Software Provenance Analysis - Acquisition Due Diligence

Agenda

- About nexB
  - What nexB does
  - Our experience
- Software Audit Due Diligence
  - Software Audit Process
  - Software Audit Tools & Techniques
  - Lessons Learned
- Why nexB

What nexB does

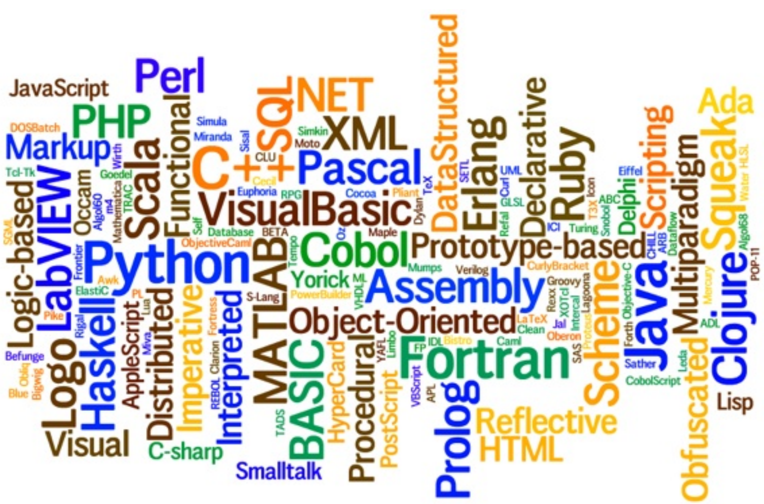
OSS Compliance Tools

- **DejaCode governance application (Cloud)**
- **AboutCode Tools**
  - For developers
  - Licensed as OSS

Software audit services

- **Software products**
- **Acquisitions**

Expertise in all languages



Active OSS developers

- **https://github.com/nexB**
- **Google Summer of Code**
- **ClearlyDefined**
- **SPDX**

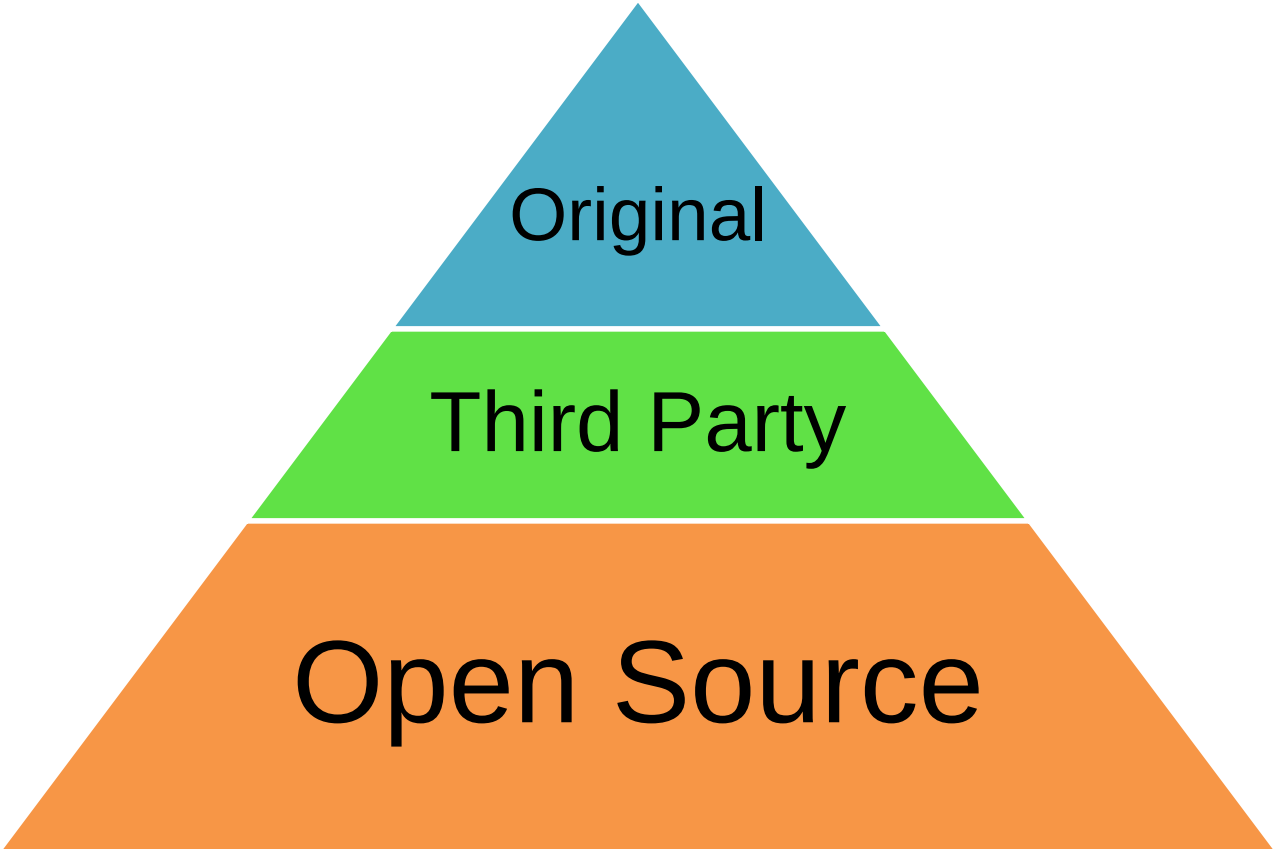
Why Software Acquisition Due Diligence

- Modern software contains on average more than 75% open source software (OSS)
- As a buyer of software assets, you need to know:
  - What specific open source components are used and how?
  - Is their any Copyleft-licensed code? and exactly how does it interact with proprietary code?
  - What will be your OSS compliance obligations? and how well does the current product comply?
- In summary, what are the risks that come with the rewards of using open source

nexB Role in Due Diligence

- Expertise in:
  - Software provenance (license and origin) analysis
  - Open source governance and compliance
  - All software languages and platforms
- Trusted third party
  - Mitigates confidentiality concerns of a seller company
  - Maintains proper segregation of information during acquisition negotiations
  - Enables objective analysis with appropriate consideration of feedback from all parties

Software Analysis Scope



Modern software products comprise 70% or more open source or third- party components

Software Analysis Options

Depending on your schedule and priorities

1/ Copyleft & Commercial issues

- Focus only on copyleft and commercial code

2/ Deployed Bill of Materials (BOM) only

- Focus on what code is actually visible to a customer

3/ Deployed BOM only with Development codebase details

- BOM of Development codebase components that are Deployed on the product

4/ Development Codebase Inventory

- Inventory of Development codebase components
- Details for Deployed components
- Summary for non-Deployed

Software Analysis Deliverables

- **Specific Action items and recommended actions for resolution** that can be factored into the deal terms
  - Including possible exposure for older product versions
  - Detailed analysis for copyleft “contamination”
- Checklist of commercial components as input to due diligence for contract review
- Analysis of how much code is original versus open source (OSS) or third-party (Commercial)

Preparation – up to 1 week

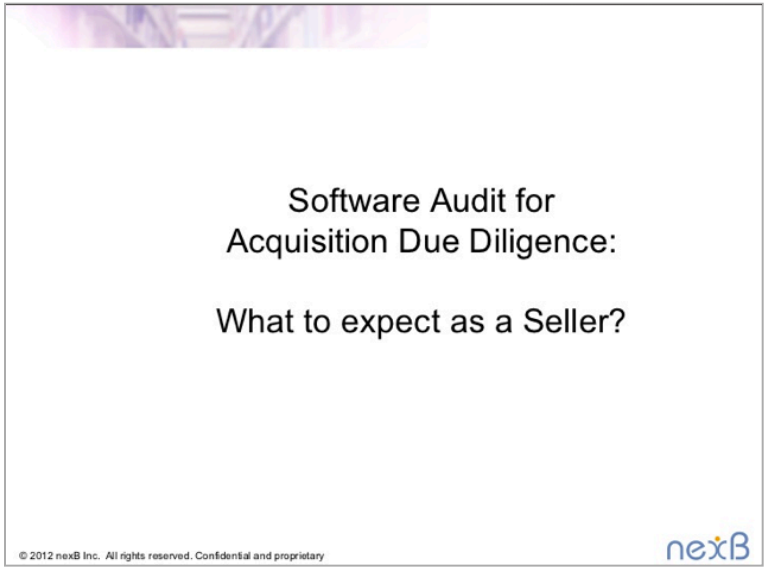
- Establish NDA with seller
  - Two-way or three-way
- Scope audit effort
  - Audit profile (questionnaire)
  - Size of code base - # files and lines of source code
  - Disclosure of known third-party and open source software
  - Onsite or remote access to the code
- Prepare/agree quote – **fixed fee, no surprises**
- Schedule project

nexB

Preparation

➔ Many targets are anxious about the process

- General level of anxiety is inversely proportional to prior M&A experience of executives
- We do some hand holding to make them feel comfortable
- Assure seller that they review all findings first so no surprises
- Explain the process and tools to the seller



Analysis – up to 2 weeks

Activities

- Scan codebases – Development and Deployment/Distribution
- Identify (conclude) open source and third-party packages components
  - Create software inventory for Development codebase(s)
  - Trace Deployment/Distribution components to Development
  - Create software BOMs for Deployment/Distribution package(s)
- Identify issues:
  - Analyze software interaction and dependency patterns for copyleft-licensed components as needed
  - Additional domain-specific investigations may be needed
  - Recommend mitigation / remediation Actions

nexB

Analysis

Results

- Software Inventory and Bill(s) of Materials
- Draft Issues and Actions

BOM Ref #	Location	Component	Version	Origin	License Style	License	Match Level	Code Type	Attribution	Redistribution	Patent Impact	Modified	Deployed	Interaction	Action	Notes	URL to Original code	Location of License Text
1000	/xxx/xxx/xx.xxx	xx.xxx	v1.6R3	Mozilla Rhino	Copyleft	Choice of NPL 1.1 or GPL 2.0 (future versions under Choice of MPL 1.1 or GPL 2.0)	Library	Binary	x	x	x		Y		Upgrade to version 1.7R1 or higher, or make sure the offending debugger binaries (which are not used in the product code base) are removed from the redistributed 1.63R Rhino jars. Rhino 1.7R1 or higher DOES NOT have proprietary Sun code in it.	There are license issues with all versions prior to version 1.7R1 leading Mozilla to remove the debugger code from the component distribution in recent releases. The debugger contains non-open source SUN proprietary binaries in: js.jar/org/mozilla/javascripttools/debugger/downloaded/ The license prohibits modification and redistribution.	http://www.mozilla.org/rhino/	http://developer.mozilla.org/en/Rhino_License
1020	/xxxxxx/xxxx/xxxx/xx/xxx/xxx/XXXXXXX/xxx/xxxxxxx/xxxxxxxxxxxx.xxxx	xxxxxxxxxxxxxx.xxxx		SwingPix Photo Organizer	Copyleft	GPL 2.0	File	Source	x	x	x		Y		Replace or rewrite to avoid GPL impact	Match with swtemary.c file from swingpix. This file is included in a standalone tool otherwise already GPL-licensed. The original notice has been removed.	http://sourceforge.net/projects/swingpix/	
1090	/xxxxxx/xxxx/xxxx/xxx/xxx/xxx/xxx/xxxx/xxxxYY.xxxx	xxxxYY.xxxx		Oracle/Sun Java JDK	Commercial	Sun Binary Code License Agreement	Snippet	Source				x	Y		Replace or rewrite or obtain commercial license from Oracle. Alternative implementations are available under non-copyleft licenses such as in Apache Harmony.	Match with JDK's java.util.prefs.Base64.java file . The JDK license does not grant rights to modify and redistribute JDK code in source or binary form	http://www.oracle.com/	
1120	/xxxxxx/xxxx-xxxxxxx-xxx/xx/xxxxxxxx/xxxxxxxxxx/	xxxxxxxxxxxxxx.xxxx		MySwing	Copyleft Limited	LGPL 2.0	Snippet	Source	x	x	x		Y	Static linking with LGPL	**Replace or rewrite to avoid LGPL impact**	120 lines match starting at line 12 with EventHandler.java from MySwing. This snippet makes the files and all the proprietary code depending on this or built in the same Jar potentially subject to the LGPL.	https://myswing.dev.java.net/	Source file header
1130	/xxxxxx/xxxx-xxxxxx/xxxxxY	xxxxxY, xxxxxY	v0.97 or earlier	GNU GRUB	Copyleft	GPL 2.0	Library	Binary	x	x	x		Y			GRUB is a bootloader and does not interact with proprietary software.	ftp://alpha.gnu.org/gnu/grub/	
1150	/xxxxx/xxxx_xx/xxxx/xxxx/xxx/xxxx/xxxxxxx/	xxxx_xxxx.x		NetBSD	Attribution	BSD-Original	Snippet	Source	x			x	Y		**Replace or rewrite or demonstrate compliance with BSD Original license**	75 lines match starting at line 23 with print-ascii.c from NetBSD. The BSD Original license has extensive advertising obligations.	http://www.netbsd.org/	
1160	/xxxxx/xxxx_xx/xxxx/xxxx/xxx/xxxx/xxxxxxx/	xxxx_xxxx.x		Libcddb	Copyleft Limited	LGPL 2.1	Snippet	Source	x	x	x	x	Y	Static Linking with modified LGPL	**Replace or rewrite to avoid LGPL impact**	70 lines match starting at line 56 with cddb_net.c, LGPL-licensed.	http://libcddb.sourceforge.net/	
1170	/xxxxxx/xxxx-xxxxxxx-xxx/xx/xxxxxx/xxxxxxxx/xxx/xxxxxx/xxxxxxxxxxx/xxxxxxxxxx/	xxxxxxx.xxxx		LizardWorks TIFF Library	Copyleft	GPL 2.0	Snippet	Source	x	x	x		Y	Linking with GPL	**Replace or rewrite to avoid GPL impact**	54 lines match starting at line 200 with IDFEntry.java from the Lizardworks TIFF library.	http://www.lizardworks.com/libs.html	Copyleft.txt in original source archive
1180	/xxxx/xxxxxxx/xxxxx/xxx/xx/xxx/xxxxx/xxxxxxxxxx.xxxx	xxxxxxxxxx.xxxx		Scientific Annotation Middleware	Attribution	Apache 1.1	Snippet	Source	x			x	Y		**Restore original license notice**	89 lines matchy starting at line 345 with file ExitStatus.java. Original Apache 1.1 software license notice missing	http://sourceforge.net/projects/sam/	
1210	/xxxx/xxxxxxx/xxxxx/xxx/xx/xxx/xxxxxxx/xxxxxx-xxxxxxxxxxxxxxxx-.xxxx	xxxxxxxxxxxxx.xxxx-.xxxx		Atlassian spring-portlet-sample	Attribution	Apache 2.0	Snippet	Source	x		x	x	Y			30 line match with file named "Uploadlib-.java" under "/spring-portlet-sample-src.zip\spring-portlet-sample\xxxx-INF\src\org\springframework\xxxx\portlet\sample"	http://opensource.atlassian.com/confluence/spring/download/attachments/10/spring-portlet-sample-src.zip	
1230	/xxxxxx/xxxx/xxxx/xxx/xxxxx-xxxY.Y/	xxxxx-xxxY.Y	vcvs-20060406	JRadius	Copyleft Limited	LGPL 2.1	Library	Binary	x	x	x		Y			LGPL 2.1 for the core development kit, GPL 2.0 for the applications. Only the development kit is used.	http://downloads.sourceforge.net/jradius/jradius-cvs-20060406-java15.zip	http://jradius.sf.net/license.html and License file in the original distribution
1240	/xxxx/xx/xxxxxxx/xxxxxxx/xxxxxxx/xxx/xxx/xxxxxx/xxxxxxxxx.xxxx.xxxx	xxxxxxxxxxxx.xxxx	v1.5.0-06	Oracle/Sun Java JDK	Proprietary	Sun Binary Code License Agreement	Snippet	Source	x			x	Y		**Replace or rewrite or acquire commercial license from Oracle.**	57 lines match starting at line 45 with Arrays.java from the JDK. The JDK license does not grant rights to modify and redistribute JDK code in source or binary form.	http://java.sun.com/javase/index.jsp	
1310	/xxxx/xx/xxxxxxx/xxx/xxxxxx/xxxxxxxxxxxx.xxxx	xxxxxxxxxxx.xxx		JRobot	Attribution	BSD-Modified	Directory	Source	x			x	Y		**Restore original license notice**	License/notices from the origin are missing in the source files.	http://sourceforge.net/projects/jrobot/	
1360	/xxxx/xx/xxxxxxx/xxxxxxx/xxxxxxx/xxxxxxxxxxxxxxxxxx.xxxx	xxxxxxxxxxxxxxxxxx.xxxx		Blitz JavaSpaces	Attribution	BSD-Modified	File	Source	x			x	Y				http://sourceforge.net/projects/blitzjavaspaces/	

Review & Report – 1 week

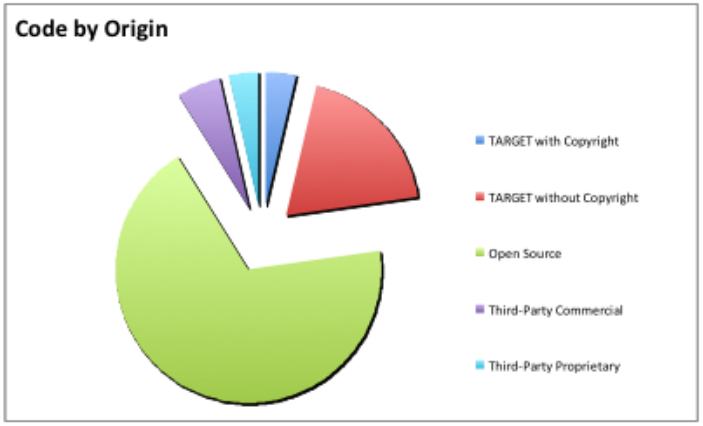
Activities

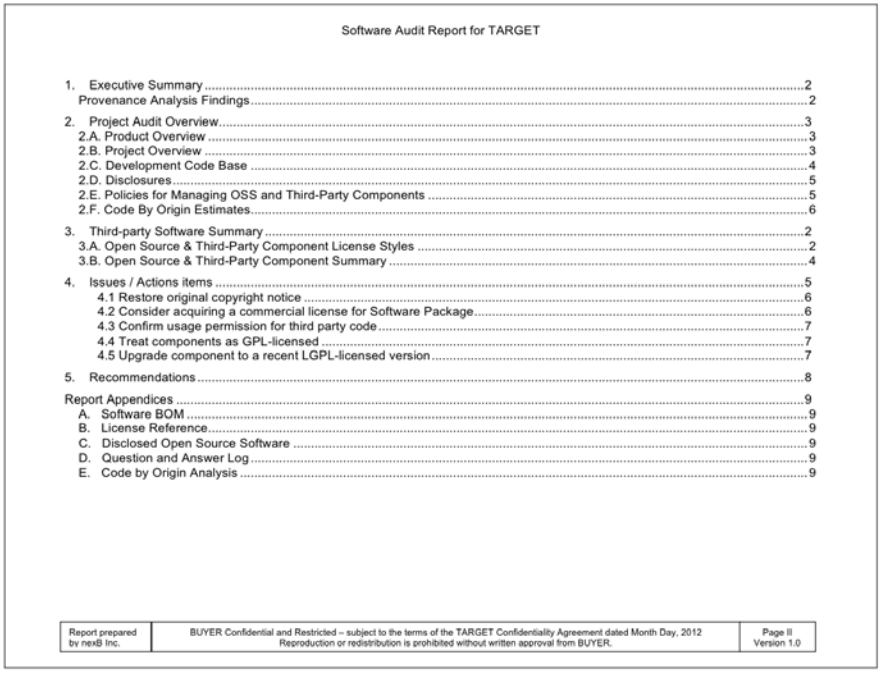
- Review draft findings with product team
  - Ask product team to respond to each Issue
    - Accept recommended solution or propose another approach
    - Acknowledge & investigate
    - Not a request to fix anything during the audit
  - Incorporate feedback and answers from product team into the Software BOM(s) and Report
  - We may “agree to disagree” – e.g. we then present two points of view: ours and the seller’s.
- Complete final report
  - Second review cycle with product team
  - Release the report
  - Conference call with buyer to present findings & answer questions

Review & Report

Results

- Final Software Inventory / BOMs spreadsheets
- Final Report - narrative with executive summary, project data and summary of Issues and Actions





Software Audit Report for TARGET

Tools

- We use tools from our own AboutCode project
  - ScanCode Toolkit - to scan code files for license, copyright and other provenance data.
  - AboutCode Manager - to review scan results and record license and copyright/owner conclusions.
  - AboutCode Toolkit - to document/track component license data in your codebase(s) and generate attribution notices.
  - TraceCode Toolkit - to identify deployed/distributed components based on tracing a product build from source to end-product.
- AboutCode is a nexB-sponsored open source project
  - Set of tools based on integrated AboutCode Data Model and industry standards – e.g. SPDX
  - Get the code from https://github.com/nexB

Techniques

- Multiple layers of analysis
  - Discovery: direct scan for license and copyright notices
  - Identification: component matching for open source and publicly available third-party components (freeware/proprietary)
  - Trace binaries back to source
  - Interaction and dependency analysis as needed
- Many utilities / tools are for specific languages / platforms
  - RPM and other package metadata
  - Docker containers
- Review and conclusion by software experts
- All require expert humans to interpret the results!

nexB

Lessons Learned – Acquisitions

- Schedule is **always** a major issue
- Initiate a software audit early because
  - Seller company will probably not have done this before
  - Negotiation of an NDA takes longer than you expect
  - Negotiation of access to artifacts and people takes longer than you think
- The review of findings and recommendations may require several iterations with target company
  - Get answers for open issues
  - Get agreement about remediation strategies
  - Get agreement that report is objective and reasonable

Lessons Learned – Acquisitions

- Identify the “crown jewels” and key platforms of the seller technology
  - Concentrate the audit on the most important parts
  - For products with multiple operating system versions, focus on the most important platforms

- Some issues can be specific to the open source policies of the Buyer
  - For instance tolerance for certain version of open source licenses or proprietary Linux drivers varies among companies
  - We apply Buyer company policies if available,
  - Otherwise we apply “conservative” community standards
  - Exceptional cases may require additional discussion with legal and and business teams to evaluate the risks

Why nexB

- We analyze Deployment/Distribution code so that you have real Software BOMs, not just an inventory
- We identify issues along with practical remediation actions
- 350+ software audit projects completed to-date

What they're saying.

> "I will most definitely be remembering nexB, as I was very impressed by your scan and how smooth the process was."

> "Again, thank you for all the work you and your team did. It was exactly what we needed."

> "It was our pleasure to work with you and your team. nexB was really professional and had an impressive analysis."

Contact us

Contact person:

Michael Herzog
mjherzog@nexb.com
+ 1 650 380 0680

More information:

http://www.nexb.com/

https://github.com/nexB