Linux Foundation Legal Summit 2023

Location: The Marker Hotel

501 Geary Street San Francisco, CA 94102

Day One – February 14	Time
Breakfast	Starts at 8:00am
Intro & Welcome	9:15 - 9:30
ML models are the new open source projects	9:35 - 10:00
What is GPL compliance for an ML Model	10:05 - 10:30
Ethics, benefits, and drawbacks of ML	10:35 -11:20
Regulation: what tools beyond licensing are in play?	11:30 - 12:00pm
LUNCH	12:00 - 1:00
Deep Dive: the RAIL license(s)	1:00 - 1:55
What is “open” in this context?	2:00 - 2:55
Data governance	3:00 - 3:55
CRA and its impact on open source	4:00 - 4:55
Concluding Q&A and Discussion	
Bus to Dinner	6:15
Dinner - International Smoke (301 Mission St, San Francisco, CA 94105)	6:30

Day Two – February 15	Time
Breakfast	Starts at 8:00am
Welcome to Day 2	8:45- 9:00
LF Research	9:00 - 9:45
Short Break	9:45 - 10:00
GPL-3.0 Proposal for Automotive	10:00 - 11:00
Litigation Update - Co-Pilot, Vizio, Stability AI	11:00 - 12:00pm
Lunch	12:00 - 1:00
AOM Panel	1:00 - 1:45
Open Source Security at IBM	1:45 - 2:30
Open Source Security at Ericsson	2:30- 2:40
Litigation Update – Tulip Trading	2:40 - 3:15
Trademarks in the Wild	3:15 - 3:30
Short Break	3:30- 3:45
SPDX 3.0	3:45 - 4:15
OpenChain Licensing & Security process standards	4:15 - 4:30
Evolving Code of Conduct	4:30 - 5:00
Concluding Q&A and Discussion	

Tuesday February 14, 2023

Please understand that we tried to take summary notes, but may have captured points incompletely or even incorrectly. We apologize in advance for any errors. If you have any changes you would like to suggest, please do so by emailing legal@linuxfoundation.org.

Legal Summit operates under Chatham House Rule,

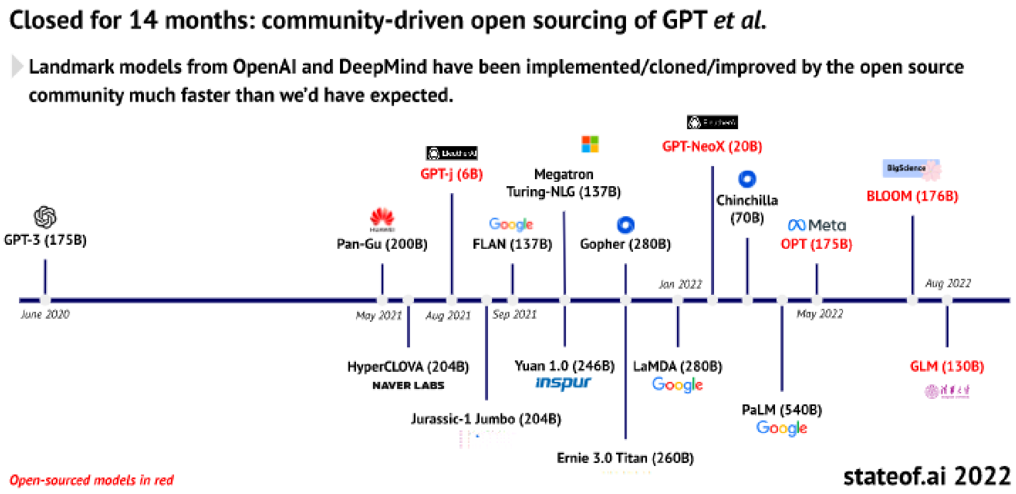
> "When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed."[1]

ML models are the new open source projects

The economic, political social, legal, and technical considerations of open source can apply to new areas such as machine learning (ML).

Generative AI gets a lot of focus as it generates new text, images, etc. but analytical AI is useful for being able to identify things or make decisions.

The models used can be thought of as the new libraries for applications. Huggingface is an ML model sharing platform, which has exponentially grown the number of shared models in the past year. Many in the room have had their internal clients asking about whether they can use models, including models shared under an open source license. A number of models have been released under "open-ish" licenses that wouldn't likely meet the OSI definition.



Closed for 14 months: community-driven open sourcing of GPT *et al.*

Landmark models from OpenAI and DeepMind have been implemented/cloned/improved by the open source community much faster than we'd have expected.

Source: https://www.stateof.ai/

A few topics have come up that warrant additional scrutiny:

- What does it mean to have a model released under a GPL license?

[1] https://en.wikipedia.org/wiki/Chatham_House_Rule

    - What is the “preferred form of making modifications to the work”?
    - Looking at the OSD #2: “The source code must be the preferred form in which a programmer would modify the program”
- How does “ethical source” impact licensing?
    - Are the key models forcing organizations to adapt to allowing a new license model?
- What does open mean for a model?
    - Goals of open: frictionless improvement, autonomy, education, community, transparency
    - What about the underlying data used to create the model?

Data in “open” ML models can be discussed along a transparency spectrum

- No data info provided
- Data source described
- Partial data available
- Data public
- Data under an open license

Code is the easy part. There are reports highlighting the implications of use of generative AI creating code in the “style of [developer name]”, which may generate more or less secure code depending on who is selected. Challenge is how account for this.

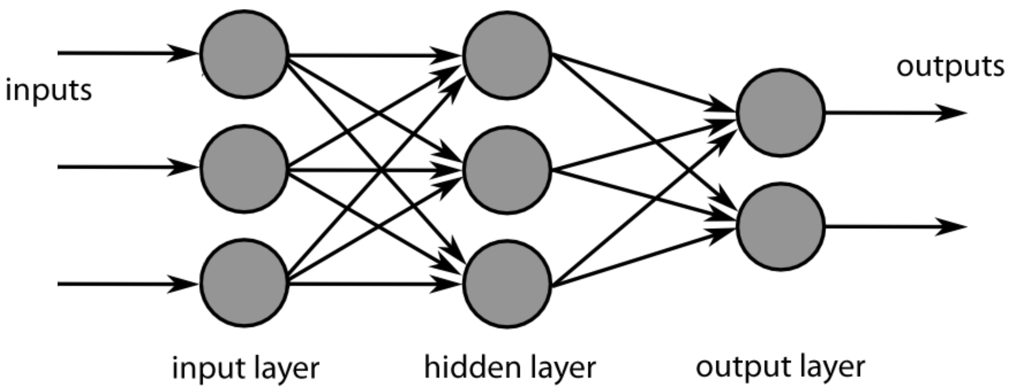
What worked for government data releases in history, doesn’t necessary apply to combined data today. Training is more costly than compiling in some cases, as is access. Nominally CC-BY-SA 3.0, combined with public domain, quickly gets messy. Data is too complicated to have a purist approach.

Discussion regarding whether there is an opportunity around community governance on data sets.

When we start talking about product liability and the potential requirement to be able to update the data that was used to train a model. Using 15 year old Java code may be fine for a bank, but using 15 year old datasets may not be ok for a model.

Large collaborative datasets are where there is an open source approach to curating, labeling, and building data together.

What is a ML model?



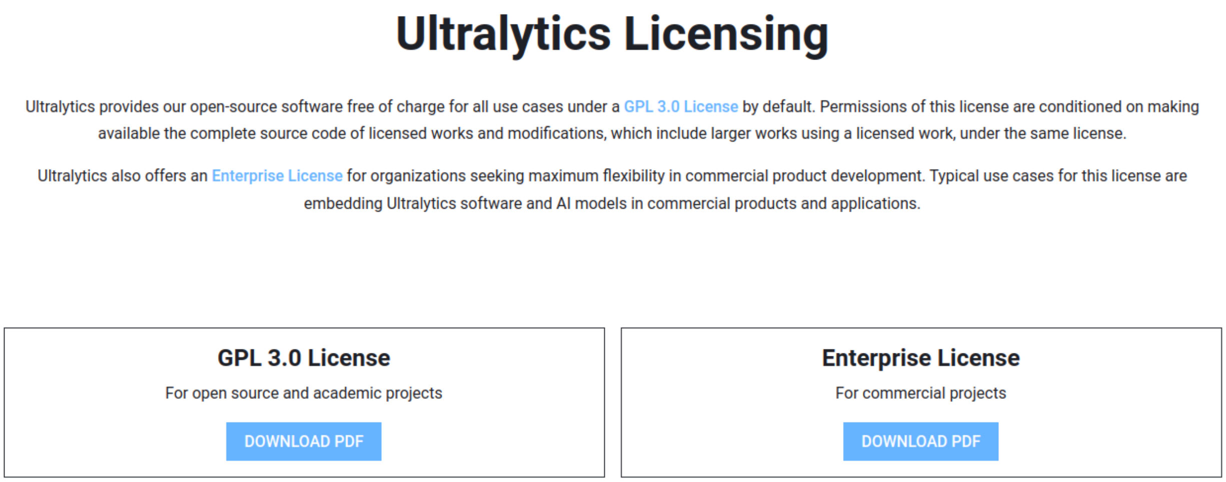
Source: https://commons.wikimedia.org/wiki/File:MultiLayerNeuralNetworkBigger_english.png

Training data is an input to the training process. Code defines the neurons (hidden layer) and the "shape" of the network. Code and software is used for training the network (e.g. PyTorch). The outputs of the process are weights (between 0 and 1) and parameters. The weights and parameters can be thought of as probabilities. There is software for running the network (e.g. ONNX, PyTorch).

ML Models have data, code (training, inference), and weights and parameters (which can be further trained). Weights, parameters and runtime are increasingly shipped as one artifact.

What is GPL compliance for an ML Model

YOLOv5 is an example of a dual license ML model published by a company named Ultralytics using dual licensing under GPL-3.0 and a commercial license.



Ultralytics Licensing

Ultralytics provides our open-source software free of charge for all use cases under a GPL 3.0 License by default. Permissions of this license are conditioned on making available the complete source code of licensed works and modifications, which include larger works using a licensed work, under the same license.

Ultralytics also offers an Enterprise License for organizations seeking maximum flexibility in commercial product development. Typical use cases for this license are embedding Ultralytics software and AI models in commercial products and applications.

GPL 3.0 License

For open source and academic projects

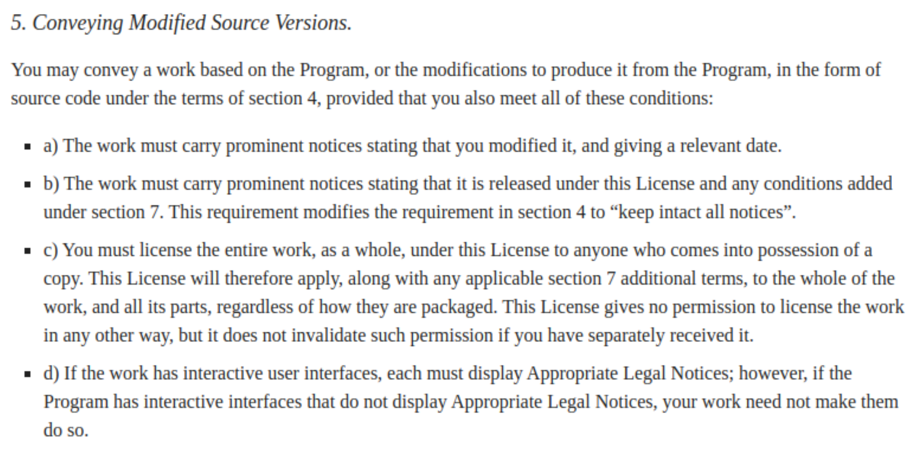
Enterprise License

For commercial projects

Source: https://ultralytics.com/license

GPL-3.0 §1 defines “source code” as “The “source code” for a work means the preferred form of the work for making modifications to it. “Object code” means any non-source form of a work.”[2]

GPL-3.0 §5 permits conveying modified source, but what is “source” for a model?



5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

- a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to “keep intact all notices”.
- c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

The code defining the network shape is generally fairly simple code. There is a possible analogy to training a model and compiling code as an iterative process. In software you start with a copyrightable work (source code) that can be outputed as a functional manifestation of that copyrighted work (object code). In the context of an ML model, the runtimes and libraries that exist are likely all copyrightable . The weights and parameters, however, are not a translation of anything in source, and there is no “object code” in the context of an ML model. The intellectual property that would be licensed under the GPL-3.0 is unclear.

The model can be a product of copyrighted material in the training data. A model can be a matrix of weights and parameters, but the model weights might be a functional encoding of the data it was trained on.

What would be needed to modify the source of a model? Some thoughts:

- The network shape is probably important, e.g. YAML files that define the shape of the network
- The training algorithms would be important to being able to modify the model

[If you setup an account on Huggingface, you are required to choose a license label.]

GPL-3.0 and Corresponding Source

- The checkpoint files (often referred to as the “model”) are generated and considered valuable but are not object code and are able to be inspected.
  - A checkpoint file is a point in time weights and parameters matrix
- It’s unclear what would be an object file

[2] https://www.gnu.org/licenses/gpl-3.0.en.html

If we’re basing licesning on copyright, and much of the ML model is not copyrightable, what is the right framework for protecting a model? Applying a copyright license to models may not work as intended.

Ethics, benefits, and drawbacks of ML

We hosted a debate on the ethical considerations of AI/ML models. One question that arises is how creators can have more control over their work and the acceptable norms of developing models. There are ideas about opt-in and opt-out, but neither gives the sense of control or credit that creators expect to have. Writers often describe it as industrial-scale plagiarism.[3]

Open source licenses enable collaboration without asking for permission. There are technological evolutions enabled and it’s hard to go back to the old way.

The key question is what is a fair value of exchange between the labor put into training models and those who use and consume the results of that labor.

Developers and governments everywhere have benefitted from open source. There is a ‘hyperscalers are exploiting us’ theme but people all over the world will benefit from models. An example raised is AlphaFold[4] - a model that “predicts a protein’s 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment.” AlphaFold saves years of research work.

We could see implications spill over from the art world, such as the Warhol[5] litigation where the United States Supreme Court will consider “whether a famous set of images that Andy Warhol based on a 1981 photograph of Prince by the award-winning photographer Lynn Goldsmith were such a “fair” use of the photograph that Warhol’s successors can license them for commercial use without the permission of (or compensation to) Goldsmith.”[6]

There are concerns with generative AI outputs being low quality and then used to train generative AI models.

Regulation: what tools beyond licensing are in play?

“Race to the bottom/top” from a licensing perspective, if there is economic incentive then people will release a permissive version. The range of regulatory instruments need to be considered

[3] https://www.wired.com/story/chatgpt-generative-artificial-intelligence-regulation/

[4] https://alphafold.ebi.ac.uk/

[5] https://www.scotusblog.com/case-files/cases/andy-warhol-foundation-for-the-visual-arts-inc-v-goldsmith/

[6] https://www.scotusblog.com/2022/10/justices-debate-whether-warhol-image-is-fair-use-of-photograph-of-prince/

from norms and voluntary practices, though this can be looked at as reputation management. State self-regulation will be an important aspect, as one of the most dangerous users of AI is the state itself. Then there is straight regulation such as the AI Act[7], banned uses, etc.

The Act[8] has Chapter 5, Article 40, when you can distribute on an open basis a general system, and you have to place restrictions on high risk scenarios (e.g., using the system to manage electrical systems). Responsible AI licenses might be one solution. The EU Commission is going to recommend model templates for stakeholders to use as data sharing agreements and it is anticipated that the EU Act[9] will include templates. It is envisioned that there will be use restrictions to be placed within AI systems.

Discussion that "open" has a seat at the table in the regulatory discussion. One concern is that European policy makers still don't understand what open source is. There should be exceptions placed within the EU Act to promote open source in AI.

Backing up, discussion of enabling AI through approaches outside of licensing, e.g., examples of codes of conduct in open source and transparency cards. Tools such as these are great ways for knowledge diffusion. Think of it as a 'nudge' or soft regulation. The role of transparency, model and/or data cards inform to use, though not all data cards are quality data cards. If you take a high quality "model card"[10], a lot of the information in the card is a core requirement in the EU Act for high risk systems. These are community mechanisms that are being steered towards regulation. The role of license-style meta data.

View of model cards being a why to get ahead of potentially dysfunctional regulation. Discussion of whether licensing has a signaling role.

DIscussion of US vs European approach to regulation and litigation. Regulation can move faster, and so Europe will probably lead in deciding what happens, and the community can play an important role in informing the European regulatory outcomes.

Licenses as been a static way of development but the AI environment requires a faster level of response.

Hearing criticism from the open software community about the responsible AI license, and yet are not seeing another open license. There have been no proposals about what an open ML license would look like. There was a response to this, a discussion about whether the current licenses can work for models as a whole with concerns that GPL does not.

Discussion that US provided more certainty for innovators early on with a more friendly regulatory environment, and now that this might be viewed as a bad thing. The "Brussels goal is

[7] https://artificialintelligenceact.eu/
[8] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
[9] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
[10] https://huggingface.co/blog/model-cards

to ‘shape the future of regulation not to innovate.” That type of international competition will be dynamic.

Deep Dive: the RAIL license(s)

The BigScience community building large language models looked at what others were doing and saw many licenses with restrictions to research only or other use restrictions. The BigScience community started assuming they would use a solely for research license, or a fully permissive license. There were some concerns raised by potential users of the large language model they were building, and started thinking about what the ideal license would look like.

Wanted to have a model license, which Apache-2.0 and others would not allow for. They also wanted to enable use restrictions. Once they released the license, they realized there will be a transformation from a license to a regulatory framework with the EU AI Act which defines the high risk scenarios which require use restrictions. RAIL started with defining community norms and now will be a regulatory compliance instrument.

3 licenses[11] were published in 2022 as the Responsible AI Licenses (RAIL): https://www.licenses.ai/ai-licenses

- End-user License “RAIL-A License”
- Source Code License “Open RAIL-S License”
- Model License “Open RAIL-M License”

One challenge has been amongst the engineering community there’s no clear standard definition of a “model”.

RAIL-M has a large footprint on a number of community-driven projects. A few of the key definitions include:

- (d) “Model” means any accompanying machine-learning based assemblies (including checkpoints), consisting of learnt weights, parameters (including optimizer states), corresponding to the model architecture as embodied in the Complementary Material, that have been trained or tuned, in whole or in part on the Data, using the Complementary Material.
- (f) “Complementary Material” means the accompanying source code and scripts used to define, run, load, benchmark or evaluate the Model, and used to prepare data for training or evaluation, if any. This includes any accompanying documentation, tutorials, examples, etc, if any.
- (e) “Derivatives of the Model” means all modifications to the Model, works based on the Model, or any other model which is created or initialized by transfer of patterns of the weights, parameters, activations or output of the Model, to the other model, in order to cause the other model to perform similarly to the Model, including - but not limited to -

[11] https://www.licenses.ai/ai-licenses

> distillation methods entailing the use of intermediate data representations or methods based on the generation of synthetic data by the Model for training the other model.

Derivatives of the model – how to draft a definition serving the goals of the license regarding our values as a community. We wanted an expansive / aggressive definition of a derivative. For example to include ‘fine tuning’, or taking the weights of one model and compressing them (distillation).

The intent was the model (and use restrictions on the model) only applies to the weights. Everything else is treated as Complementary Material. Distillation refers to refinement of the weights, fine tuning, compression, etc. The intention was to ensure the users could have all the rights they need. BigScience took inspiration from the Apache-2.0 license and oriented towards permissive licensing. Some discussion opened up around whether under 1(e) there was any protectable IP rights in the output of a model if a party has no rights on the inputs to the model.

The intention was to require compliance with the license while *using* the model, but the output wouldn’t have any restrictions. There is no claim to any IP rights extending to an Output.

> 6. The Output You Generate. Except as set forth herein, Licensor claims no rights in the Output You generate using the Model. You are accountable for the Output you generate and its subsequent uses. No use of the output can contravene any provision as stated in the License.

Use restrictions are required to be passed to downstream recipients in §4(a), but are not required to be enforced. The use restrictions do not apply to the Complementary Material.

The community saw distribution inclusive of as-a-service or API delivery.

> (g) “Distribution” means any transmission, reproduction, publication or other sharing of the Model or Derivatives of the Model to a third party, including providing the Model as a hosted service made available by electronic or other remote means - e.g. API-based or web access.

Contribution’s definition hinges on “any work of authorship” which may . Within BigScience they’re discussing whether to even define Contribution at all.

Requirement to undertake reasonable efforts to use the latest version of a model.

> 7. Updates and Runtime Restrictions. To the maximum extent permitted by law, Licensor reserves the right to restrict (remotely or otherwise) usage of the Model in violation of this License, update the Model through electronic means, or modify the Output of the Model based on updates. You shall undertake reasonable efforts to use the latest version of the Model.

RAIL community recently issued a call for input to responsible AI licensing. The community is developing a license for datasets, RAIL-D. RAIL is seeing more interest from public institutions vs private industry stakeholders. Governments have their own software teams, research labs, open source offices who are producing software and machine learning models and are looking for license options.

What is “open” in this context?

“Open” shouldn’t require permissions and roadblocks to prevent your use or frustrate your modification. If you receive a pie that has nuts, but you have an allergy to nuts, for it to be “open”, you need the recipe and instructions to make a new pie without the nuts.

ML systems are dynamic and not the same as software. They are not static and can relearn all the time. We may need new definitions. When promoting “The right to read is the right to mine”[12] there wasn’t an anticipation of AI.

One view of what’s needed:

1. Accessibility of datasets, which is a range of access based on interaction with privacy and other laws
2. Learn the social norms that have allowed for open source to flourish, aware that there is this fear in the AI community of doing harm; they want to have ways of controlling deployment of these systems
3. Ways to help policymakers, the regulations are coming really fast. Fast moving regulation can have damaging side effects.

Model cards disclosing minimum information about the source of data could help with transparency when you cannot share the data.

One challenge is while core concepts of copyright law are global, data privacy and other regulations are most definitely not, so a globally usable “license” is more challenging.

So much of the acceptance criteria for ML/AI will depend on trust and transparency of the data.

Question on the security component, discussion that there is more focus on the security elements of open source, also tied to AI/ML Example of DARPA interested in whether models have been tampered with. Innovation is moving very fast and we need to help the regulators understand the environment.

Takeaways about how fear has become part of the technology narrative and driven a change in the conversation from ‘open sharing’ to ‘better sharing’.

[12] https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/

Data Governance

Discussion that “licensing is not governance.”

The panel discussed the need for collaboration on data, including the need to synthesize multiple sources, licensing, and make decisions. Open Streem Map data is released under the ODbL.[13] Other sources of data from outside OSM with relevant data to mapping were not acceptable to upstream back to OSM because those data sources were not ODbL-licensed.

ODbL sharealike provisions can be difficult to implement and resulted in minute parsing. Overture was not going to require the share-back. The assumption is the best data is used and feedback is provided upon. This is not license mandated.

Overture uses either / or in order to account for the future, for example layers of the map that don’t exist yet. We settled upon CDLA version 2 b/c it is listed as compatible by OpenStreetMaps and was acceptable to the companies that could address layers of the map that were not addressed by OSM.

Question where national legal systems forced projects to do something, and as an example was quotations and copyright of quotations. International disputes on borders mentioned.

Governance discussion on Wikimedia regarding arbitration committees such as the ‘meta oversight board’ which deals with disputes on content on the projects, other aspects of governance deal with permissions.

The Overture Maps Legal committee discusses data governance items around disputes. Overture deals not just with baseline map data but additional data with potential global complexities. If you think about places data. E.g. The Marker Hotel, there’s an address, a building footprint, there’s a restaurant, restrooms. If you have a sharealike license, like ODbL, the license doesn’t tell you and you’re left with IANAL (I Am Not A Lawyer) debates in Internet forums.

Humanitarian OpenStreeMap Team[14] (“HOT OSM”) helps in disaster situations to bring real time changes to map data and is a use case that was not originally imagined. HOT OSM is being used in the Turkey earthquake recovery.

OSM has traditionally believed in the “ground truth” where someone walks a path and maps out what is there, similar to a historical survey. There are possibilities now to fill in the gaps with AI/ML generated map data. That data can come from smartphones which become a two-way communication of data back to the map provider. These signals can indicate there are new roads somewhere. But there are privacy implications - you can’t share someone’s phone location data - but can you aggregate the path 1,000 phones took and share that aggregate

[13] https://opendatacommons.org/licenses/odbl/

[14] https://www.hotosm.org/

route? Maps are created by people, which means there are mistakes. You can have the best sensors and AI/ML but you need to handle data vandalism, mistakes. If someone is navigated into the woods due to bad data, or a lake disappears from the map[15], what's the policy and approach to fixing that data? There are governance questions also around changes to data - if a shoe store changes the data on a competing shoe store, should that be allowed with 1 person changing the data? How do you establish who owns the shoe store?

Discussed copyright isn't for everything. Statutory damages are not always the right remedy. Open licenses shouldn't cover everything and there is a role in governance and policies that exist independent of the license.

CRA and its impact on open source

There are positive approaches to dealing with open source security challenges.

- Soverign Tech Fund (Germany) https://sovereigntechfund.de
- Open Technology Fund (US): https://www.opentech.fund

The United States Congress is debating a Securing Open Source Software Act.

- https://openssf.org/blog/2022/09/27/the-united-states-securing-open-source-software-act-what-you-need-to-know/

Europe has been looking at a Cyber Resiliency Act[16] (CRA). The scope covers products with digital elements, does not cover non-commercial products (including open source, sort of), services, and excludes certain products sufficiently regulated on cybersecurity (e.g. cars, medical devices, etc).

A number of open source organizations have published reactions and commentary on the CRA. OSI has published a list.[17]

The OpenSSF's policy committee drafted and submitted a response to the EU.[18] The response highlights the ubiquity of open source, the diversity of community participants, and key issues in the CRA draft.

- Improve clarity of Recital 10 Exemption for open source due to a lack of clarity about "commercial"
- Exempt open source in Article text to improve certainty
- Clarify the intended scope of "Distributor" with regards to software development platforms

[15] https://help.openstreetmap.org/questions/68553/lake-michigan-disappeared-from-openstreetmaphow-long-until-this-is-corrected

[16] https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

[17] https://blog.opensource.org/the-ultimate-list-of-reactions-to-the-cyber-resilience-act/

[18] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13410-Cyber-resilience-act-new-cybersecurity-rules-for-digital-products-and-ancillary-services/F3376650_en

    - The CRA focuses on physical store distribution
- Clarify scope for software components not intended for stand alone end-use
- Align any SBOM requirements with industry best practices and standards
- Revise Annex I requirements for delivery “without any known exploitable vulnerabilities” and handling vulnerability disclosure - to apply industry best practices
- Revise Annex II which was developed focusing on tangible products (e.g., consumer products such as a toaster or in the enterprise field, construction contracts). Applying these concepts to the intangible, digital environment raises enormous risks of confusion and unintended consequences, especially for open source software.
- Recommended the EU engage the broader open source software community, including the Commission’s own OSPO.

The CRA was recently a topic at FOSDEM[19] with an awkward moment when it was revealed the drafters had not consulted the EC’s own OSPO in developing the CRA draft.

Wednesday, February 15, 2023

An introduction to the day was provided with news that a guide for member counsel would be included with the slides in the post-Summit materials.

LF Research

Objective: describe the impact of open source projects, technologies and standards to solve the world’s most pressing challenges. LF Research was founded in 2021 on a strong tradition of research: 2018 Linux Kernel History Report, 2020 FOSS Contributor Survey. Empirical research methodologies, with 27 reports published to date on a wide variety of topics.

Deliverables:
- Written report
- Infographic
- Blog
- Webinar / Panel
- Open Data – published on opendata.world

Research is organized around frameworks, with a section on management and legal issues. Reviewed software and cyber security readiness projects. Highlighted in the technology framework reports on resiliency in multi-cloud environments, Data and Storage trends, Reviewed industry dynamics as a focus, with recent examples including the Academy Software Foundation research piece on telling the story as to why collaborating in open source is valuable. Review of examples in financial services industry and examining ways to created

[19] https://fosdem.org/2023/schedule/event/cyber_resilience/

shared value around common pain points. European focus and opportunities in the public sector in Europe was discussed.

Important Findings in Research in Open Standards

The survey included 497 respondents across company size

- Royalties are less important than selling services or products
- Nearly 75% believe that open or partially open are beneficial
- 40% believe royalties are required for standards to have value

Role of supporting maintainers and dispelling the myth that open source is a charity activity. Dr. Henry Chebrough and Innovation Economics report demonstrated that it is highly valuable and a lever to get out of an economic downturn.

Special Recognition of Karen Copenhaver's Contribution to Open Source

Mike Dolan highlighted the importance of understanding the open source context for providing legal advice in this space. Mike highlighted the important contributions of Karen Copenhaver to the open source community. Jim Zemlin highlighted Karen's contributions to the community and support of the Linux Foundation and her understanding of the importance of developers. Karen highlighted the importance of open source attorneys letting go of the concerns about risk and realize the potential through collaboration.

GPL 3.0 proposal for Automotive

What does it mean to put GPL-3.0 components in a car when there's a requirement that owners be allowed to modify that product?

One of the goals was to restrict Tivo-ization, restricting users to do what they wanted with their own device. There were terms put into GPL-3.0 to address this. As operating systems are coming into cars, cars are one large computing devices with wheels. Automakers are making software subsidiaries to address software in a car.

Is it safe to allow GPL-3.0 code into a car if you have to allow it to be modified? The other key question is whether you can ship a car with GPL-3.0 components. Most Linux-based operating systems include GPL-3.0 components, generally low-level operating system libraries.

GPL-3.0 has an "installation information" requirement plus automotive manufacturers have to deal with regulatory requirements.

> *If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for*

> *a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information.*

https://www.gnu.org/licenses/gpl-3.0.txt

> *A “User Product” is either (1) a “consumer product”, which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, “normally used” refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.*

Installation information.

> *"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.*

GPL-3.0 does accommodate proprietary software in various ways. It also includes limitations on the scope of GPL-3.0.

1. Installation Information is limited to “Covered Works+”
    a. Covered Works not the entire stack
    b. “+” installation information “must suffice to ensure that continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.”
2. User Product with modified GPL-3.0 binaries does not need to be fully functional
    a. Has to work on any equivalent hardware platform
    b. Does not mean a car as a whole needs to continue to function as a car
3. Installation Information need not be keys
    a. No need to provide owners of consumer products with signing keys - it’s possible to enable a mode or state that does not prohibit modified software from running
    b. Installation information can be provided in various forms, appropriate to the product
    c. For example, some Chromebooks make it possible for a user to override lockdown by engaging a normally-inaccessible physical switch or by accessing a normally-hidden menu

4. GPLv3 does not require continued network access
    a. *"Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network."*
    b. There's no affirmative requirement to provide access to a network. Cars today often have access to networks (e.g. cellular) and also operate networks within the car.
    c. Network access likely only applies in the limited sense of access to network interfaces to install modified binaries.
    d. Network connectivity for non-GPLv3 works is beyond the scope of GPLv3 because it's not under "Covered Works".

Auto makers can set requirements for network access.

Discussion: Cars are increasingly software-defined

Examples:

- Third party applications
- ADAS (assisted driving)
- Telematics
- Infotainment

Various regulatory bodies are involved: NHTSA, EPA, CA, etc. Cybersecurity is a risk as cars enable more interconnectivity and connectivity to cloud systems.

Right to repair laws require OEMs to provide consumers and independent dealers with information, manuals, diagnostics needed to repair devices. Massachusetts in 2013 which was followed by a 2020 Data Access ballot initiative to enable access to telematics. NY DIgital Fair Repair Act law extends to consumer devices and electronic equipment, but not automotive, medical devices, or where the "risk of improper installation heightens the risk of injury." NHTSA has raised concerns about safety and cybersecurity with right to repair laws.

Approaches that have been examined:

Conclusion – requirement under the license is only to provide license information for Covered Work and we believe would apply to the Linux stack to ensure that GPLv3 is not prevented or interfered with. That is required but nothing else, maybe around network but that is unclear.

Discussion regarding providing a dev board.

Litigations update Co-pilot, Vizio, Stability AI

Docket review current summaries.

1. SFC vs. Vizio

Announced in the press in October 2021. Alleged GPL violations. State court case in the Superior Court of California, County of Orange, Central Justice Center, Santa Ana, California, Civil Case No. 30-2021-01226723-CY-BC-CJC

“Plaintiff will… amend this Complaint to reflect the true names [of the] Does when their identities become known.

Claim is breach of contract (GPL-2.0 and LGPL-2.1), not copyright infringement. SFC says that they are suing in the capacity of a third party beneficiary of these ‘contracts’. Asking for source code that may be compiled without undue difficulty.

SFC uses language implying requirements for Installation Information under GPL-2.0. Not discussed in the session, but there is a JOLTS law review article on this topic at: https://jolts.world/index.php/jolts/article/view/149/270

Vizio first filed for removal to federal court, rational being this is really a copyright infringement claim.   Vizio cited Jacobsen v. Katzer, SFC objected. The Federal court cited Versata Software, Inc. v. Ameriprise Fin., Inc. in holding a contract claim is not preempted by a potential copyright claim.

Vizio filed an answer in June, 2022 with any and all possible defenses.

Why interesting:

1. Idea that GPL licenses are enforceable as a contract
    a. See this eBook: https://global.oup.com/academic/product/open-source-law-policy-and-practice-9780198862345?cc=us&lang=en&
    b. It’s a free PDF download if you click “Open Access”->PDF
2. Third party beneficiary theory
3. SFC appears to be testing its new theories of what is meant by “Corresponding Source”
    a. See LF Open Source Summit presentation here: https://osselc21.sched.com/event/IASE/in-person-installation-information-and-gplv3-gplv2-what-information-must-you-provide-and-what-do-you-need-not-provide-for-your-embedded-systems-mccoy-smith-lex-pan-law

The jury trial is schedule for September 25, 2023. Concern that this case creates a template for trolling actions.

Co-Pilot / Codex Cases

Federal case, GitHub, Microsoft, 6 OpenAI entities.

Under federal rules you are not supposed to use anonymous filings.

Expansive claims but there is no claim of copyright infringement.

- DMCA violation of “copyright management information”, 17. U.S. C. 1202
- Violation of the Lanham Act, 15 U.S.C. 1125
- Contract related conduct

Defense has responded:

- FRCP 12(b)(1) motion to dismiss for lack of Article II subject matter jurisdiction for using anonymous/unidentified plaintiffs without good reasons and motion for anonymity
- FRCP 12(b)(6) motion to discussofr for failure to state a claim
    - A multitude of arguments to support the motion
- Why is this interesting?
    - Plaintiff class action is a new enforcement model and could be dangerous path and generally only benefits the attorneys without any beneficial remedy
    - Enforcing FOSS compliance due to attribution-only license omissions
        - For a CMI claim, there’s no need to file a copyright registration in order to make a claim.
        - Statutory damages for removal of CMI are higher than copyright infringement
        - Plaintiffs claim damages of $90m to $900m
    - Issues will involve whether computer generated code is copyrighted and whether a copyright infringement is necessary to seek damages
        - Are LLMs merely extracting the functional rules for any particular programming language, and applying them to use prompt?
        - Are LLMs merely reproducing expressive, existing, programming text that best fits the requirements of a particular prompt?
        - What are the legal rules around the use of FOSS for training AI & ML code:
            - Depend on the tool?
            - … the license?
            - … particular piece of FOSS…

Andersen et al. v. Stability AI, Inc.

- Filed by the same lawyers as the Copilot case
- Has direct and indirect copyright infringement claims

Getty Images v. Stability AI, Inc.

- Copyright infringement, DMCA CMI violations, trademark infringement, deceptive trade practices

- Has interesting factual information in the complaint, looks like a watermark replication

There is another interesting case at the SCOTUS that may have implications for AI cases:

- https://www.scotusblog.com/case-files/cases/andy-warhol-foundation-for-the-visual-arts-inc-v-goldsmith/

Also had discussion of eBay vs Bidder's Edge: https://casetext.com/case/ebay-v-bidders-edge

AOM Panel

DG Comp received a formal request to review AOM's IP Policy relating to the AV1 video codec.

AOM's IPR Policy[20] is royalty free based on the W3C IP Policy[21] and contains a universal reciprocity clause. DG Comp discussed concerns raised about implications of AOM non-members who implement AV1. The panelists provided an update on the discussions with DG Comp in Fall 2022 and the danger of new, uninformed DG Comp personnel revisiting competition issues that were settled 15+ years ago.

There was discussion of alternative codecs (see for example, https://en.wikipedia.org/wiki/List_of_open-source_codecs). There was discussion that most implementers implement many video codecs in solutions, and wouldn't just implement AV1. Discussion was raised regarding other universal reciprocity IP approaches used by Bluetooth, USB-C, and the Apache-2.0 license. Discussion of the importance of universal reciprocity to small and medium sized companies who don't have a large patent portfolio or the ability to negotiate patent licenses. Studies have been commissioned showing benefit of having AV1 available as an option.[22]

Discussion of historical US views of standardization and the early rules around licensing just the minimal necessary claims, not patents. Implementation of the standards doesn't have any concept of necessarily infringed.

Discussion turned to all the pro-open source European Commission policies and that DG Comp intends to operate outside of the rest of the Commission's activities.

There have been multiple advocacy efforts with letters and a panel hosted by Open Forum Europe in Brussels. The panel video is available at https://www.youtube.com/watch?v=QTQEzKQFjXg&t=390s

[20] https://aomedia.org/license/

[21] https://www.w3.org/Consortium/Legal/2002/ipr-notice-20021231

[22] https://www.unifiedpatents.com/insights/2021/7/14/vvc-frand-royalty-report-by-charles-river-asso-estimates-lower-rates-because-of-av1-adoption?rq=VVC%20report

A summary of the Linux Foundation letter to DG Comp was discussed.

Open source security at IBM

Cybersecurity is a top priority and a huge challenge including customer impact, legal and reputation risks, limited budgets, skills and talent scarcity, and misunderstanding in policy making circles.

A top priority and a large challenge. Potential significant customer impact, serious reputational and legal risks, limited budgets, speed wins in the marketplace, skills scarcity, a technical subject matter and public/policy maker misunderstanding regarding open source.

Organizations must be compliant with regulations (enforcement risk) and be cyber secure (regarding internal security policies) (business risk).

Stakeholders must navigate the tension between compliance and security. Various teams will have different goals:

- Development
- Brand
- Legal
- Compliance and government
- Leadership
- Security professionals
- Financial professionals
- Country/international stakeholders

Review of the Log4j Vulnerability

Is the narrative accurate that community developed code is insecure? On the contrary it shows the resilience of the open source community. The problem was that people didn’t know what was in their code, they didn’t know about their dependencies. Even though there was a fix available if you didn’t know you had log4j you didn’t know you needed the fix. Also some people needed to effect a number of log4j upgrades in order to fix the issue.

Point of view on open source security. Open source is not inherently less secure than other software, it has advantages and disadvantages. We need to do more to support OSS security initiatives as do others. SBOMs is an area of focus, we believe they are critical because if you don’t have an SBOM you don’t know what you have in the code. Contribute directly to the open source projects, and if you don’t, then you should acquire paid support from the likes of SUSE or Red Hat.

Our focus has been around:

- Increased security scanning
- Focus on projets that we rely upon for increasing contributors
- Need for SBOMs
- Reliance upon tooling and automation, OpenSSF Scorecard

We are ingesting more open source, not less. Security requirements are going up and so the need for automation goes up.

Open source is important, not inherently less secure. SBOMs are critical but we are pushing back on making these things public, open source works only if users contribute back – either with engineering resources or paying a company who contribute.

Security needs to work with the open source model, for example policing the identity of contributors will not work.

We have worked to increase the number and diversity of professionals in this space, to do other things around education.

Open Source Project Office – under the Supply Chain Security team in the CISO

OSPO “Systme Managers” provide unit specific guidance and advice
Specific legal team focused on legal reviews in processes.

Further structural optimization necessary?

Open Source Security Foundation discussion, this project of the LF can help all of us be better stewards of Open Source Security. Quote from Jamie Thomas, “... a community mentality is needed to understand that security is the responsibility of each developer and not someone else’s problem.”

https://www.ibm.com/policy/open-source-security/

Open source security at Ericsson

Appreciative of the cooperation between the leaders of engineering and legal teams, and sees this as more important in the next 10 years than the last 10 years. To get the nuance about open source across, it’s not just the case that 80% of any product is open source but “modern software development is based on open source”. No developer today starts developing any application without going to the platform (e.g., operating system, ML framework, etc.). It is the essence of where our developers start, we have that level of dependency on open source now.

US legislatures don’t understand how essential open source is to our businesses. Open source is as water, and while there are sharks in the water, you don’t get rid of the water.

Ericsson is a telecom provider, and the industry is in a transformation involving disaggregation of the telco stack. This opens opportunities but also attack vectors.

On the topic of the CRA, we must educate our legislatures on how open source works. There will be another log4j. When we open our systems and tie it to a network there are vulnerabilities. Cyber threats will target that network and that software. It will take a lot of education.

At the same time the open source community can evolve. Attack surfaces within open source need to be managed. There is a risk point when there is a declining developer base, and this creates risk. We have work to do, and what we do will matter for all the software that we are writing.

Litigation update - Tulip Trading

Review of developer liability to users case. Tulip Trading has a number of cases. Claims there is a fiduciary duty of bitcoin developers to him. Question: can this theory be extended to other projects.

The court of appeals decision said it is possible that the developers have a fiduciary duty. Ruling was that there would not be a fiduciary duty unless code were maliciously changed. Professor Walch published an article on fiduciary duty of developers[23]. Professor Tamar Frankel argument regarding fiduciary duty when people in power have been entrusted with property. These arguments are refuted in a Stanford Blockchain Journal article[24].

Trademarks in the Wild

Example 1: provided of a project that was forked while the project name was maintained. While the "project" is owner of the mark, who is that if there is no legal entity for the project? Clarity on this point is important.

Example 2: project hosted by a company, but commercial offerings leveraged the same brand. Where we get into trouble is when the project wants to move to an independent home. Keeping the same branding as commercial products fights against that. If you convince the cojmmunity that the use of the marks are not confusing, you will then have to convince trademark offices (if you are interested in registrations) need to be convinced. But in the enforcement context, how do downstream users know that what is permitted by the project is also permitted by the company? A license does not solve the control issue.

Example 3: Formalities matter.

[23] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3203198
[24] https://assets.pubpub.org/jmjwk3m8/31573938342619.pdf

Project comes over but the company had registered the marks but not transferred them and then sold the project trademarks. This created a serious issue for the project but called into question about code contributions. It became more than a trademark issue which could have been avoided by completing the work of recording the assignment. When you really need a certificate (in an enforcement situation) the details in an enforcement matter.

Example 4.

Company developing a specification that could be developed in various languages. The good or service analysis might not be clear, it might be the specification being developed versus something else such as software implementation.

SPDX 3.0

A review of recent milestones from the SPDX Specification project.

SPDX consumption tooling review, including validator and translator tool, vulnerability lookup, minimum SBOM fields present.

SBOMs were very topical in 2022. OpenSSF had a SBOM Everywhere SIG. Discussion of when/how SBOMs should be ingested/outputted.[25] Review of the rewriting of the internal core model of SPDX.

The SPDX 3.0 model is anticipated to be released March / April. Expect to take 3.0 to ISO in the latter part of the year.

After 3.0, there are various groups working on additional functionality. Highlighted the safety standards automation efforts. There is a Functional Safety working group working on this. There is also an opportunity to maintain a hardware profile.

OpenChain Licensing and Security process standards

Update on trends seen by the OpenChain license and security compliance standard. There are now 10 third party OpenChain certifiers.

[25] https://ntia.gov/sites/default/files/publications/sbom_formats_survey-version-2021_0.pdf

In addition to standards, OpenChain produces free online training courses.

Evolving code of conduct

A review of issues in addressing codes of conduct violations. 10 years ago it was difficult to adopt a code of conduct, but now they are widespread but how they are enforced is still controversial.

Bad behavior can be common in open source, the most common is rudeness. But there can be more extreme behaviors such as sexual harassment.

A welcoming community does matter, and codes of conduct are important to underrepresented groups. HR departments do not have jurisdiction, and so codes of conduct are important. There are legal risks when enforcing codes of conduct.

Employees that volunteer on codes of conduct committees need access to company legal counsel. The legal risks increase with the severity of the behavior or if the consequences could impact someone's career.

Lawsuits can be brought by the accused, by someone injured, or by the employee of a hosting foundation. Codes of conduct violation can create a hostile work environment.

A review of examples:

1. A failure to perform a thorough investigation by the code of conduct committee (failure to present evidence, notice, etc.);
2. Thought policing (e.g., saying things others find offensive through social media);
3. DEF Con Code of Conduct litigation, a permanent ban coupled with naming online. Publicly naming someone is almost an invitation for a lawsuit. Dismissed without prejudice, and plans to refile in another jurisdiction.
4. Liability for personal injury, asserting that the project leaders or committee did nothing to stop the harassment.

Challenges of Enforcement

Time consuming, it can be difficult to find community members who are willing to serve on a code of conduct committee. Sometimes the only ones that step forward bring their own biases. Sometimes there are conflicts of interest.

An alternative would be to have foundation staff run enforcements. That has challenges too, it can be seen as 'corporate'. Criticisms will get directed at the foundation. Also, there is a shortage of professional service providers who know how to do this work. It isn't like a typical HR investigation.

Hybrid Model of Enforcement

An enforcement model that several LF communities have adopted involves:

- A code of conduct committee consisting of both foundation staff and community members; and
- Engagement of external service providers (legal counsel, professional investigators)

How can Member Counse Help?

- Volunteers for code-of-conduct committees needed;
- Contribute to improving the code of conduct documentation in project you are involved with;
- Educate your employees on code of conduct committees regarding issues and risks.
- Consider a project’s code of conduct as a health indicator for the project.

Establish a project for Code of Conduct Standards, a current idea being proposed by David Rudin of Microsoft.