Making PostgreSQL accessible to your non-technical colleagues

Pavish Kumar R

Mathesar Foundation Inc.,



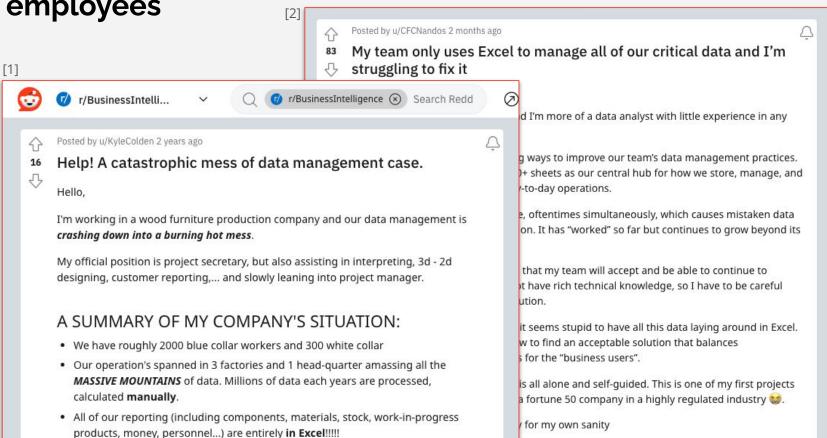
Plan for the talk

- Data management problems faced by non-technical employees.
- How it affects technical employees.
- What it means for the PostgreSQL ecosystem.
- Steps to help your non-technical colleagues.
- > A new open source software project to address this problem: Mathesar.

Terminology

This presentation uses "non-technical" to refer to "non-IT" people. This includes domain experts such as scientists, investment bankers etc.,

Data management problems faced by non-technical employees



Real-life example: Non-technical employee's perspective (1/6)

Reddit Thread: Help! A catastrophic mess of data management case. r/BusinessIntelligence, u/KyleColden, Mar 2022, [1]

"...working in a wood furniture production company and our data management is crashing down into a burning hot mess ... My official position is project secretary ... slowly leaning into project manager"

"Our operation's spanned in 3 factories and 1 head-quarter amassing all the **MASSIVE MOUNTAINS of data**. Millions of data each years are processed, **calculated manually**."



Real-life example: Non-technical employee's perspective (2/6)

"All of our reporting (including components, materials, stock, work-in-progress products, money, personnel...) are **entirely in Excel**!!!!!"

"If there is something wrong ... manually go and find the problems in hundreds of excel files."

"The data between departments ARE NOT LINKED. We send excel files to each other"

"The managers, and also the majority of the staff, do not know where or what to start with."

"Tensions between and inside departments are growing."



Real-life example: Non-technical employee's perspective (3/6)

"What our company desperately needs right now:"

"A tool to **sync data from all the departments and branches**. We have warehouse, exporting team, importing team, factory, fittings, materials,... We really need a tool to sync the data from different teams **and keep them being editable**."

"Linked data. When a data is updated, the others depending on that data will also be updated"

"Access to **history of data changes**, who changed it, at what time"



Real-life example: Non-technical employee's perspective (4/6)

Inferences:

- OP is a domain expert working in a managerial position.
- Doesn't have an IT background.
- Recognizes the need for better data management.
- Company relies heavily on spreadsheets to store and manage all their critical data.
- Data is decentralized and team coordination is tedious.
- High manual labour required to ensure data is valid, organized, and not-duplicated.



Real-life example: Non-technical employee's perspective (5/6)

Some comments on the reddit thread:



tech4ever4u · 2y ago

All this is possible with an in-house app developed by your requirements. At the first iteration, this could be a system that handes existing Excel files (import/export) and keeps one source of truth in the central database (most likely PostgreSql can handle your volumes easily). Once you have all data in a single database, you can start using BI tool(s) to get rid of Excels that are used for data retrieval. Then, by writing data management UIs Excels may go away as a horrible dream.

Build an in-house app, with PostgreSQL as the DB.



Real-life example: Non-technical employee's perspective (5/6)

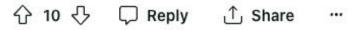
A common comment:

"Hire someone"





It sounds like you need critical rework of how data is handled. Like everyone else has said you need to find a consultant or hire someone who will benefit you in the long run.





JohnHazardWandering • 2y ago

Short term, hire consultants to automate that.



Real-life example: Technical employee's perspective (1/5)

Reddit Thread: My team only uses Excel to manage all of our critical data and I'm struggling to fix it.

r/dataengineering, u/CFCNandos, Nov 2023, [2]

"My manager tasked me with finding ways to improve our team's data management practices. We use **one Excel workbook with 40+ sheets as our central hub** for how we store, manage, and interact with our **data critical to day-to-day operations**."

"Most of our team uses this Excel file, oftentimes simultaneously, which causes **mistaken data entries, conflicting filtering**, and so on."



Real-life example: Technical employee's perspective (2/5)

"My issue is I need to find a solution that my team will accept and be able to continue to maintain should I leave. **They do not have rich technical knowledge**."

"...I don't know how to find **an acceptable solution** that balances sophistication and **user-friendliness for the "business users"**.

"Btw, this is a fortune 50 company in a highly regulated industry"



Real-life example: Technical employee's perspective (3/5)

Inferences:

- OP is technical, works with a team that's non-technical.
- Team uses large spreadsheet as central hub for storage and for managing critical operations.
- Team coordination is tedious, there's conflicts, mistaken data entries etc.,
- Tasked with improving this process.
- Needs to find a solution that the team will accept.



Real-life example: Technical employee's perspective (4/5)

Some comments on the reddit thread:



techmavengeospatial - 3mo ago

I recommend switching to postgresql Developing proper tables and relationships

Then use nocode/lowcode solutions like APPSMITH or Budibase or Tooljet connected to those postgres tables to build search forms/screens and attribute table views And full CRUD apps

Also find nocodb to be useful since it also exposes REST API of the data And even webhooks

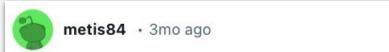
Move to a relational DB, PostgreSQL, and use available user-friendly tools to make it easier for the non-technical people.



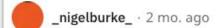
Real-life example: Technical employee's perspective (5/5)

A common comment:

"Quit your job"



TBH - strongly consider leaving. You can nothing good nor useful there, and you need to start your career off on a good



If you're working in a place that uses an excel spreadsheet for all it's data and you're making it up as you go along you might as well give up now before all these bad habits become ingrained







What do these exchanges tell us?

- Never ask for help on Reddit.*
- Spreadsheets are often the go-to tool for data management among non-technical employees.
- Spreadsheets end up being misused for scenarios that require a DBMS.
- People don't streamline their data management process until it's too late.



Why are Spreadsheets misused when DBMS have existed for 6+ decades?

- Non-technical employees are often the real owners of data in an organization.
- The typical ratio of IT to non-IT employees in a company is 1:27. [3]
- Spreadsheets are popular, familiar, and help users hit the ground running.
- There's limited knowledge of DBMS and data management best practices among non-technical employees.
- Spreadsheet companies build features that falsely mimic database functionality to satisfy the growing requirements of these users.



Some popular news stories

- Scientists rename human genes to stop Excel from misreading them as dates. [4]
 - Spreadsheets used to track scientific research and conduct clinical trials.
 - Study from 2016: 1/5th of published papers with genetic data contain Excel errors.
 - Data Integrity & Consistency were missing here.
- Uncontrolled data lead to 16000 UK Covid-19 test results being lost for a week. [5]
 - Size limit in Excel lead to losing over 16k test results.
 - Situations like these affect government policy and is potentially life threatening.
 - Data Durability was missing here.



How it affects technical employees

- Technical employees are:
 - Hired to solve problems with data management.
 - Expected to solve problems with the organizational process.
- All "acceptable" solutions tend to revolve around spreadsheets.
 - Affects technical decisions and goes against good practices.
- Intermediary to data rather than maintainers of technical infrastructure.
 - Querying data to answer questions, producing reports, building inhouse tools.
- Assisting non-technical employees doesn't get recognized as a role of its own.



If non-technical employees used databases directly

- Reduced dependency on IT teams for their day-to-day work.
- Improved communication with IT teams:
 - Effective discussions about data requirements, constraints, and capabilities.
 - People who model the business process will also be able to model the data.
- Better management:
 - Feasible requirements, realistic expectations for timelines and capabilities.
 - Better understanding of system limitations and capabilities
- Technical employees can focus on building technical infrastructure.



What it means for the PostgreSQL ecosystem

- PostgreSQL has existed for 35+ years:
 - o Popular, favoured, and loved among technical people.
 - Unrecognized by non-technical people.
- Primary focus of the ecosystem has been on technical audience, and for good reason.
- Huge unrealized potential among non-technical users.
- Not enough general purpose tools that provide direct access to PostgreSQL's power and functionality to non-technical users.



A bit of history: Ingress during the database wars (1/3)

- Ingress was a database that started as a research project in UC Berkeley.
- Ingress used QUEL.
- In the early 80s, commercial version of Ingress gained popularity.
- Oracle was losing market share to Ingress.
- Oracle 4 introduced a new query language, SQL: A seQUEL to QUEL?
- SQL gained recognition and acceptance. ANSI accepted SQL as the standard and databases followed suit.
- Oracle regained market share. Ingress was out of the database race in a few years.



A bit of history: PostgreSQL comes into existence (2/3)

- In 1985, Michael Stonebraker returned to Berkeley, began work on a post Ingress database. **Post Ingress = POSTGRES**.
- POSTGRES introduced several improvements and used PostQUEL.
- In 1995, POSTQUEL was replaced by SQL, creating Postgres95.
- In 1996, the Postgres95 project was renamed to PostgreSQL to reflect it's support for SQL.
- SQL was so important, several popular relational database projects have `SQL` explicitly mentioned in their names.



A bit of history: QUEL vs SQL (3/3)

- QUEL was considered superior by many.
 - It closely followed Edgar Codd's relational model than SQL.
- QUEL was super composable, SQL was not.
- Why did SQL prevail?
 - SQL was more "english" than QUEL.
 - Did not require mentioning how to query something, only what to query.
 - Welcome by novice users and experienced technical users alike.
 - In other terms, SQL was more accessible.



Addressing the next accessibility concern

- The power of databases needs to be in the hands of users of all skill levels.
- The PostgreSQL ecosystem is best suited to take leaps in this direction.
- Here are some things we can do:
 - More advocacy towards database literacy
 - Community driven general purpose user-friendly interfaces to PostgreSQL
 - Training and documentation aimed towards non-technical people
 - Simplifying migration from existing spreadsheet based workflows to databases
 - MS Access deserves a shoutout.



Steps to help your non-technical colleagues

- Raise awareness:
 - Conduct events, talks, and workshops in your organization.
 - Create content focusing on your non-technical colleagues.
 - Explain when to use databases and when to use spreadsheets.
- Advocate for database literacy:
 - Talk to your management on the cost and benefits of database literacy.
 - Use real-life examples to showcase the dangers of using misusing spreadsheets.
 - Target non-technical managers.



For the enthusiastic

- Help them set up a PostgreSQL database on their machine.
 - EDB offers installers for Mac and Windows. Postgres.app is a simple app for Mac.
 - The options are all clearly documented on the PostgreSQL site.
- Provide them with easy instructions to connect to the database and start querying.
- Here's the hard part: Get them to learn SQL.
 - Getting started with SQL is easier compared to the excel formulas they're used to.
 - Help setup an initial data model for their use case and provide common queries.
 - Resources on SQL are abundant, reducing their dependency on you.



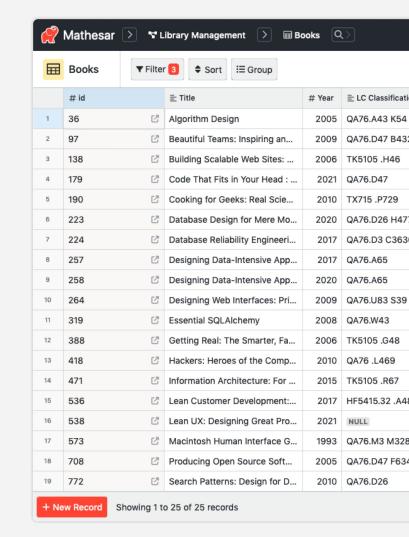
For the less enthusiastic

- The best way to tackle them is to install and manage the database yourself.
- Rely on FOSS tools and the PostgreSQL ecosystem. Do not build custom solutions.
- Do not simplify database terminology, let them get familiar with the terms.
- Some tools for reports/visualization:
 - Metabase, Evidence, Apache Superset
- Some tools for Dashboards/UI/Data entry:
 - Mathesar, NocoDB, AppSmith, Budibase
- Find a list of tools <u>here</u>.

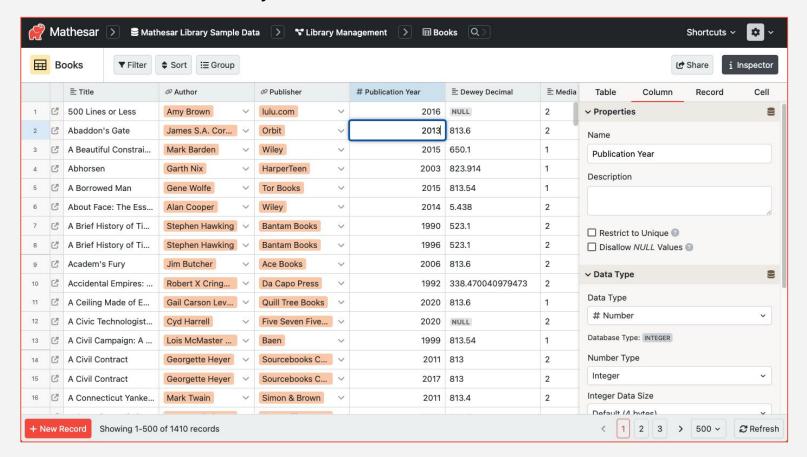




- Powered by PostgreSQL, aims to make databases accessible to users of all skill levels.
- Mascot is a young and friendly elephant,
 symbolic of what we're trying to do. :)
- Open source, Community-led, Non-profit.
- Connect to your PostgreSQL database and it works right away!

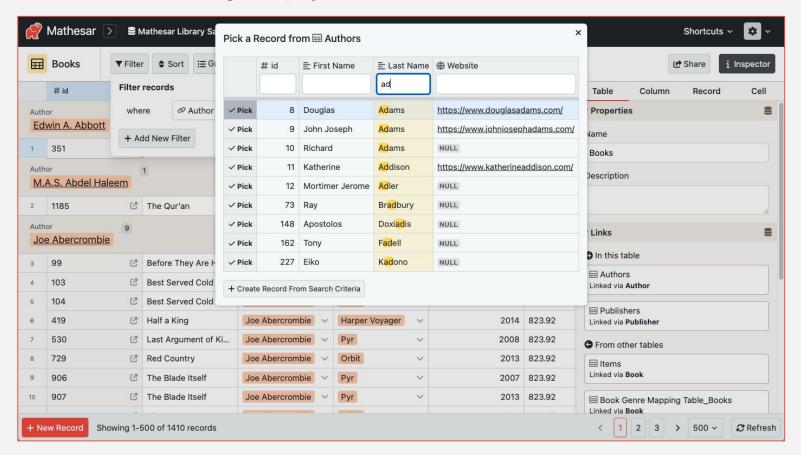


Perform data entry



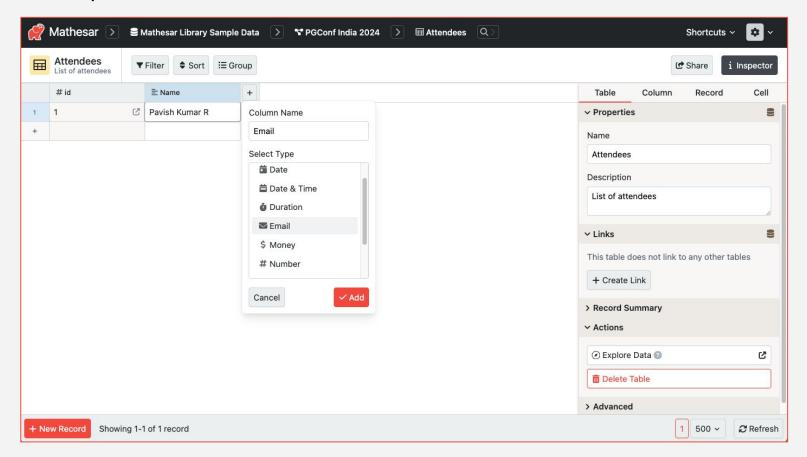


Filter, sort, and group your data



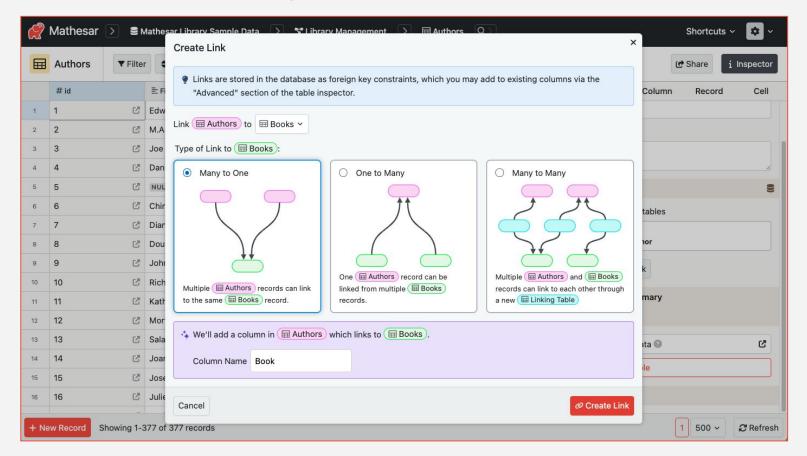


Set up data models



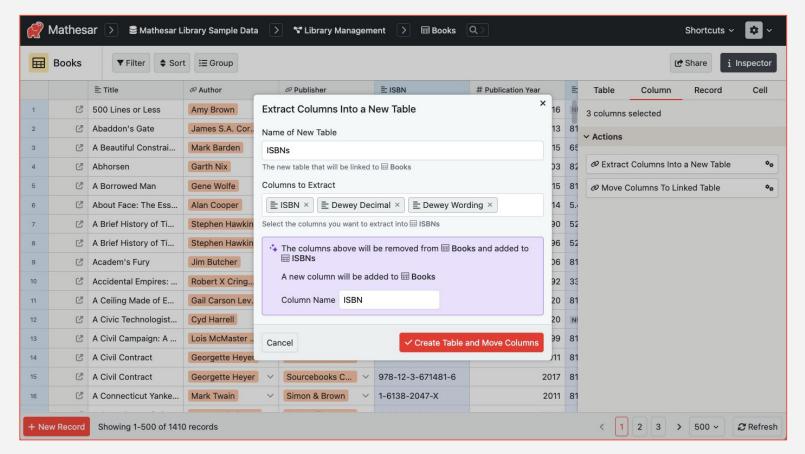


Establish relationships



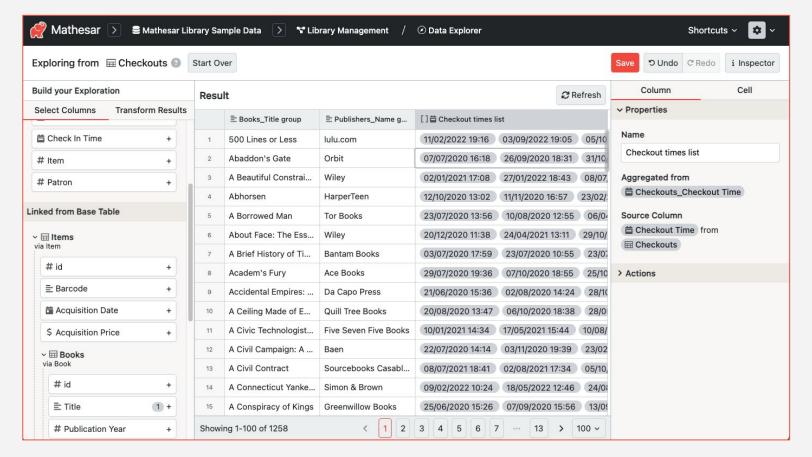


Perform schema migrations





Join tables, explore data, and construct reports





Emphasis on Interoperability

- General purpose, direct interface to PostgreSQL.
- Works on top of your existing infrastructure.
- High degree of interoperability with other tools on the ecosystem.
- Does not have its own set of abstractions.
- Does not reinvent terminologies.



Our journey so far

- In active development for 2+ years.
 - ~2100 GitHub stars, 90+ contributors.
- Launched our MVP on Hacker News in March 2023.
 - 24 hours on front page, 75k+ views.
- We're still in our alpha stage, with 5 releases.
- We've been talking to early adopters and users from all sorts of backgrounds.
- Inching closer towards our beta.



How you can help

- Check out Mathesar.
 - Install it, attempt to use it, star us on GitHub if you appreciate it.
- Talk to us.
 - Tell us if you think Mathesar can help solve a problem for you.
 - Every comment helps us understand your use case and prioritize our work.
- Let us know if you'd like a new feature.
- Become an early adopter.
- Contribute to the repo.



Thank you!

Pavish Kumar R pavish@mathesar.org



References

- 1. https://www.reddit.com/r/BusinessIntelligence/comments/tbke1f/help-a-catastrophic mess-of-data-management-case/
- 2. https://www.reddit.com/r/dataengineering/comments/186zm6c/my team only uses excel to manage all of our/
- 3. https://workforce.com/news/ratio-of-it-staff-to-employees
- 4. https://www.theverge.com/2020/8/6/21355674/human-genes-rename-microsoft-excel-misrea ding-dates
- 5. https://www.bbc.co.uk/news/technology-54423988
- 6. https://eusprig.org/research-info/horror-stories/

