DREAM Principles and FAIR Metrics

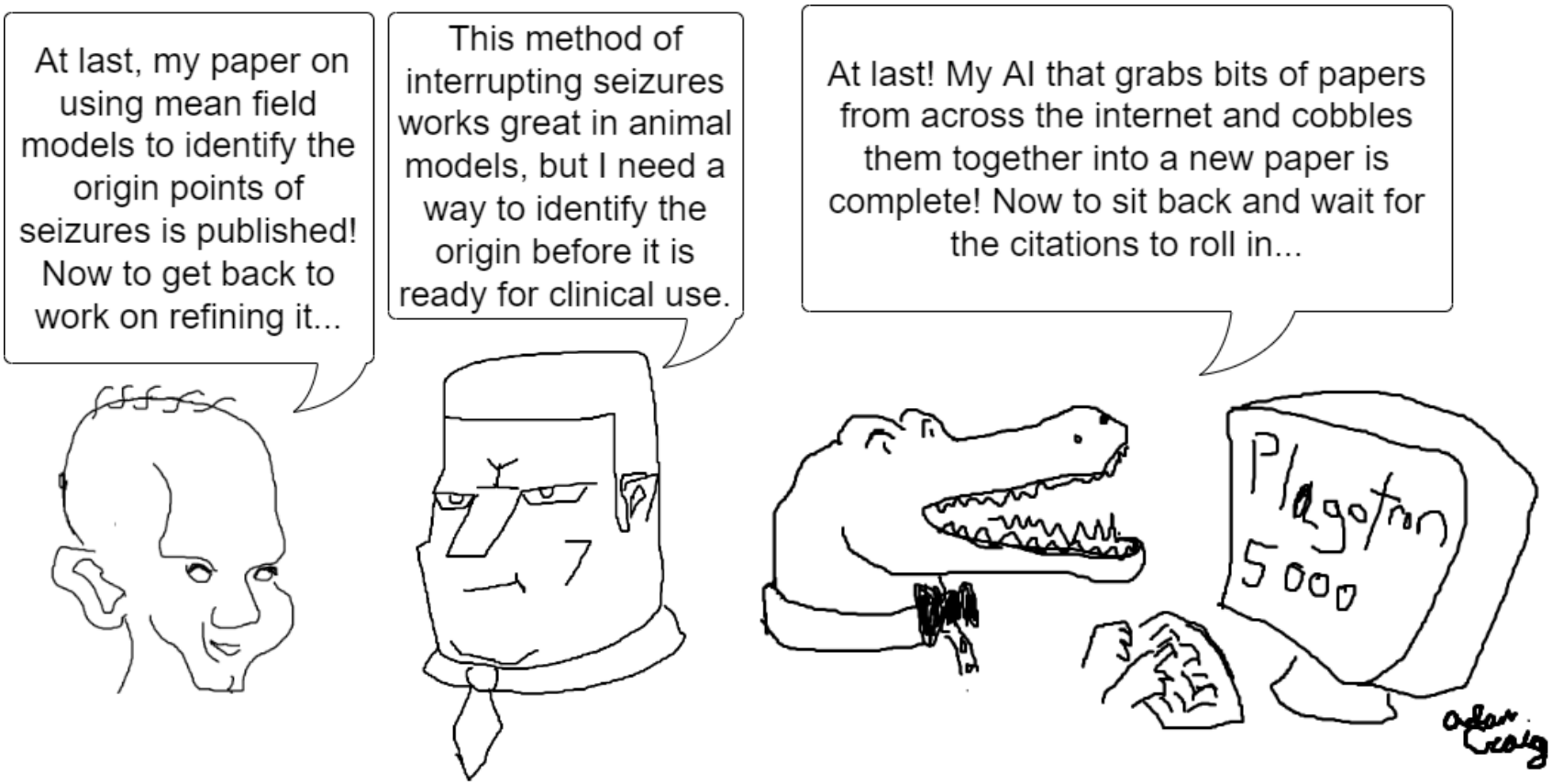
Adam Craig, S. Koby Taswell, Anousha Athreya, Carl Taswell

Brain Health Alliance Virtual Institute, Ladera Ranch, CA, USA

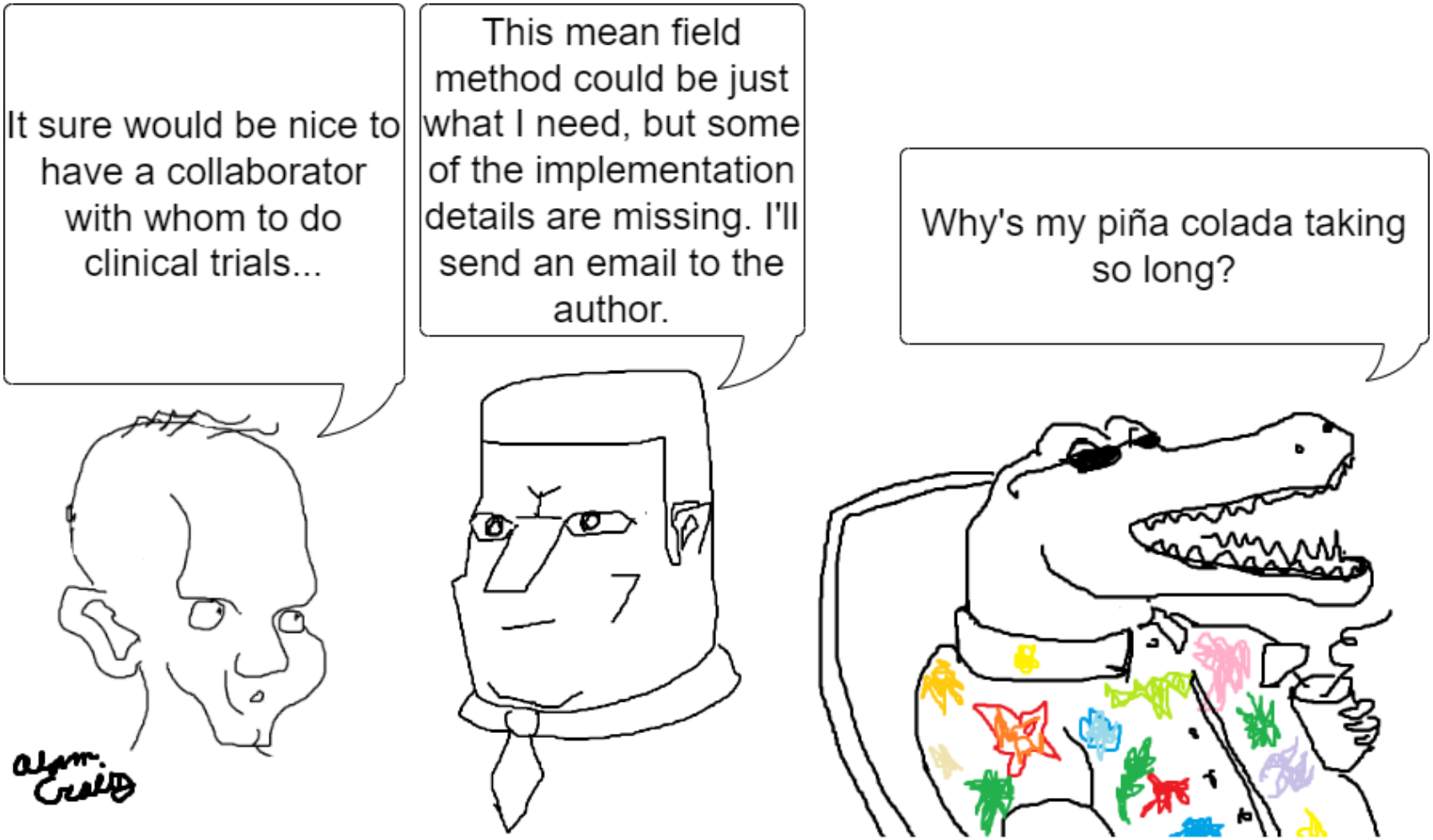
2022 ASIS&T Workshop: October 9 @ 13:00 EDT
Who Are the Guardians of Truth and Integrity?

A brief what if...

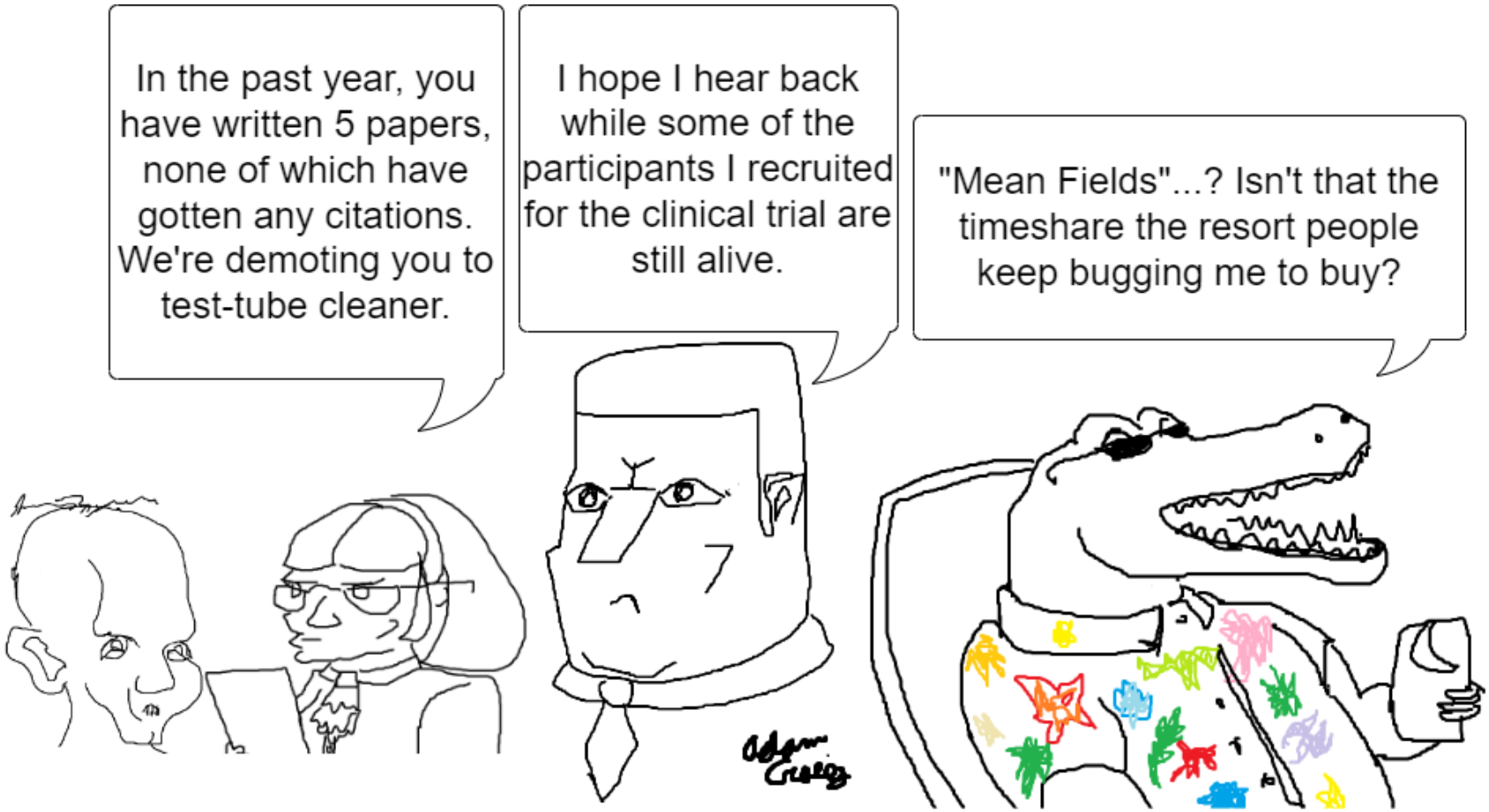
very loosely inspired by the Retraction Watch article "'Tortured phrases', lost in translation: Sleuths find even more problems at journal that just flagged 400 papers"



A little while later...



Later still...


In the past year, you have written 5 papers, none of which have gotten any citations. We're demoting you to test-tube cleaner.
I hope I hear back while some of the participants I recruited for the clinical trial are still alive.
"Mean Fields"...? Isn't that the timeshare the resort people keep bugging me to buy?

Victims of plagiarism in science

- The author: denied credit, recognition, opportunities for further work and collaboration
- The institute: duped into funding a goldbricker (worker just pretends to work)
- Other scientists: cannot trace an idea back to its source, leading to problems with reproducibility
- Patients and other stakeholders: denied the benefits of new discoveries the real researcher and collaborators could have made

Plagiarism beyond copy-and-paste

- Even if the words are different, if the ideas are the same, it's still plagiarism.
- Developers of plagiarism detection software are in an ongoing arms race with developers of paper-generating software that automatically paraphrases plagiarized content, e.g., SCIgen.
- See the previously mentioned article for further discussion.
- Even if you replace the word "ginger" with "asparagus", it's still plagiarism (as seen in the Retraction Watch article "When is asparagus not asparagus? Why, when it's ginger, of course!").

Ways ideas can be the same

1. Verbatim copying
2. Paraphrasing: synonyms, restructured sentences
3. Summarizing: cutting out the details, keeping the big picture
4. Expanding: embellishing, recontextualizing, etc. while keeping the old core idea
5. Republishing the same idea or set of ideas under a different name

- If it does not cite the original, it is still plagiarism.
- For more, see Athreya et al., 2020, "Essential Question: 'Equal or Equivalent Entities?' About Two Things as Same, Similar, or Different"

Editorial oversight: velociraptors guarding the henhouse?

- Former editors Kuo-Chen Chou (investigated 2020) of *Journal of Theoretical Biology* and Artemi Cerdà of *Artemi Cerdà* (investigated 2017) and *SOIL* are both former due to having pressured authors to cite more of their papers (Lockwood, 2020, *Proceedings of the Royal Society A*).
- In 2016, Thomson Reuters suspended the journal *Applied Clinical Informatics* for one year, in the belief that editors artificially boosted the impact factor of sister journal *Methods of Information in Medicine* (Lehmann et al., 2022, *Applied Clinical Informatics*).

The need for open and transparent peer review

- Authors may avoid citing the work of potential rivals, leading to idea-laundering plagiarism.
- They may also misrepresent the content of prior work.
- Peer reviewers and editors may have their own biases or perverse incentives, leading to idea-bleaching censorship.
- Institutional ethics committees may care more about avoiding damage to the institute's reputation than about righting wrongs.
- Open peer review forces reviewers and editors to be accountable for their judgments.
- Transparent peer review forces them to make their reasoning clear.
- For further discussion, see Taswell et al., 2020, "The Hitchhiker's Guide to Scholarly Research Integrity".

Open and transparent peer review on the NPDS cyberinfrastructure

With Nexus-PORTAL-DOORS-Scribe (NPDS) repositories, you can manage...

- multiple versions of the same document
- references to the same document in multiple repositories defined by problem domain
- concept-validating constraints (automatic checking for relevance to scope)
- reviews from multiple contributors
- systematic reviews and comparisons of articles using embedded microformats
- public and private records

The FAIR Metrics of good citation practice

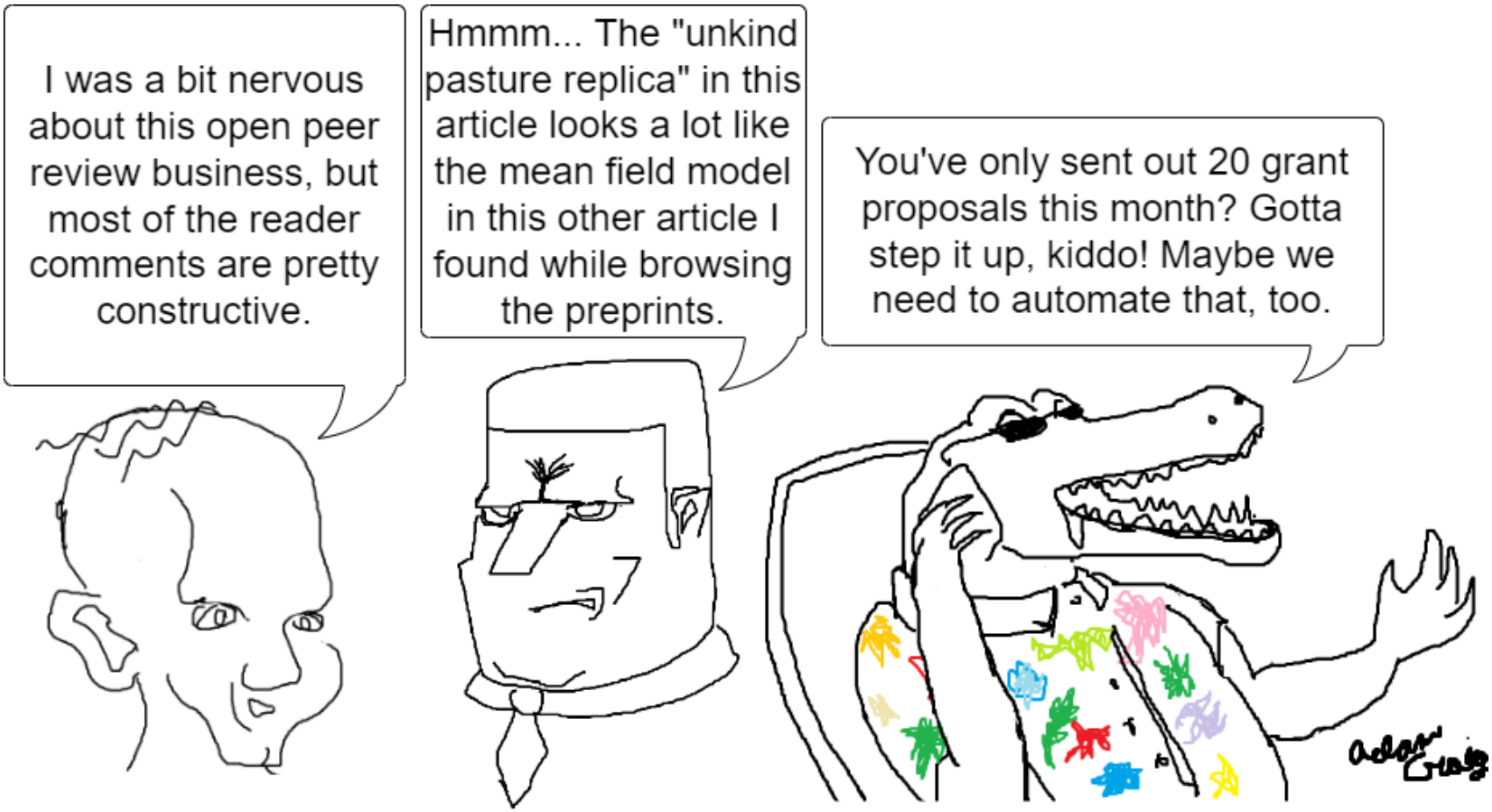
- The Fair Attribution to Indexed Reports (FAIR) Metrics evaluate how appropriately researchers attribute substantive claims to their sources.
- In (Craig & Taswell, 2018, *BIBM 2018*), we defined 4 counts:
- Q for Quoted: claims correctly attributed to prior work
- M for Misquoted: claims misrepresenting the content of prior work
- P for Potentially Plagiarized: claims presented as novel but matching those in prior work
- N for Novel: claims presented as novel and not found in prior work
- In (Craig et al., 2019, *ASIS&T 2019*), we introduced 4 ratio FAIR Metrics:
- F_Q: overall frequency of valid attributions to prior work
- F_M: emphasis on misrepresentation of prior work
- F_P: emphasis on plagiarism
- F_N: balance between new ideas and prior context

The DREAM Principles

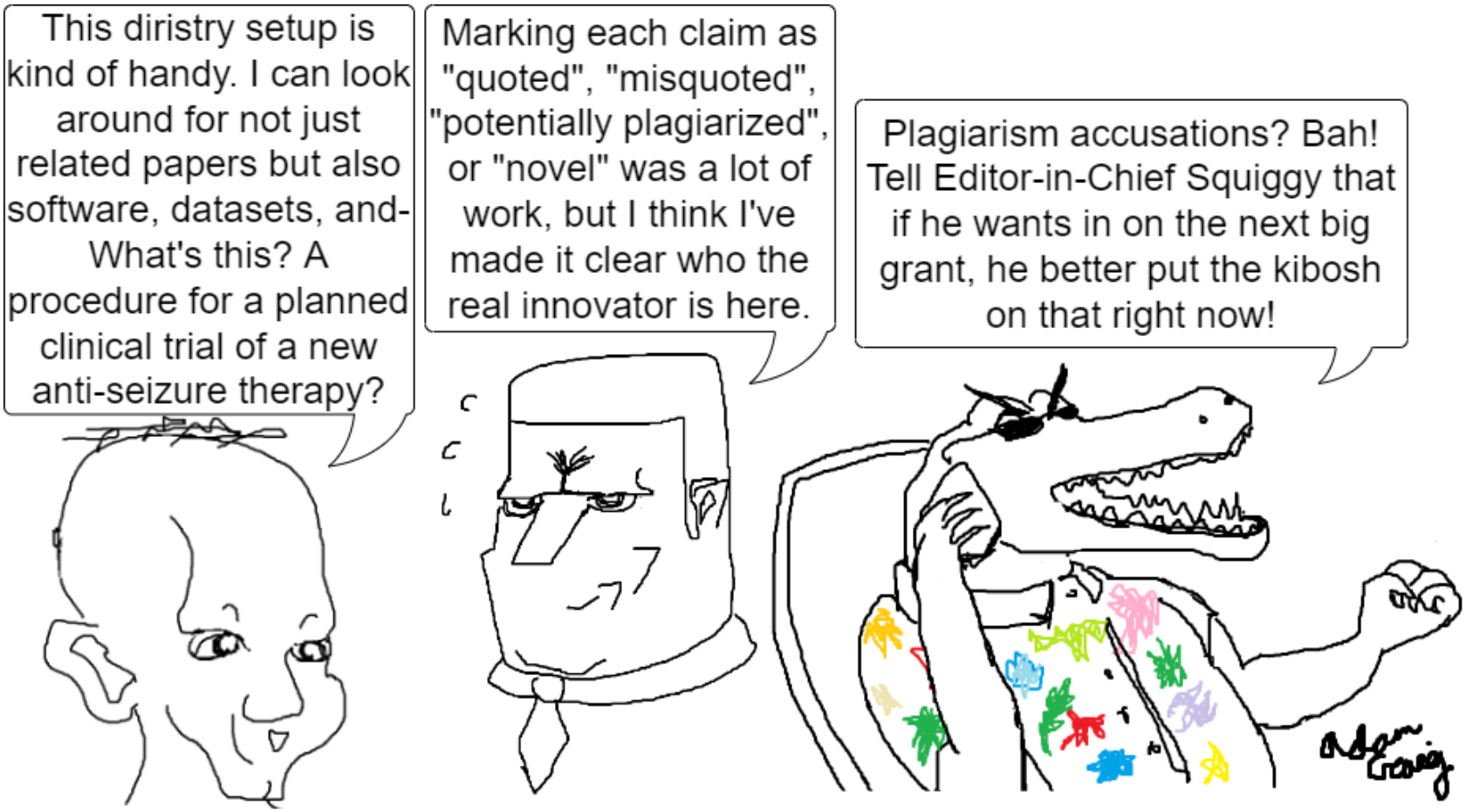
An evolution of the original PORTAL-DOORS Project design principles:

- **D** Discoverable Data (also Decentralized, Distributed, Democratized Diristries)
- **R** (with) Reproducible Results (also Repositories, Records, Resources, Reusable)
- **E** (for) Equivalent Entities (also Enhanced, Extensible)
- **A** (with) Accessible Attributes (also Accurate, Available)
- **M** (and) Manageable Metadata

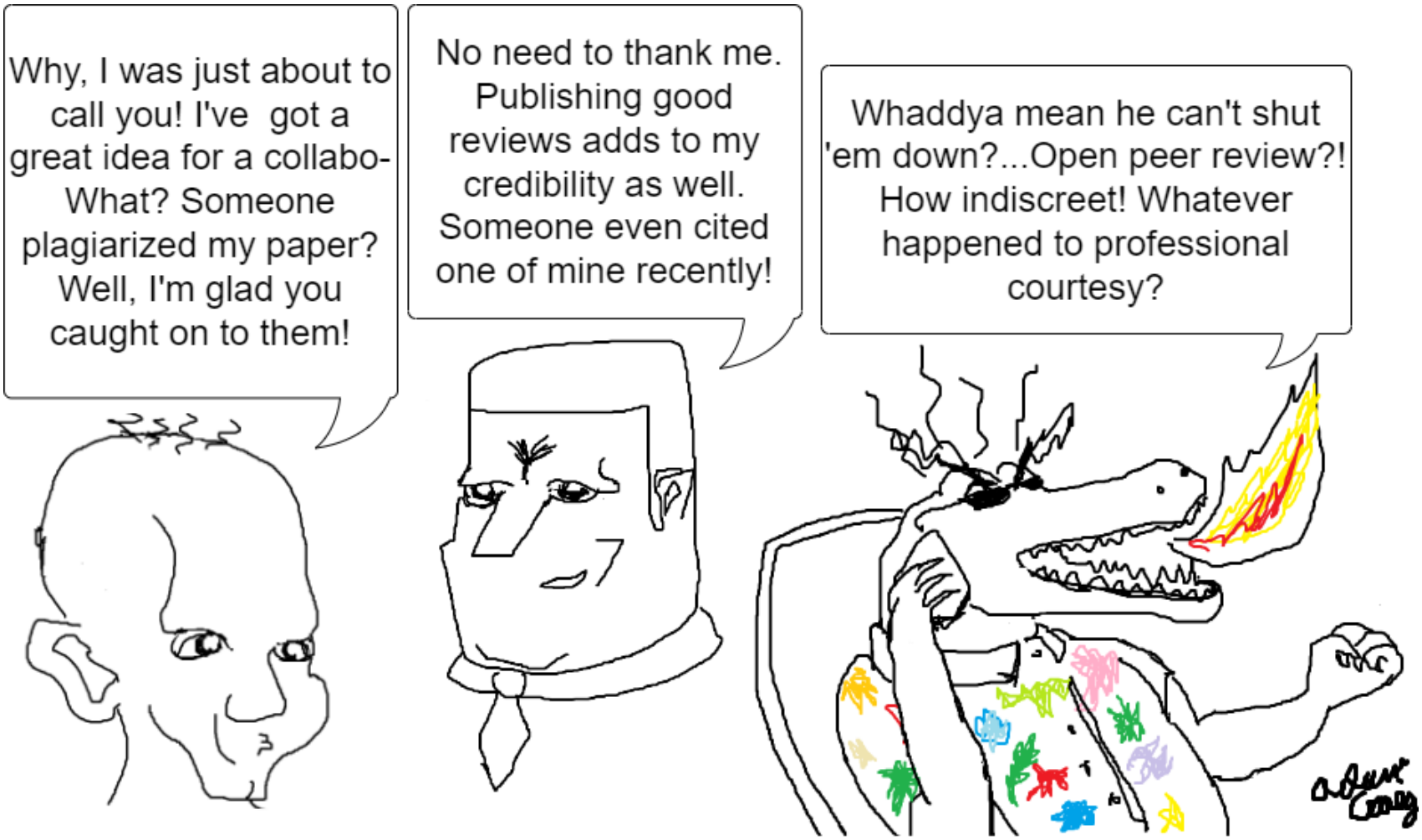
Another what if...



Some hours later...



After much ado...



Conclusion

- Plagiarism hurts not only the true authors but the entire process of science and those who depend on its results.
- Open and transparent peer review is necessary to combat plagiarism and other forms of misconduct.
- The NPDS cyberinfrastructure can help you keep track of manuscripts and reviews, making open and transparent peer review manageable.
- The FAIR Metrics provide a more systematic way to evaluate how honest authors are about the provenance of their ideas.

References, part 1

All PDP literature available at
https://bhavi.us/NPDS/PdpSite/Papers

- Marcus, A. (2021).'Tortured phrases', lost in translation: Sleuths find even more problems at journal that just flagged 400 papers *Retraction Watch* (website) (Retrieved 2022-10-06).
- Marcus, A. (2018). When is asparagus not asparagus? Why, when it's ginger, of course! *Retraction Watch* (website) (Retrieved 2022-10-06).
- Athreya, A., Taswell, S. K., Mashkoor, S., & Taswell, C. (2020, September). Essential question: 'equal or equivalent entities?' about two things as same, similar, or different. In *2020 Second International Conference on Transdisciplinary AI (TransAI)* (pp. 123-124). IEEE.

References, part 2

- Lockwood, M. (2020). Citation malpractice. *Proceedings of the Royal Society A*, 476(2242), 20200746.
- Lehmann, C. U., Ball, M. J., Haux, R., & Lehmann, J. S. (2022). Applied Clinical Informatics Journal: A Brief History. *Applied Clinical Informatics*, 13(02), 516-520.
- Taswell, S. K., Triggle, C., Vayo, J., Dutta, S., & Taswell, C. (2020). The hitchhiker's guide to scholarly research integrity. *Proceedings of the Association for Information Science and Technology*, 57(1), e223.

References, part 3

- Craig, A., Lee, C., Bala, N., & Taswell, C. (2022). Motivating and Maintaining Ethics, Equity, Effectiveness, Efficiency, and Expertise in Peer Review. *Brainiacs Journal*, 3(1).
- Craig, A. & Taswell, C. (2022, September). Inclusion and Exclusion Criteria for Automating Adherence to Scope of Conference Calls for Papers. In *2022 4th IEEE Conference on Transdisciplinary Artificial Intelligence* (TransAI). IEEE.
- Craig, A., Ambati, A., Dutta, S., Kowshik, P., Nori, S., Taswell, S. K., ... & Taswell, C. (2019, June). DREAM principles and FAIR metrics from the PORTAL-DOORS Project for the semantic web. In 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1-10). IEEE.

Contact Info

- ctaswell@bhavi.us
- www.BHAVI.us
- www.BrainHealthAlliance.org