**The GitHub Archive Program - Script**

**Presenter 1**

**Slide 1: Title**

**Script:** “Hello, everyone. Today, our group will be talking about the GitHub Archive Program. Our presentation will explore what the program is, why it matters, and how it works.”

**Slide 2: What is it?**

**Script:** “The GitHub Archive Program is an initiative that serves to protect and preserve the world’s open-source code so that it isn’t lost to time, disaster or organizational collapse. GitHub recognizes that code is more than just lines typed into a computer. It’s a product of cultural, technical, and historical evolution and should be treated as such. Just as we preserve books, art, and architecture, this program treats code as something future generations should have access to study and learn from it.”

**Slide 3: Why It Matters**

**Script:** “This program matters because so much of our modern infrastructure depends on open-source projects. The preservation of open-source code ensures the longevity of its digital heritage allowing future scholars, developers and researchers to study, understand and build upon past software. It also reinforces the idea that open source is not just a licensed model, but part of our shared culture.”

**Presenter 2**

**Slide 4: Key Initiatives Overview**

**Script:** “To achieve its mission, the GitHub Archive Program uses several flagship initiatives. These include the Arctic Code Vault, which is a long-term cold storage solution; the Greatest Hits Archives, which preserve the most important projects in world-class libraries; and partnerships with organizations like Software Heritage and the Internet Archive. Together, these initiatives create multiple layers of protection.”

**Slide 5: Arctic Code Vault**

**Script:** “The Arctic Code Vault is perhaps the most famous part of the program. On February 2, 2020, GitHub took a snapshot of every active public repository – which equated to about 21 terabytes of data – and stored it on archival film designed to last up to 1,000 years. This film is stored deep inside a decommissioned coal mine in Svalbard, Norway, one of the most remote and geologically stable places on Earth. The idea is that even if the internet collapsed tomorrow, future generations could still recover this code.”

**Slide 6: Greatest Hits Archives**

**Script:** “Another initiative is the Greatest Hits Archives. GitHub identified the most relied-upon projects – those with the most stars and dependencies – and preserved them in museum-quality archives. These were deposited in institutions like the Bodleian Library at Oxford, the Bibliotheca Alexandrina in Egypt, and Stanford Libraries. By placing code in these trusted cultural institutions, GitHub ensures that it’s preserved alongside humanity’s greatest works.”

**Presenter 3**

**Slide 7: Partners & Collaborators**

**Script:** “GitHub doesn’t do this alone. It also partners with leading archival organizations. Software Heritage preserves full development histories, not just snapshots. The Internet Archive provides frequent web backups. Microsoft’s Project Silica is experimenting with storing data in quartz glass, which could last thousands of years. And groups like the Long Now Foundation bring expertise in long-term stewardship. Together, these partners create redundancy and resilience.”

**Slide 8: Technical Details**

**Script:** “Technically, the program uses some clever strategies. The Arctic Code Vault uses Piql archival film, which encodes data in QR-like patterns. Each reel also includes a human-readable recovery guide, so even if future technology is very different, people can still decode the data. The program follows the LOCKSS principle – which stands for Lots of Copies Keep Stuff Safe – and uses a multi-layer strategy: hot storage on GitHub itself, warm storage in the Internet Archive, and cold storage in the Arctic.”

**Slide 9: Strengths of the Program**

**Script:** “By spreading copies across different media and geographies, the program is reducing the risk of total loss. The human-readable recovery guides make it more likely that future generations can decode the data. Trusted partners like libraries and archives add credibility and stewardship. And the durability of the storage media – from film to glass – ensures that the code could last centuries, if not millenniums.”

**Presenter 4**

**Slide 10: Challenges & Criticisms**

**Script:** “But of course, there are challenges. The Arctic Code Vault is just a snapshot from 2020, so it doesn’t capture ongoing development. Future decoding could be complex, especially for experimental media like Project Silica. There are also legal and copyright questions, like who owns the code, and under what terms can it be accessed? As well as governance being an open question: who decides how and when the archives are updated or accessed?”

**Slide 11: Broader Significance**

**Script:** “Despite these challenges, the broader significance is undeniable. The program reframes code as a cultural heritage, worthy of preservation alongside books and art. It also provides lessons for digital preservation more generally – showing how redundancy, diverse media, and trusted institutions can work together. And it inspires future archival projects, reminding us that what we create today may be studied centuries from now.”

**Slide 12: Conclusion & Q&A**

**Script:** “In conclusion, the GitHub Archive Program is a bold and innovative effort to safeguard open-source software. By balancing immediate access with long-term preservation, it ensures that the software shaping our world today will remain accessible to future generations. It’s a reminder that code is not just functional – but also cultural.”