

Constructor Influence

BUILD. RE-IDENTIFICATION.

Details Claudia Whitehouse, a developer at Big Data Inc., was given the task of building a web page that would allow an organization's data scientists to combine a number of disparate sources of information together with the aim of gaining greater insight into the organization's customers.

To ensure that what they developed would actually work Whitehouse asked the head of product development, Diane Ermel, for samples from actual customers of Big Data Inc.¹

Ermel, ever conscious of privacy laws, approached two groups, Laws - What Laws! and Reel Estate Submarines, both longstanding customers of Big Data Inc., and asked for de-identified data.^{2,3} Germaine Miller (customer liaison) of Laws - What Laws and Seumas Greer (also customer liaison) of Reel Estate Submarines both agreed to the request and concurred with the assessment of the effectiveness of the proposed anonymization. Both were keen to see the new product.

Whitehouse proceeded with the development and began testing using the data provided. They found that, in many cases, with as little as three non-randomized bits of common data across the disparate sets, the randomized data could be unscrambled and undo the de-identification. Whitehouse informed Ermel and said that they would continue to work on the problem. "Problem, what problem?" asked Ermel. "Surely no one else could possibly think of doing this."

Once Whitehouse discovered that there was no way to prevent this from happening⁴ they proposed that the project be scrapped. This did not go down well with Ermel who had talked up and promised the new web page to many customers, especially, Laws - What Laws and Reel Estate Submarines.

"Continue" instructed Ermel: "No." replied Whitehouse. At this point Susan Wittgenstein, Whitehouse's boss and Ludgwig Haack, the head of HR, became involved. Ermel would not back down neither would Whitehouse⁵.

The project was postponed indefinitely.

Cast	Role
Diane Ermel	Head, Product development
Seumas Greer	Customer liaison of RES
Ludgwig Haack	Head of HR
Germaine Miller	Customer liaison of LWL
Claudia Whitehouse	Big Data Developer
Susan Wittgenstein	Whitehouse's Boss

Table 1: C₂ Cast

¹ For testing purposes only, of course.

² The de-identification consisted of keeping the actual values and randomizing the names, dates of birth, and addresses of the records.

³ Ermel thought that by getting de-identified information from two such different organisations anonymity would be guaranteed.

⁴ It turned out that if the original data existed anywhere in the common data being combined it could be found and reattached to the anonymized data.

⁵ Who threatened to resign, waiving about a piece of paper with visible but unreadable writing.