04:46): Okay. It the cuckoo clock has rung and we are 20 minutes out. We are 18 minutes out. I stand corrected.

****Brian**** (01:05:00): We have two options because it looks like groups are really into this. One, each of you can tell the other, each group can give a brief report to the others about what they're thinking about, or two, we can just keep talking and keep taking notes and we'll write up our notes in an email and send it to everyone who's interested.

****Christina Drummond**** (01:05:24): Okay.

****Kiera McNeice**** (01:05:25): I vote for one so we know if we're talking about the same things.

****Brian**** (01:05:29): Fantastic. Um, in that, in that case, I will regretfully interrupt you all. Yes.

****Christina Drummond**** (01:05:34): Um I would suggest an addition on this, after we share around so we know what we talked about and we also talk about what are next steps for this.

****Brian**** (01:05:41): Yes. We explicitly need to take next steps for an interest group. And you will see that I have started that with the most fun part of PII. And if you're interested, we can signal our interest there, and I'll start talking to the RDA people about what that entails. Let me be clear. I will use what is in our notes as the basis of our remit, not what I came up with. If there are parts of what I came up with that you think are worthwhile, make sure they're in our notes. This is how democracy works, I think.

****Brian**** (01:06:20): All right. Um, so I will choose by proximity this group first. Who is your spokesperson? This is literally undergrad group conversation. Who's your spokesperson?

****Andrew Treloar**** (01:06:36): Um, which group? Andrew Treloar. Which group? Whatever. Um, standardization. Great.

****Andrew Treloar**** (01:06:47): Uh, Andrew Treloar, Australian Research Data Commons, but you probably know that already. Uh, so I guess the key, we had a fantastic conversation. The key insight from it would be that essentially we're just talking about methods. So rather than talk about transparency or reproducibility, just talk about document your methods, guys. Uh, and so then the question is, great, but how? Uh, and so the what could be extended, captured by my beautiful assistant on the left. Uh, kind of stealing from research software, so FAIR for research software is a stuff there we can filch, CodeMeta's another one. Um, Jens suggested, yeah, but aren't we just talking about workflows? So the workflow languages that we could steal stuff from, if we capture this in a method section, where's a good place to



standardize it? Protocols.io is one of the suggestions. Um, so that looks like it's a tractable problem. The challenge is going to be working out what's the minimum information that's needed for the method section and how do you do this in a way that doesn't just impose more unfunded work on the researchers? So how can you, how can you make it as easy as possible? And one of the, one of the cooler ideas was at the start of your interaction, you tell the LLM, at the end of this interaction, I'm going to ask you to document it in this format, and then at the end you say, now spit it out for me in this format. Um, I mean it's still more work for the researcher but it's much less small work.

****Brian**** (01:08:16): That's a terrible sentence.

****Andrew Treloar**** (01:08:17): Yes.

****Brian**** (01:08:18): There is some discussion that I want to have around meta-work, like is copy editing in the method section? So I think that there is, but that is, that is, it's a live question. And I would say that ideation and everything up to copy editing so it's not into an infinite recurse. Um, excellent. So that group is interesting. We are going.

****Kiera McNeice**** (01:08:42): Actually, a comment on that point because, oh, all right, but also you've volunteered. Yeah.

****Brian**** (01:08:49): I am in, I am in full classroom mode right now.

****Kiera McNeice**** (01:08:55): Yeah, on on the copy editing question, it depends. I mean sometimes when people use a full-on external English language editing service or whatever to publish a journal article, we do ask them to declare it or if someone's been extensively involved. So yeah, it depends to what extent, um, yeah, and it varies a lot by discipline again, is what I've seen. Um yeah. Um, we sort of had a pretty broad ranging conversation. Um, uh, I mean my experience is that it would be very difficult to standardize a lot of these things across all publishers and fields. There are really different views even within disciplines. Um, but there are some things that we can all agree on like AI can't be listed as an author of an article because the author's responsibilities are more than just, yeah, writing the article.

****Brian**** (01:09:34): AI cannot be fired, therefore it cannot be an author.

****Kiera McNeice**** (01:09:37): We talked a lot about what people are using AI for and what's counts as like a benign use as like just spell checking as opposed to actually using AI to do significant parts of the research. Um, and how that might be declared, which again would probably vary across different disciplines a lot, but, um, and we talked a lot about how the researchers are still ultimately responsible for checking the outputs of an LLM and checking that it's actually valid work and that very few people would just take something and paste the whole output from a prompt word for word into a research article or something like that. Uh but on the other hand we talked about the tensions of the publisher parish mentality and how people are have motivations to just put out as many articles as possible and how publishers also if they're reliance on the APC model of publishing have an incentive to publish as many



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



articles as possible, accept as many articles as possible. Um and so there's yeah, work for journals there too to make sure we're properly doing peer review. Um, I'm sort of jumping around here, sorry. Um, we talked about the, yeah, the extra effort required of researchers and how, you know, if you're writing an article, it's an iterative process you go through prompt after prompt and maybe hundreds or thousands of prompts, is it really sensible to document every single one of them and who's going to have the time to do that? Um, so without machine actionable ways to, yeah, get an output from the AI potentially, it's going to be difficult to encourage people. Um, we talked also a bit about using AI in research and having access to content, so access to, you know, information from different databases, if they're all building their own LLMs and saying if you want to use AI to analyze our content, you must use our LLM and that creates a fragmented landscape, much like the old text and data mining problems where if you wanted to do a systematic review of a corpus of journal articles, you'd have to go to 15 different publishers and get agreement from each one of them. And so if yeah, each different uh, database publisher, whatever has their own LLM and says you can't combine their stuff with other stuff in an LLM, that yeah, causes issues for research. Um, what else is, anyone, was there anything else key here that I haven't touched on yet? And yeah, how many authors are actually acknowledging AI use, because the comment here was everyone's absolutely using LLMs.

****Brian**** (01:12:47): And definitionally, because most almost all authors are using Word.

****Kiera McNeice**** (01:12:51): Yeah. Yeah, as I said like Office 365 explicitly markets their AI like their spell checking as an AI service. So everyone's using it, um, and I know from our perspective, it's it's really difficult for to actually to actually get an answer to that question, how many declarations are there because there's no standard way to declare it yet. Um, we don't like data availability statements we've gotten pretty standardized on, but we've had some journals ask us, you know, can we have an a like an AI statement as a separate thing in an article? Does that get captured in metadata? Because at the moment some people put it in as part of the methods, some people use an acknowledgement, um and yeah, so there's part of the discussion is what you should declare and why and where is also to be able to report on who's declaring it and what they're declaring. Um yeah, okay, sorry I'll stop.

****Brian**** (01:13:33): No, no, no, it's So what I'm going to say, given that we're 10 minutes out and I'm just trying to make sure that I'm not keeping you for beer time, is for the groups who have already presented, if there are topics that you think the interest group should be interested in based on your presentations or your discussions, drop them in the interest in interest group section or bold them or somehow signify them so that we have useful metadata from the output of today. You folks are close, therefore, who's your volunteer?

****Charlie Lembit**** (01:14:04): Um, hello, I'm Charlie. I work at the Murray Darling Basin Authority, um, from Australia, I guess. Um... Uh, so yeah, we sort of talked about a range of things. We first talked about, um, well, we we're talking about ethics, sorry, first of all. Uh, so we talked about, you know, governance, um, protocols, what they might be for even using um, LLMs to even uh... I guess use sensitive data, um, in that context like what would even be allowed, what was appropriate with that. Um, you know, it needs to be very specific to the context is what we came to the, you know, conclusion of, you can't really have a universal



um... standard-ation of that. Um, and then we sort of talked about the levels of trust, um, within that. So I think, um, how would we know it was trustworthy? Someone made a was it you said prompt, uh process and provenance is the three Ps, which I thought, you know, that was quite a good uh way of looking at it. So, you know, where did the, where did it come from? What process was used? What prompts were used? And that would be a way to know that it was trustworthy. Um, and then whether putting the sensitive data in as a prompt, what could actually be disclosed? Cuz obviously you couldn't disclose the sensitive data. So whether that would be just showing uh, the process of of it as opposed to the final data. Um, and then the last sort of thing was talking about, uh, the expensive, uh commercial AIs, um, and but... well not sorry, um, but basically people saying sometimes the better uh, LLMs could be cost prohibitive and potentially having something in between that's more of a collective ownership to make it, um, more accessible, um, especially to groups that uh more likely to have sensitive data, Indigenous people, um disabilities, all those things. So, medical data, um, things where we wouldn't want it to be cost prohibitive. Was there anything else anyone want to add? I feel like...

****Brian**** (01:17:39): Based on just that last one alone, I will say one of the things we might want to say is our um requests to NCI for serving large language models might be part of this. I'm sure that they would love to hear what we have to say and have no problems whatsoever with that. But yes, uh cool. Uh and I think the last group is over there.

****Speaker X**** (01:18:04): Yeah, I have to say we were kind of slightly naive version of the last group. That's fine. Thinking about publisher requirements. Several of us didn't really have much practical experience or knowledge of having to do that professionally. We document all this.

****James Wilson**** (01:18:22): Yes. Um, so we've come up with just uh, it was a a simple exercise really, just to think of the kind of kinds of metadata we might want to try to capture, um, without really going on to think about how easy or difficult it would be to actually capture the effort required to record it or anything else. But, um, we started off talking a little bit about the need to capture motivations, so the motivation for the research topic, the motivation for why you're using AI, the motivation for choosing the model, all very important contextual information, um, which you couldn't otherwise capture easily. The model used, when, who, um, how the app model was accessed, whether it was via um, a standard kind of prompts in a via web interface or the API. Um, something about the, the data used for training. I'm not sure how possible it is to even know that in sort of black box models, but the kinds of responses you're going to get are going to be so dependent on the training data. How you go about capturing that? No idea. Um, but it seemed important to us in our naive way. Um...

****Amy Nurnberger**** (01:19:34): There is a question, I believe.

****James Wilson**** (01:19:36): Oh, I'll just interrupt myself then.

****Amy Nurnberger**** (01:19:37): Sure, yeah.



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



****Amy Nurnberger**** (01:19:39): Live. Amy Nurnberger, Massachusetts Institute of Technology. On the question of the training data, I think there are two separate questions there, right? You have your large language models that many of them are quite opaque, but how I see researchers using large language models is actually not in the form of authorship, but in the form of other types of research where they will take a smaller model that they run locally, train it on their own set of data, right? And thinking about how LLMs are being used in that way, which I want to make sure is included in this overall conversation as well, since it seems to have been fairly focused on LLMs for authorship or for ideation. But thinking about fine-tuning for other purposes too. So not at all unreasonable.

****James Wilson**** (01:20:22): No, fine-tuning and basically anything after the production of the foundation model can be in our remit. So, please continue. I think that was actually pretty much it. I don't think uh, we spoke that much about any of the other things. Does anybody else in the group have something they'd like to mention I'm missing in our uh, relatively short and sweet. I'll get it. System prompts. System prompts. That's true, actually, we did talk a little bit about the system prompts, not least because I didn't really know what they were. Um, but now I do. I feel much better informed. Um, but trying to capture, yeah, the system prompts that have gone into it that are lying behind things would also be very useful.

****Brian**** (01:21:07): Okay. So, thank you all. You have accomplished my goals exactly. Um, the rough pattern is I will be using, sorry, I need to be using my indoor voice and the microphone. Um, I will be using this as a pitch to turn this group into an interest group for RDA so that we can start talking about XKCD's 18th standard, uh, because that is, let's be clear, what we're proposing, but hey, tradition. Um, I would love if as thoughts occurred to you live topics of interest to the interest group, not right now, drop it in the doc. I will figure out some form of communication for these emails. Honestly, probably just a mailing list because anything else will be too annoying, but I'll figure it out. Um, and I will start gathering us together. So, I've promised a number of you beers. Um, and to talk about AI and education policy and my various spicy takes thereof. Um, as soon as we're done with that welcome reception, that that is on offer. So thank you all and please email me with whatever questions you have around AI, research, education, policy, et cetera. Thank you. (applause)