Advocating for free software in healthcare

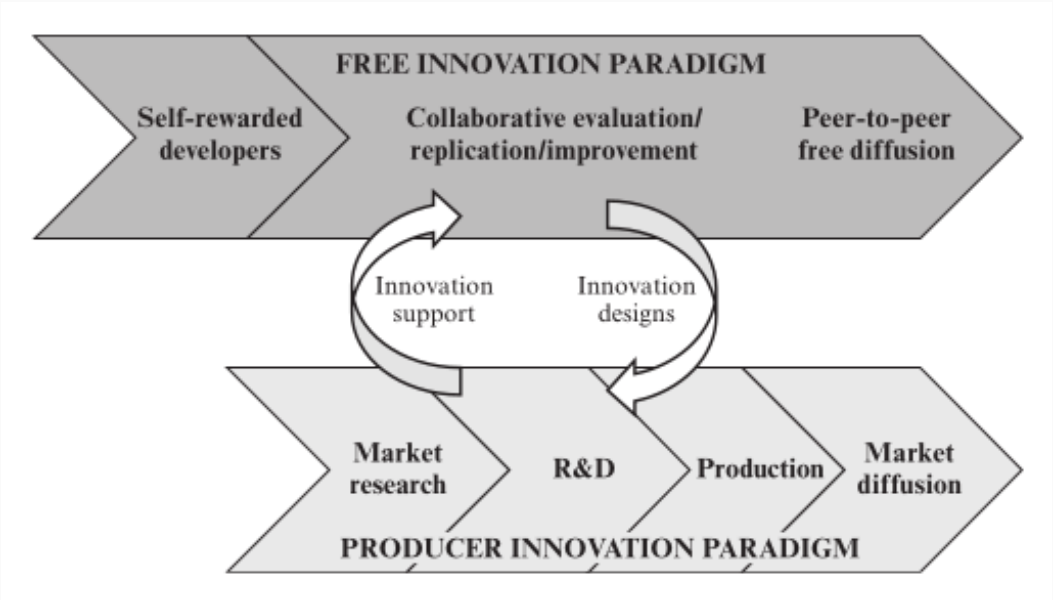
Leonor Palmeira

Open Science Day, November 7th 2024

Introduction

Sources of innovation

Driver	Funding	Motivation
Manufacturer driven	Private	Prospect of reaching a market
Science driven	Public/Private	Solve a scientific problem + publish/perish model
User driven	None/Indirect	Solve a personal problem



von Krogh and von Hippel, 2006 ; von Hippel, 2016

Innovation and intellectual property (IP)

Protecting IP comes with both values and costs

Intentional and routine free-revealing

Collective invention 1800s iron and mining industry

Professional societies Meetings and public reports by employees of competing firms on performance data and design improvements

Free revealing innovation Chemistry analyzers for medical diagnosis (Technicon Corporation)

User expertise Medical personnel freely revealed major improvements (analyzer and tests)

Allen, 1983; von Hippel and von Krogh, 2006

Overview of proprietary/open source software

Table 1. Comparison of proprietary and open source software development methods

Aspect	Proprietary software	Open source software
Software owner	Company, shareholders	Community, citizens
Foundations of product	Other products on market with a few distinct changes (analogy: me-too drug)	Existing tested code base (analogy: generic drug)
Pricing model	What the market will bear	Cost recovery
Development team	Professional programmers isolated from user base	Mix of professional and amateur programmers, often including users
Development team strategy	Cut and run, lock in to proprietary code	Code reuse, continuing quality improvement
Development team dynamics	Small, centralized, managed	Large, decentralized, meritocratic
Developer incentives	Salary, internal promotion	Community, recognition, contribution to application area ± salary & promotion
Method to test and assure quality	Internal synthetic test cases, team integrity	Real cases, user testing, open inspection by community
Driver to respond to user needs and requests	Market share	Community-prioritized need
Intellectual input	Small team, distorted by team dynamics	Wisdom of crowds

Reynolds, 2011

Situated knowledge

Overview

CHU of Liège - Human Genetics Department

Down's syndrome's

Non-invasive **prenatal** diagnosis

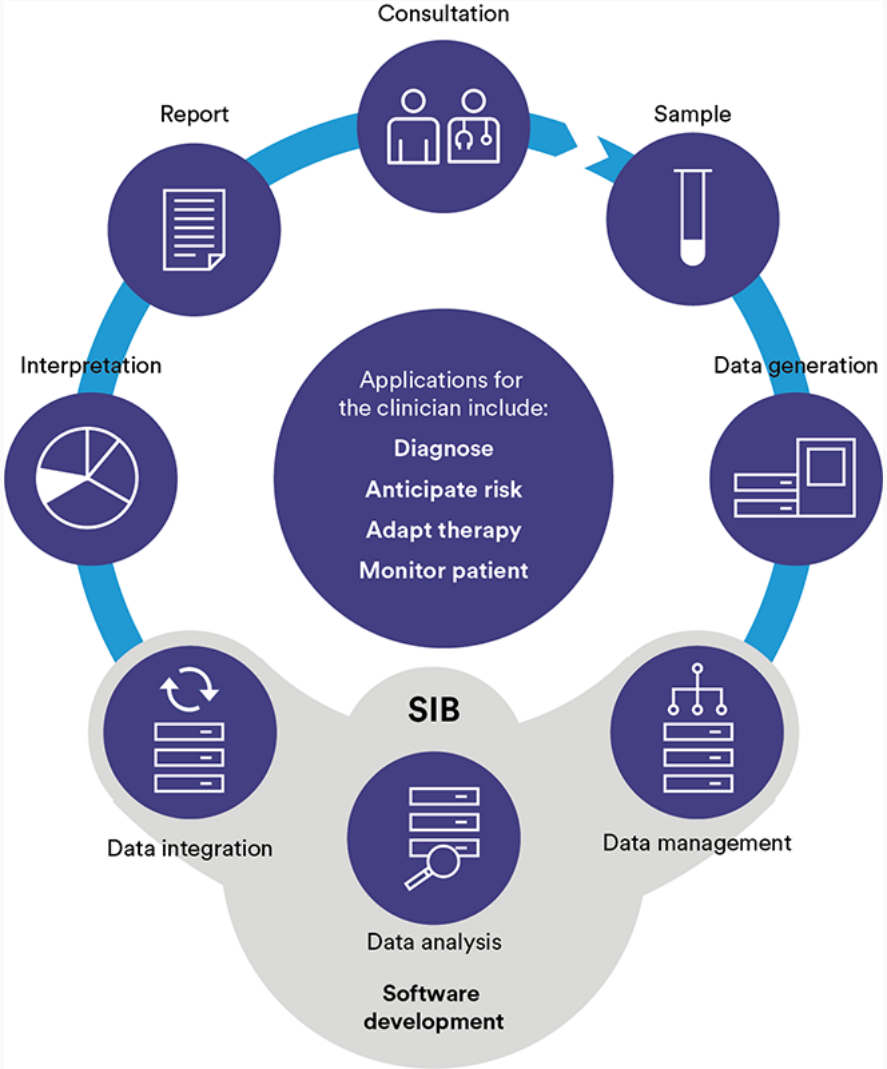
Diagnostic of **cancers**

Detection of mutations

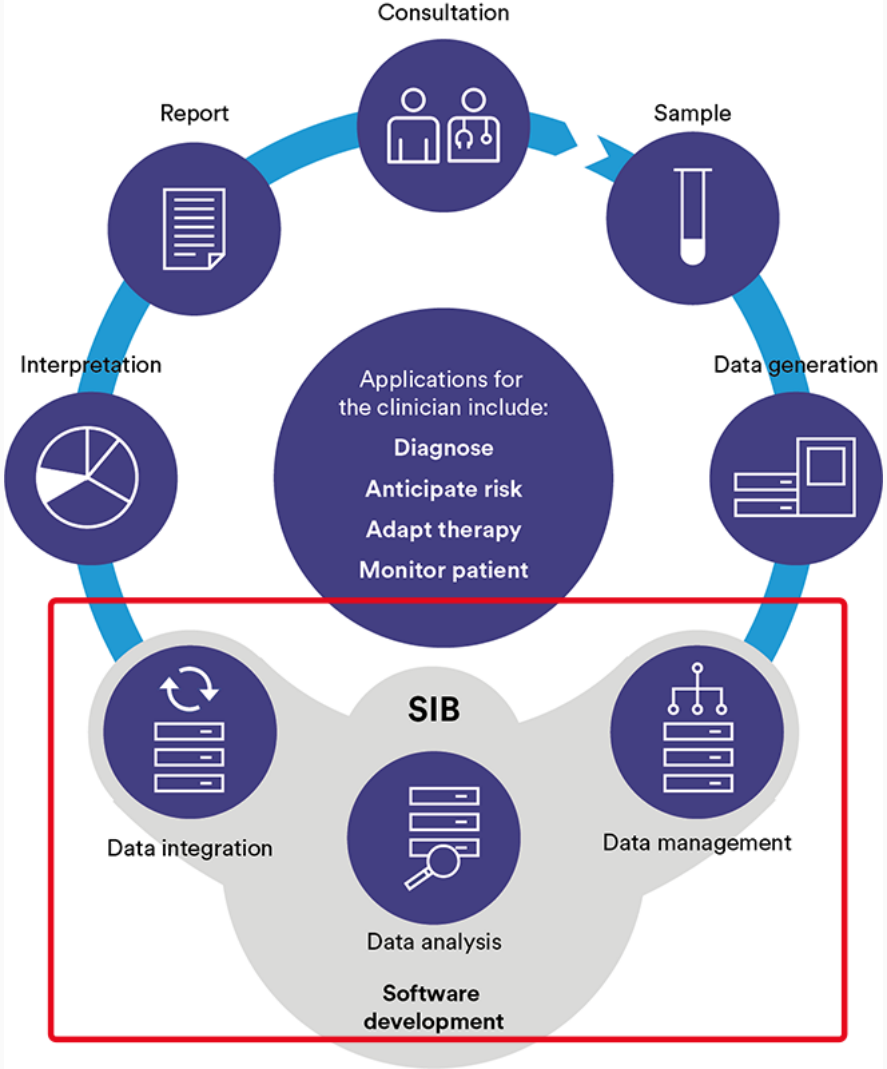
Diagnostic of **genetic diseases**

Postnatal detection of mutations

Overview


Consultation
Sample
Data generation
Data management
SIB
Data analysis
Software development
Data integration
Interpretation
Report
Applications for the clinician include:
Diagnose
Anticipate risk
Adapt therapy
Monitor patient

Overview


Consultation
Report
Sample
Interpretation
Data generation
Applications for the clinician include:
Diagnose
Anticipate risk
Adapt therapy
Monitor patient
SIB
Data integration
Data management
Data analysis
Software development

Bioinformatics for genetic diagnostics 1/2

Bioinformatics team @CHU, Liège

- Scientists with experience in
  - Open science
  - Open data
  - Open source
- A wide variety of scientific profiles:
  - Biology
  - Bioengeneering
  - Physics

Bioinformatics for genetic diagnostics 2/2

Clinical diagnostics constraints

Normative ISO 15189:2022 (Medical laboratories)

- Traceability
- Validation/re-validation
- Continuous improvement

Budget Limited and highly monitored (public tenders)

Production Turn-around time + Routine execution of analyses

Innovation From bench to bedside: urge to develop new applications

Commercial software in healthcare

Significant power imbalance

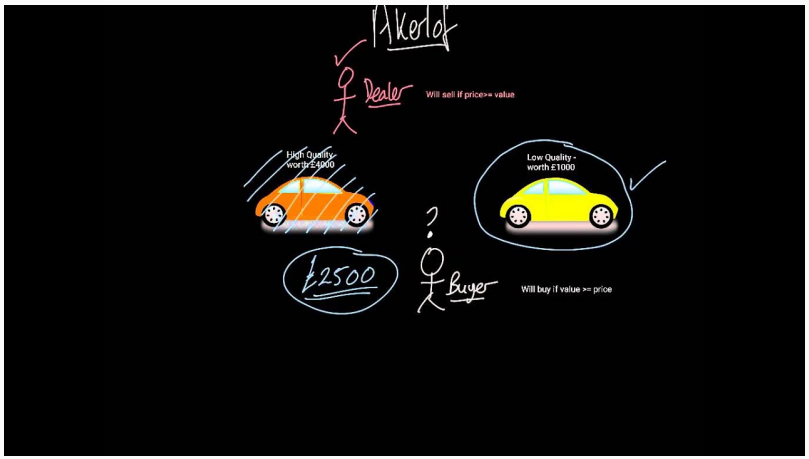
The Lemon Market Theory

Asymmetrical information Buyer lacks technical knowledge to discriminate between products

Reducing asymmetry? Gain internal technical knowledge

Reducing asymmetry? Software implements standards

Reducing asymmetry? Buy maintenance contract



The risk of lock-in in healthcare software industry

- Lack or poor support of **open scientific standards**
- **Poor interoperability** between software bricks/steps
    - Different users/institutions cannot communicate
- Production-acquired knowledge cannot be exported/imported (**data loss**)
- High cost of product switching due to **ISO constraint**
- + Risk of non-neutrality when accessing the commercial market (risk for public tenders)

Public funding absorbed into closed source code 1/2

Spin-off model (funding: public ↪ mixed ↪ private)

- ↪ *Example: Alissa Interpret, Agilent Technologies*
- ↪ Laboratory development (KU Leuven)
- ↪ University spin-off (branded as Cartagenia)
- ↪ Bought by a commercial company (re-branded as Alissa)
- ↪ Discontinued by commercial company

Public funding absorbed into closed source code 2/2

Collaboration model (source: open ↪ mixed)

- ↪ *Example: DRAGEN-GATK, Broad Institute + Illumina*
- ↪ Laboratory development (Broad Institute)
- ↪ Community building: reputation+improvements (Best Practices)
- ↪ Commercial company collaboration (Illumina)
  - Gained reputation is used to promote commercial product
  - Improvements no longer transparently fed back to community (2019/2024)

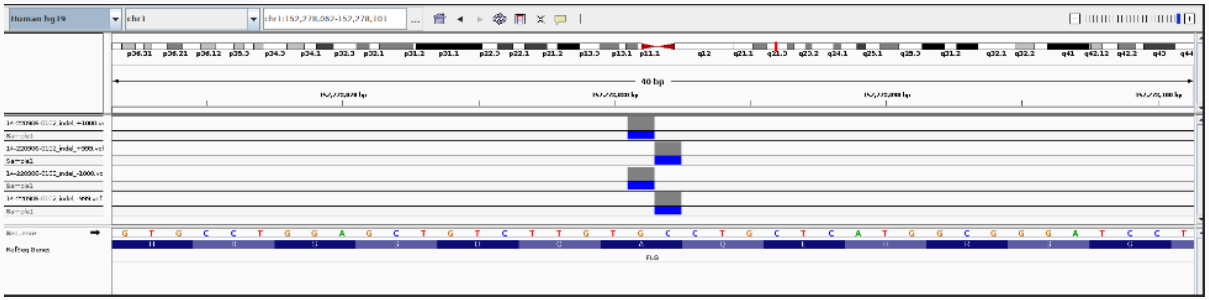
Use cases

Use case 1: Bug fix *Closed source*

chr17	30897	31441	NM_003585.5_exon_0	0	-
chr17	22327	22407	NM_003585.5_exon_1	0	-
chr17	13920	13995	NM_003585.5_exon_2	0	-
chr17	11871	11981	NM_003585.5_exon_3	0	-
chr17	11205	11332	NM_003585.5_exon_4	0	-
chr17	6010	6168	NM_003585.5_exon_5	0	-
chr17	-1097	-1015	NM_003585.5_exon_6	0	-
chr17	-2035	-1938	NM_003585.5_exon_7	0	-
chr17	-7419	-2630	NM_003585.5_exon_8	0	-

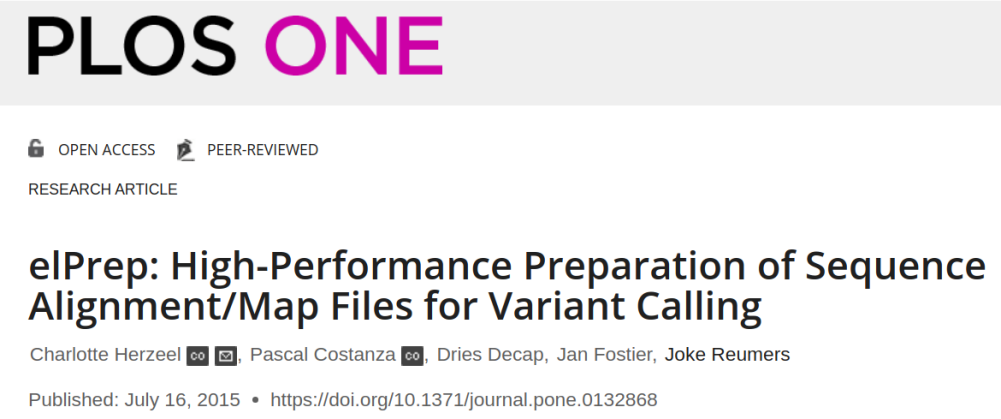
- *Example: Alissa Interpret, Agilent Technologies*
- Identification of a **bug** generating negative genomic positions for some transcripts
- Reported in Feb. 2020 + root-cause identified
- **Bug still present** and reported again in May 2022
- Software discontinued in 2024 without bug fix

Use case 2: Bug fix *Open source*



- *Example: VarScan2, McDonnell Genome Institute*
- Identification of a **bug** where a deletion was wrongly identified as a point mutation
- Reported with root-cause and fix (March 3d 2023)
- Local fix implemented for **CHU production**
- New official version with this and other fixes (March 28th 2023)

Use case 3: New features *Open source*



PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

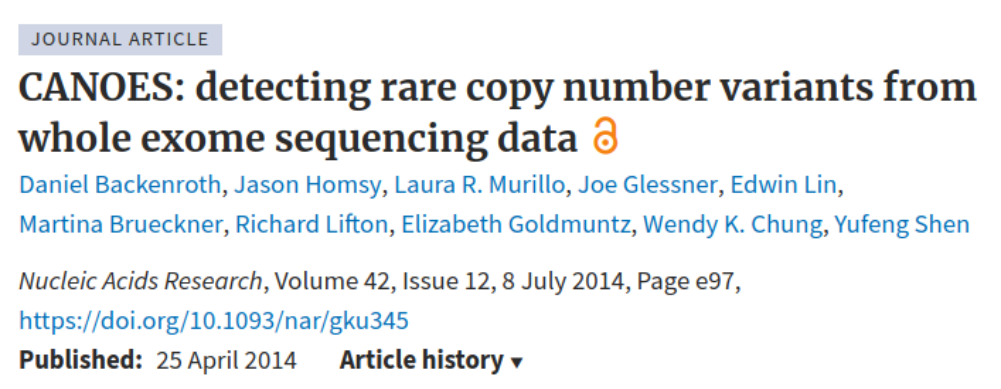
elPrep: High-Performance Preparation of Sequence Alignment/Map Files for Variant Calling

Charlotte Herzeel, Pascal Costanza, Dries Decap, Jan Fostier, Joke Reumers

Published: July 16, 2015 • https://doi.org/10.1371/journal.pone.0132868

- *Example: elPrep, imec*
- Added filter to select perfectly matching reads
- Code audited + feature implemented (Oct. 3d 2017)
- Directly merged into public repository
- New official version released (Oct. 5th 2017)
- Integrated into NIPT **CHU production** (Feb. 2018)
- Fostered a **collaboration** between institutions

Use case 4: New features *Open source*



JOURNAL ARTICLE

CANOES: detecting rare copy number variants from whole exome sequencing data

Daniel Backenroth, Jason Homsy, Laura R. Murillo, Joe Glessner, Edwin Lin, Martina Brueckner, Richard Lifton, Elizabeth Goldmuntz, Wendy K. Chung, Yufeng Shen

Nucleic Acids Research, Volume 42, Issue 12, 8 July 2014, Page e97, https://doi.org/10.1093/nar/gku345

Published: 25 April 2014 **Article history**

- *Example: CANOES, Columbia Genome Center*
- Added structural variant detection on sex chromosomes
- Corrected for pseudo-autosomal regions
- Corrected for maximum likelihood copy number estimation
- *Collaboration* with UZ Antwerpen
- Integrated into **CHU production**
- No contribution back to community due previous lack of **licensing information**
- CANOES is now distributed under MIT License

Advocating for free software

Hospital budget allocation

- Hospitals have very **limited budgets**
- Significant reduction of **software acquisition** costs
- Development of **in-house expertise**
- Software can be modified for **internal applications**
- Software contribution between healthcare institutions **distributes development costs**

Adoption of open standards and open formats

The minimal working block to allow:

Developers $\hookrightarrow$ freedom to innovate

Users $\hookrightarrow$ freedom of software choice

Specifications	Publicly available for free
Process	Procedure open to all for adoption and maintenance

Levels of openness

Java Sun leads the process but other entities can join the format decision process

PDF: Adobe controls the format but publishes it which allows other companies to implement PDF readers

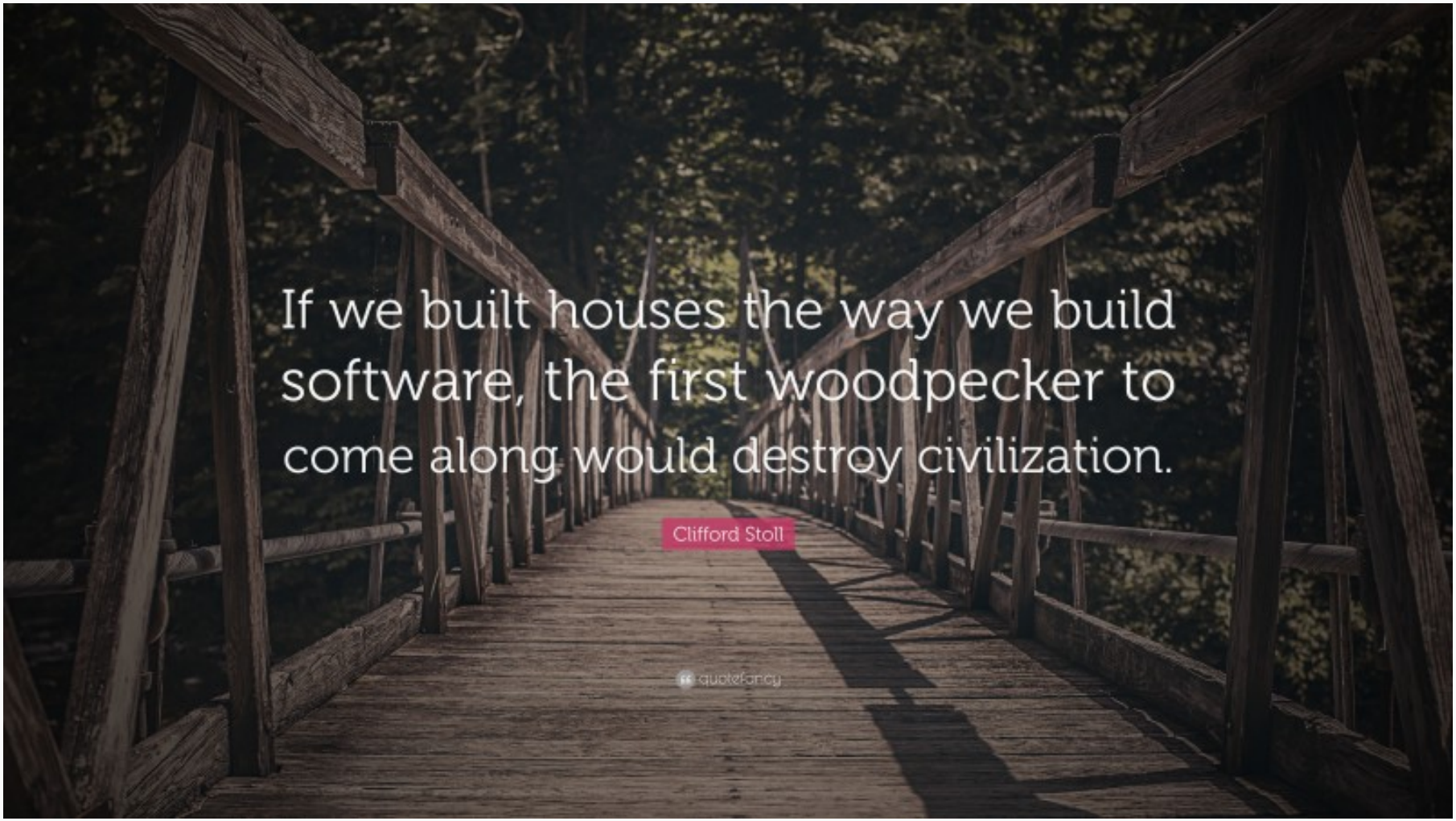
Microsoft C# Submitted to the international standardization organization ISO/IEC 23270

Whatever the level of openness, standards are *de facto* needed to avoid lock-in.

Public scrutiny as a model for accountability

Open data	Citizen data science
Open code	Citizen scrutiny Citizen contribution

- Role of customers in corporate innovation
- Ex: NightScout software by parents of diabetic children
- Ex: DIYPS.org for exocrine pancreatic insufficiency symptom score (developed by a non-healthcare professionel)
- Faster translational medicine?
- Bypasses market research steps


If we built houses the way we build software, the first woodpecker to come along would destroy civilization.
Clifford Stoll
quotefancy

Vincent Bours

Benoit Charloteaux

Victor Grentzinger

Laura Helou

Lennart C. Karssen

Thank you!