

Authors want to deeply thank the recommender as well as the two reviewers for their suggestions and advice. We believe the manuscript is now way more straightforward and accessible which was the main objective when writing it. When provided, their names were added to the acknowledgements.

Additional information : there was an issue with the upload on the preprint server and the interactive PDF file could not be kept, the EcoEvovRiv team is working on it

Decision for round #2 : *Revision needed*

Both reviewers are satisfied with the improvements, and are only recommending (very) minor wording / clarification changes. I will not send this preprint for review again when they are done.

by [Timothée Poisot](#), 12 Aug 2024 14:19

Manuscript: <https://doi.org/10.32942/X2G033>

version: 2

Review by Nick Isaac, 01 Jul 2024 11:56

This revised ms is much clearer and I am pleased to see that many suggestions from the first round of review have been adopted. I would be happy to see this work published: in addition to providing an advertisement for the Galaxy platform, it makes a number of really good points about computational practices in ecological research. Most of my recommended changes (below) are aimed at clarifying the message and simplifying the language:

Some of the new text appears to have been rather hastily inserted and could be improved. For example, on line 162: "propositions have been delimited within several thematic communities in ecology to evaluate and enhance best practices application, notably the SDM communities" should be simplified to "Individual subdisciplines have taken initiative to assert community best practices, e.g. in SDMs" (or similar)

Thank you very much for your help on these enhancements. Modification included.

Line 138: "although data accessibility": this paragraph is not just a single sentence. Please combine with previous or next para.

Combined with next paragraph.

Line 181: replace "the expertise and research" with "the scientific process"

Modification included.

Line 182: replace "FAIR research data" with "the adoption of FAIR practices"

Modification included.

Line 244: the heading "frameworks towards best practice" is a bit confusing. Perhaps replace with "Principles for best practice"

Thank you, as "principles" do not necessarily imply a methodology authors opted for "Guidelines" instead.

Line 247: would benefit from a clearer definition of atomisation, perhaps with an example. Something like: "In a maximally-atomised workflow, each small step would be conducted by its own bespoke function"

Thank you, example added.

Line 195-7: replace “mechanically reduces the number of potential users ... fragilizing ...” with “creates barriers to uptake and challenges for peer-review.”

Modification included.

Line 220: replace “single” with “distinct”. Also insert “each” between “steps” and “constituting”.

Modification included.

Line 224-6: Around here it would be appropriate to have a sentence explaining how/why atomisation is part of the solution. (I realise much of this is covered in the later section from line 245, so overlap between sections should be minimised). My own recent experience of working on a colleague’s code has been that atomisation makes it much easier to alter the code, to make it do something different from what the original coder intended. Altering how atomised code is used is, of course, the first step towards generalisation. However, as written the two steps appear to be quite distinct.

As this paragraph is part of the introduction and the “how/why” of the atomisation-generalisation framework seems closer to “results/discussion” in our opinion, we decided to add a short straightforward sentence to illustrate your point: “Therefore, generalisation cannot be efficiently achieved without prior atomisation.” We hope it is sufficient to clarify this.

Line 227: replace “this framework The Galaxy-Ecology” with “Atomisation and Generalisability are central organising principles in the design of the Galaxy-Ecology”

Modification included.

Line 229: I don’t think that “sharing and processing data” does justice to what Galaxy is aiming for. How about “analysing data and sharing outputs”.

Modification included.

Line 284: again, an example would help to illustrate this point more clearly: “This means trying to avoid hard-coding anything that is specific to the structure of the original dataset (e.g. number of years).”

Definition included.

Line 290 the word “step” appears twice in this sentence. Replace the second with “element” or similar word.

Modification included.

Line 301: the section title is quite clunky: how about “Practical steps towards atomised and generalised coding”.

Modification included.

Line 319: the figure legend needs more detail here. Explain that the different colours refer to different scripts/categories (1-3) and that the boxes refer to functions/scripts.

Thank you, more details were added to the legend.

Line 321: re “code-writing habits”: I think it would be useful to make some kind of comment about how ecologists learn to code. Most of them learn by analysing their own data, and/or from a statistical ecology module that focusses on specific applications, in which the statistics and the ecology are given primacy but the computation. Few are taught formal programming skills in the way that computer science students would be. This means that most of us have generated a large number of “bad habits”!

Thank you, the comment has been added on lines 162-166.

Line 345: can you explain why someone should feel less embarrassed or fearful out sharing code if it has been atomised and generalised? Is it because they will feel confident about having followed best practice (I doubt it) or because they will feel confident that someone else will be able to actually read and implement it (more likely)?

Yes, it is something like the second reason. In the article by Gomes et al. (2022), they suggest sharing codes with trusted peers for review before submission to overcome the embarrassment/insecurity barrier. As trusted peers (or even the author of the code) can read and re-use codes more easily, the author can be better reassured when validation is provided for their work. This explanation has been included at the end of the paragraph.

Table 2 legend: would be more informative to write as “illustration of how Galaxy implements and confirms to best practice”. The second sentence about “limitations ...” makes no sense to me. Perhaps it is a reference to the fact that Galaxy is a work-in-progress that the table is a snapshot of current status.

Thank you, the sentence on limitations was mostly to acknowledge that not all best practices are attained simply through the atomisation-generalisation framework and that advices are included in the table to do so (e.g. “One must keep track of different parametrisation and input settings at each computation”). As it seems confusing, it was removed.

Line 444: can you elaborate on the figure 3 legend to relate it back to the concepts of atomisation and generalisability? In particular, it is appropriate to describe the named items with checkboxes as atoms in the workflow?

Direct links with the atomisation-generalisation framework has been added in the legend.

Line 459: this is very unclear. Does “eventually” mean “when the user becomes expert” or “there is an aspiration for Galaxy to have this new functionality”. If the former then perhaps replace with something like “In addition to using existing tools, users may develop and upload entirely new tools to the Galaxy server”.

Yes it was the former, thank you for this proposition it was included.

Line 462: “utterly” is superfluous

Word removed.

Line 464: “notably” is superfluous

Word removed.

Line 515: there is perhaps another level in this hierarchy: I have authored papers that were fully executable at the time of publication. However, we did not user Docker or other tools to account for changes in the underlying software, so the code no longer works and the work is

therefore not reproducible. Distinguishing between “reproducible now vs reproducible forever” might be helpful.

*Thank you, it is indeed an interesting comment to make on executable papers !
We added these elements.*

Line 531: “additionally”

Thank you, mistype corrected.

Line 550: replace “correctly” with “appropriately”

Modification included.

Line 552: “heavier” is not clear. I get that the “heaviness” refers to the amount of time investment required to realise the advantages of using Galaxy (but this could be clearer – please use a different word). What I don’t understand is the comparator: is it heavier for experienced than non-experienced, heavier to learn Galaxy than the principles of atomisation and generalisation, and are you referring to absolute or relative terms (i.e. cost vs cost:benefit ratio).

It is true that the mention of “experienced” is unnecessary here, it was removed. We wanted to highlight that scientists with programming experience might be reluctant to use an interfaced solution to make analyses, given their ability to conduct these analyses through computer codes. This paragraph justifies the formalisation of the atomisation-generalisation framework (that is helpful for scientists that remain unconvinced by Galaxy, which is valid) while also addressing these potential concerns.

Review by anonymous reviewer 2, 09 Aug 2024 16:24

This draft is greatly improved in structure from the previous version and I found it significantly easier to follow. The presentation of reproducibility as the goal and the Galaxy workflow as a solution is strong and clear, and generally the message is more concise. The section describing Galaxy-E and the discussion however have some issues with redundancy of some topics, and the absence of other topics, that make them harder to follow.

The “Entering a new dimension” section lists many of the ways Galaxy meets the criteria for reproducibility, or the benefits it might bring, but it doesn’t describe what it *is*. That piece is critical for this type of introductory paper – where does it live? How does the user engage? What are the key pieces? This part might include existing parts of the section such as the description of the community, or how a user uploads data, but should be more comprehensive and systematic. I think in the previous version many of those ideas lived in the “methods” section, which has rightly been moved to a different venue, but a paragraph or two of description is still necessary. Starting with a clear description also gives grounding for the platform’s benefits, as the reader already has clear evidence of how the tool might achieve those things. One possible structure for the “Entering a new dimension” section would be:

- What Galaxy-E is (how users engage with it)
- How Galaxy-E follows the atomisation/generalisation and reproducibility/fairness philosophy
- Other benefits
- Examples of its success

Thank you for this structure, it is very helpful. We followed it to reorganise the paragraphs. As you suggested, the description of Galaxy that was in the first section was moved to the beginning of the Galaxy section. With all the alterations you proposed, it seemed to us that the description of the platform was enough for first-users. As the Galaxy initiative exists since 2009, many papers cited in this article give many details on the platform. In addition, as mentioned on line 370, the GTN provides many tutorials explaining how to engage with the platform for many publics.

Regarding the examples of its success, it is briefly mentioned on line 380 with two references that gives more specific examples. Additionally, the Galaxy community (2022) publication is updated every year and gives many examples of successes and new developments.

I think improving the structure will also reduce some of the redundancy of ideas and language throughout the section. For example, the paragraph starting at line 403 is a repetition of the general philosophy of Galaxy-E that echoes a similar sentiment given many times throughout the paper, I don't think It's necessary here.

Thank you, this paragraph has been shortened.

Similarly, the discussion repeats many of the basic concepts of reproducibility or atomisation/generalisation without linking them back to the platform. The first two paragraphs in particular could either be cut, or should be edited to be directly relevant to the platform and its strengths. The discussion could also benefit from a more detailed comparison to other existing platforms.

Thank you, we edited these paragraphs according to your detailed feedback and this particular comment.

Detailed feedback:

Line 227-243: This description feels out of place here, as it's followed by a more detailed description of the atomisation/generalisation framework rather than the Galaxy-E approach, I'd move it to the "Entering a new dimension" section

Done

Line 342: How or why might that be true?

As responded to reviewer #1, In the article by Gomes et al. (2022), they suggest sharing codes with trusted peers for review before submission to overcome the embarrassment/insecurity barrier. As trusted peers (or even the author of the code) can read and re-use codes more easily, the author can be better reassured when validation is provided for their work. This explanation has been included at the end of the paragraph.

Line 346-360: This feels redundant, it could be integrated more concisely into the earlier paragraphs.

The bullet point list has been removed, it still seemed crucial to posit the complementarity of atomisation and generalisation as highlighted by reviewer #1 as well.

Line 376: This feels redundant after the previous paragraph

Thank you, the sentence has been shortened.

Line 501: I don't think this bullet list is appropriate for the discussion, maybe a summarized version in the reproducibility section, but as it stands it's relevance to the Galaxy-E tool is not clear.

Thank you, we altered the bullet list to these sentences "The levels of application of these best practices fall within a continuum offering a range of possibilities from the sole sharing of processed and interpreted results with a brief description of methods to an executable paper published within a container and emulated virtual machine (Strijkers et al., 2011; Grüning et al., 2018). Situated somewhere in between the aforementioned extremes, the atomisation – generalisation framework and the utilisation of the Galaxy platform might represent viable solutions offering a satisfactory level of best practices."

Line 527-535: These things are all true, but what are their relevance for Galaxy-E? How does it help a user achieve this?

In this paragraph, we highlight limitations of the atomisation-generalisation framework which are partially addressed by Galaxy as explained in a summarised manner in the following paragraph. The tools are versioned, all the analytical procedure is tracked in the Galaxy history, at the end of the procedure one can extract a workflow that can be shared, visualised and explored by anyone etc... as it is explained in the section on Galaxy and is synthesised in table 2. Sorry, we are not sure what should be specified with this paragraph, we added some alterations that hopefully will clarify the overall message of these sentences.

Line 530: I don't think arborescence is the right word here, are you trying to describe the relationships of analysis pieces to one another?

Sorry, it was a wrong translation, it was simply referring to file/folder structure (or tree).