Software management plans in research software

Shoaib Sufi, Community Lead, Software Sustainability Institute
shoaib.sufi@software.ac.uk

Supported by EP/H043160/1 & EP/N006410/1 & EP/S021779/1
The Software Sustainability Institute

• A national facility for cultivating world-class research through software

• “Better Software, Better Research”

• Software code/processes/community reaches boundaries in its development that prevent improvement, growth and adoption

• Providing the expertise and services needed to negotiate to the next stage

• Programmes, events, policy and tools to support the community developing and using research software

• We advocate for all things Research Software
Teams

Software
Helping the community to develop software that meets the needs of reliable, reproducible, and reusable research

Policy
Collecting evidence on and promoting the place of software in research & sharing with stakeholders

Outreach
Exploiting our platform to enable engagement, delivery & uptake

Training
Delivering essential software skills to researchers, partnering with institutions, doctoral schools and the community

Community
Developing Communities of Practice by supporting the right people to understand and address topical issues

Activities

Software
70+ project consultancies
200+ evaluations
4 surgeries

Policy
650+ RSEs engaged
2k signatures
13 issues highlighted

Outreach
170+ external contributors
20k unique visitors/month
6.5k followers (Twitter)

Training
200+ Carpentry workshops
5000+ learners, 190 instructors
80+ guides

Community
129 Fellows
25+ workshops organised
Types of Management Plans

- Data Management Plans (DMP)
  - Where it all began - stating how you will manage the data produced on your project
- Software Management Plans (SMP)
  - How you will manage the software produced on your project
- Output Management Plans (OMP)
  - How you will manage data, software and other resources produced on your project
The idea of managing software outcomes is not new

- Defense System Software Management Plan - 1976
  - Importance & costs of software
    - Acquisition, development, maintenance guidance
  - [https://apps.dtic.mil/docs/citations/ADA022558](https://apps.dtic.mil/docs/citations/ADA022558)
- Advanced Composition Explorer Project - 1994
  - Contract, QA, Responsibilities, Scope
- Idea of Software Management Plans not necessarily new but also meant slightly different things to different people
Software Management Plan (SMP)

What is it
• A statement of intent around how you will manage your research software

Why do we need it
• The normal research process can squeeze out time/thinking for the proper management of software
  ▪ Collaboration
  ▪ Papers
  ▪ Proposals
  ▪ Meetings/conferences
SMP: What does it contain

• Describe what the software does / problem it solves
• Who the software is for (even if it’s just for you)
• How you will make your software available
• How it will help you / other users
• How you will assess how it’s helped you / others
• The level of support you are willing to offer
• How the software fits into the broader ecosystem of software in the problem space (e.g. what does it add)
• How you intend to make your software available beyond the life of the project
SMP: Guidance

• A checklist is available produced by the UK Software Sustainability Institute
  ▪ Caveats
  • Not all questions relevant for all projects
  • Depends on nature of research software
  • Depend on state of development
• Checklist for a Software Management Plan - https://zenodo.org/record/2159713
  ▪ (pdf, docx, md, odt)
SMP: Checklist Questions

• What Software will you develop? - *greenfield, trademarks*
• Who are the intended users of your software? - *skill level, extensibility*
• How will you make your software available to your users? - *licensing, containers*
• How will you support those who use your software? - *setting expectations, issues, forums*
• How will your software contribute to research? - e.g. *novelty, speed, accuracy, ease*
• How will your software relate to other research objects? - *relating, FAIR Digital Objects, www.researchobject.org*
• How will you measure your software’s contribution to research? - citation, surveys (remember privacy)
• Where will you deposit your software to guarantee its long-term availability? - digital repositories (e.g. GitHub-Zenodo integration), identifiers, longevity
Questions?
SMP: when to write one

• Normally at the start of a project
  ▪ Been Mandated in some (UK) funding calls:
    • EPSRC High End Computing (HEC) Consortia Call 2017 -
    • Computational Science and Engineering: Software for the Future II -
      https://epsrc.ukri.org/files/funding/calls/2014/computational-science-
      and-engineering-software-for-the-future-ii/
    • Both cases as part of ‘Pathways to Impact’ set of document
      ▪ “is primarily for detailing the activities that will increase the likelihood of
        potential economic and societal impacts being achieved.”
SMP: when to write one ... 2

• During a running project
  ▪ If you have not made one at the beginning
  ▪ As a review or audit of software assets
  ▪ Once known a more consistent approach could be taken
    • e.g. putting them under an organisation on GitHub rather than individual repos tied to staff
    • Consistency of documentation / licenses
    • A consistent approach towards credit
SMP: Living documents

- Software Management Plan are living documents
  - The first one is your baseline
  - They should be revisited e.g. every 3-6 months or annually
  - They should be versioned
  - Project lead will ultimately be responsible for making sure they are implemented / used
- Software Evaluation can help!
Software Evaluation

• Software Management Plans are an intention around what you PLAN to do
• Software Evaluation is about what you ARE DOING.
• Software Evaluation can thus help with delivering your SMP and adjusting it where necessary
  ▪ Assess code quality
  ▪ Usability
  ▪ Overall Sustainability
Software Evaluation Approaches

https://www.software.ac.uk/resources/guides-everything/software-evaluation-guide

• Criteria-based (https://software.ac.uk/sites/default/files/SSI-SoftwareEvaluationCriteria.pdf)
  ▪ Quantitative assessment:
    • Sustainability
    • Maintainability
    • Usability
  ▪ Can inform high-level decisions on specific area for software improvement
  ▪ Basis of Online Sustainability Evaluation (OES) -
    https://www.software.ac.uk/resources/online-sustainability-evaluation
  ▪ If you don’t have an SMP the OES can be a great way to bootstrap one
• Tutorial-based (https://software.ac.uk/sites/default/files/SSI-SoftwareEvaluationTutorial.pdf)
  ▪ User/developer subjective experience of
    • learning
    • building
    • installing
    • configuring
  ▪ Outcome: a practical guide for getting the software to work in the way it should more about Quality than Planning (e.g. SMP)
• In Either case (i.e. Criteria or Tutorial) judgement needed about what to include based on type of software, environment (e.g. open development) and personas of those doing assessment.
The Recently (2019-08-08!) released CHA OSS Metrics are another way you could assess your software. They are potentially more suited to larger pieces of software which have started to have a community. The Guidance is here:

- [https://chaoss.community/metrics/](https://chaoss.community/metrics/)

- Working groups with Focus Areas
  - Common Metrics
    - Organizational Affiliation
  - Diversity and Inclusion
    - Event Diversity
    - Governance
    - Leadership
  - Evolution
    - Code Development
  - Risk
    - Business Risk
    - Code Quality
    - Licensing
    - Transparency
  - Value
    - Labor Investment
    - Living Wage
  - Goal-Question-Metric format

SSI participating via Director, Neil Chue Hong
Questions?
SMP: modern examples

- Laurent Gatto (SSI Fellow), Open Science advocate and Group Leader at Du Duve Institute, Belgium
- Difference between DMP & SMP and the proliferation of plans and the need for Output Management Plans - https://lgatto.github.io/output-management

<table>
<thead>
<tr>
<th>Data</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Dynamic</td>
</tr>
<tr>
<td>Meta-data</td>
<td>Documentation</td>
</tr>
<tr>
<td>Large</td>
<td>Small</td>
</tr>
<tr>
<td>QC</td>
<td>Testing</td>
</tr>
<tr>
<td>DMP</td>
<td>SMP</td>
</tr>
<tr>
<td>DB, public repo</td>
<td>Version control in public repo</td>
</tr>
</tbody>
</table>
SMP: Laurent Gatto example

- An example of one from a funded project
- An Output Management Plan
  - Combine Software, Data and Materials
- Original wanted a DMP but included information about software
  - Dissemination - *BioConductor*
  - Release schedule
  - Development - *GitHub*
  - Licensing
  - Documentation
  - Reproducibility framework
- Not all things nailed down but a strong intention - it’s a living document or should be.

https://riojournal.com/articles.php?id=11624
SMP: private example

- SMP are part of grant proposals - usually (in the UK) part of Pathways to Impact
- So they are **not public**
  - Ideally there would be a repo of good ones event if redacted
  - SMPs check to see if the right things are being done, not new things, hence the could well be public.
- Software Management Features:
  - Training
  - Hackdays
  - Documentation
  - GitHub
  - git, issues, wiki
  - Continuous Integration
    - Unit tests
    - Hudson CI
  - Extensions
    - Examples of use
Criticism of SMP’s

- It sounds like a proposal
- It’s not specific enough
- Basic information is missing
- Size and type of data is missing
- Programming language used is missing
- Necessary infrastructure is missing!
- Exact licenses what are they!
- What’s your preservation duration!
- Who are the people responsible?!
- It’s easy to read but useless to me
- It’s about principle but it’s not machine actionable
- If this is a draft of the planning phase it’s ok - but we want a living doc
- The more advanced the research the more information is needed!

Is it a wonder only the brave would share their SMP/DMP with comments like these!

Perfection is the enemy of the good (enough)

A more constructive approach:

- Equally applicable to DMP’s
- SMPs are self regulated (a good thing - otherwise overbearing?)
  - Maybe useful for project which are contracts e.g. EU H2020
  - valid thing to ask at a review
    - How are you keeping your SMP/DMP up to date?
    - How are you evaluating that you are following your SMP/DMP?

Imposter syndrome or trolls should not stop you from trying to improve and adopt better practices!
Questions?
SMP: Funders perspective

- Wellcome Trust - [https://wellcome.ac.uk/funding/guidance/how-complete-outputs-management-plan](https://wellcome.ac.uk/funding/guidance/how-complete-outputs-management-plan)
  - Output Management Plan
    - Data and Software
    - Research Materials
    - Intellectual Property
    - Resources required for the above
  - Wellcome exists to improve health by helping great ideas to thrive, they are a politically and financially independent foundation in the UK that plan to spend £5B over the next 5 years.
  - In 2018 it was the 4th wealthiest charitable foundation in the world.

Examples of applications that require an outputs management plan:
- studies producing whole genome/exome sequence data, whole genome genotype or other omics datasets generated at scale
- genome-wide or large-scale functional genomic studies in a specific organism
- longitudinal studies of patient and population cohorts
- clinical trials
- large-scale neuro-imaging studies
- development of viewers and annotation tools that allow visualisation and analysis of DNA, cells and other biological components
- computational models and simulations of neurological, physiological or other biological systems
- creation or development of a database, materials collection or other research resource.
The NIH in the US have many requirements around data sharing (effective 2003)


- Software Management has been mentioned in some calls.

"Are the data and software management and sharing plans adequate to make these resources available within the LINCS consortium and to the larger research community?"


"As appropriate, applicants should also describe data and software management and provenance, software development and testing practices, software toolkit development and deployment, application programming interfaces (APIs), and human subject data privacy and security protections."

Development of a Knowledge Management Center for Illuminating the Druggable Genome (US4), RFA-RM-13-011, 2013


NIH default Resource Sharing Plan mentions Data and not software (2018)

[https://instr.iastate.libguides.com/dmp/NIH](https://instr.iastate.libguides.com/dmp/NIH)

Software is not regarded as "data" but it is recognized that access to software and other tools may be necessary to to access and interpret the data (i.e. they may need to be covered in your plan).
Institutional Perspective

- Data Management Plans
  - Cambridge: [https://www.data.cam.ac.uk/](https://www.data.cam.ac.uk/)

- Software Management Plans
    - A note on commercialisation
    - Pointing to SSI advice
  - York: [https://www.york.ac.uk/library/info-for/researchers/data/planning/](https://www.york.ac.uk/library/info-for/researchers/data/planning/)
    - mainly about DMP’s but mentions the SSI SMP template at DMPOnline
  - UCL: [https://blogs.ucl.ac.uk/rdm/tags/sustainable-software/](https://blogs.ucl.ac.uk/rdm/tags/sustainable-software/)
    - SMPs in the context of software preservation
  - STFC: [https://edata.stfc.ac.uk/page/policy](https://edata.stfc.ac.uk/page/policy)
    - Infrastructure for supporting actions (e.g. deposition) of DMP and SMP
Advocacy for SMP

- Making Software a First-Class Citizen in Research - [https://software.ac.uk/blog/2018-11-28-making-software-first-class-citizen-research](https://software.ac.uk/blog/2018-11-28-making-software-first-class-citizen-research)
  - [http://wssspe.researchcomputing.org.uk/wssspe6-1/](http://wssspe.researchcomputing.org.uk/wssspe6-1/)
  - Recognition of research software is lagging research data
  - The need for culture change around software credit (applies to data also)
  - Some recommendations:
    - 4OSS recommendations - [https://softdev4research.github.io/recommendations/](https://softdev4research.github.io/recommendations/)
    - SSI guidance - [https://software.ac.uk/blog/2018-11-28-making-software-first-class-citizen-research](https://software.ac.uk/blog/2018-11-28-making-software-first-class-citizen-research)
    - Netherlands eScience Centre Guide - [https://guide.esciencecenter.nl/](https://guide.esciencecenter.nl/)
  - Raised with the Head of the Netherlands funding agency at the eScience 2018 conference in Amsterdam at an open Q&A

<table>
<thead>
<tr>
<th>Type of Support</th>
<th>Data</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutional</td>
<td>Research Data Management</td>
<td>Research Software Engineer</td>
</tr>
<tr>
<td>Plans</td>
<td>Data Management Plans</td>
<td>Software Management Plans</td>
</tr>
<tr>
<td>FAIR</td>
<td>FAIR Data</td>
<td>FAIR Software?</td>
</tr>
<tr>
<td>Open</td>
<td>Open Data</td>
<td>Open Source?</td>
</tr>
</tbody>
</table>
Python tool for SMP

- [https://github.com/softwaresaved/software-management-plans](https://github.com/softwaresaved/software-management-plans)
  - YAML file
    - Single source of ‘truth’
    - Questions
    - Full guidance
  - Python script
    - Produce full guidance doc (e.g. in .docx or .odt)
    - Produce a checklist
  - Example of use
    - Used to create the v 1.0 docs in Zenodo
  - Intended use
    - Adapted by service providers (e.g. DMPonline or DMPtool)
    - Adapted by institutions / funders / domains
  - Possible extensions
    - Push templates to Google Docs
    - Create GitHub issues
    - Push to service providers via their API
Tools for making DMP/SMP

• UK main site:
  ▪ [https://dmponline.dcc.ac.uk](https://dmponline.dcc.ac.uk)
  ▪ SMP’s being made available on DMPonline

• US main site:
  ▪ [https://dmptool.org/](https://dmptool.org/)
    • Does not serve SMP’s specifically at this time

• ‘Golden’ examples
  ▪ Does not exist as SMP often bundled in a private way
  ▪ Good examples exist for research data / DMPs - useful by analogy
  ▪ LIBER DMP Catalogue useful in this case - [https://zenodo.org/communities/liber-dmp-cat](https://zenodo.org/communities/liber-dmp-cat)
    • eight examples, various disciplines, reviews and highlights

• Future
  ▪ ‘Machine Actionability’
Take Home Messages

• Data Management Plans are the most common type of plan
• Software Management Plans are starting to become more common/mandated
• There are moves to have combined plans in the form of - Output Management Plan
• The is an interplay between funders, institutions, tool/guidance providers and researchers
• Sustained culture change needed to move towards SMP/OMP
• Something is better than nothing ... don’t fear the trolls
• SMPs should be living documents
• Software Evaluation can help keep SMPs fresh or even bootstrap them
• You can comment on the SSI SMP guidance guidance via the GitHub project
• There is a need for an open repository of SMP’s to help the community formulate their own - due to privacy/permission issues this does not exist yet cf. DMPs
• Primary benefit of this if for you!
Acknowledgements

- Mike Jackson, EPCC, Edinburgh, UK

- Neil Chue Hong, Director Software Sustainability Institute

- Sarah Jones Associate Director of Digital Curation Centre

- Giacomo Peru, Project Officer, Software Sustainability Institute

- Marta Ribeiro, former staffer at the Digital Curation Centre

Supported by the UK Research Councils through grants EP/H043160/1, EP/N006410/1 and EP/S021779/1. Additional project funding received from Jisc.
Questions