UNIVERSITÄT
LEIPZIG

Tales of doing Research with Video Game Fan Databases

A data-driven Approach

Nottingham, 21.08.2018

Peter Mühleder
Tracy Hoffmann

DFG

UNIVERSITÄTS BIBLIOTHEK ubl LEIPZIG

Ostasiatisches Institut der Universität Leipzig

diggr.link

Cooperation between University Library Leipzig and Japanese Studies (Institute of East Asian Studies) of Leipzig University

Q: How can we use online data sources for research on (Japanese) video games?

Agenda

Video game fan databases as data sources
Linking data sources
Conceptual issues
Conclusion

Fan Databases

- Fans collect, organize and share an enormous amount of information about video games.

  e.g. Mobygames, GameFAQs, Wikia

- Highly specialized communities
- (Mostly) easily accessible

Fan Databases

- Metadata
  Company and developer credits, technical information, ...

- Discursive data
  User reviews, discussions, walkthroughs, ...

- Community practices

Data Sources

- Different sources provide different information

Data Source	Records	Language	Scope	Japanese Release Date(s)	Credits	Companies	Alternative Titles	Links to Knowledge Base	Walkthroughs
Media Art DB	38.068	Jp	Japan						
Mobygames	81.609	En	Worldwide						
GameFAQs	55.834	En	Worldwide					(Wikipedia)	

=> Data integration

Challenges

- No unique identifiers
- Datasets in different languages
- Heterogeneous data models

Q: How can we create links between these data sources?

Record linking

Linking based on game titles

=> Fan databases often feature alternative titles (in different languages)

=> Game title matching algorithm:

- preselection based on platform
- probabilistic ratio based on title similarity
- extraction and comparison of numbers; subtitles
- basis for machine learning model

https://github.com/diggr/diggrtoolbox

Result: Match probability 0 <= x <= 1.0

[Mobygames entry X] ← 0.9 → [Media Art DB entry Y]

Linking results / clusters

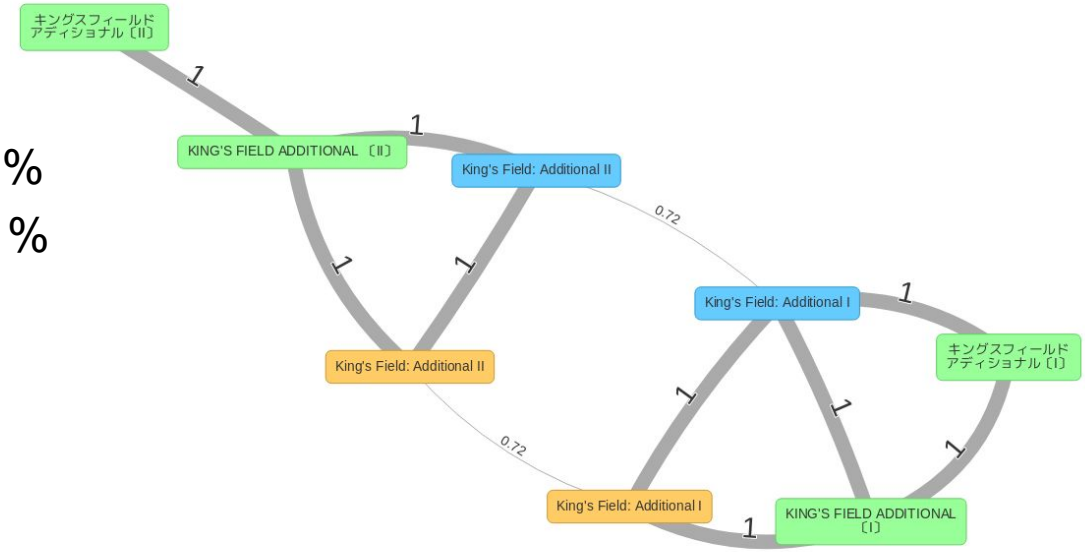
Media Art DB -> Mobygames: 48 %

Media Art DB -> GameFAQs: 85 %

Linking results / clusters

Media Art DB -> Mobygames: 48 %

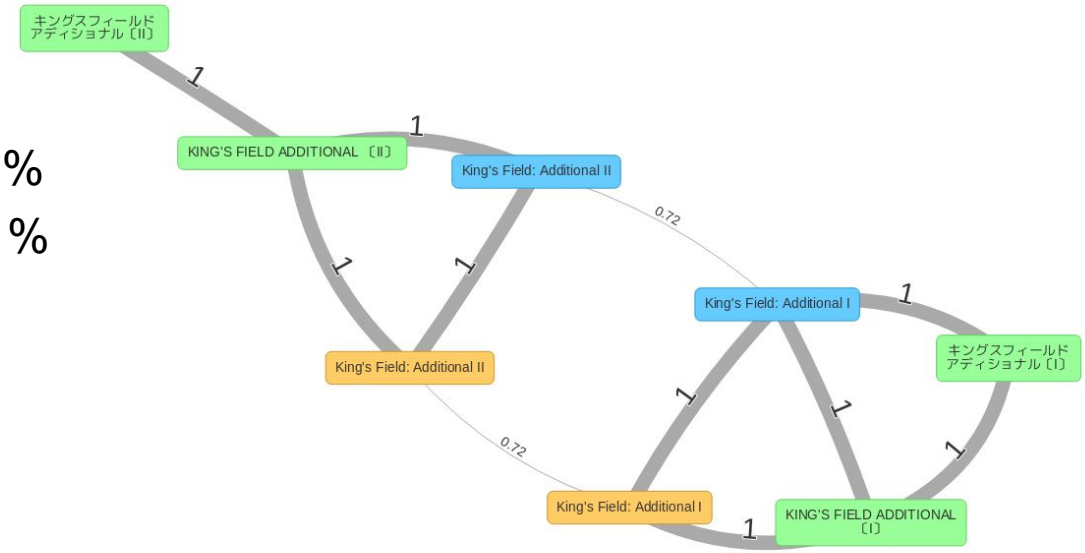
Media Art DB -> GameFAQs: 85 %



Linking results / clusters

Media Art DB -> Mobygames: 48 %

Media Art DB -> GameFAQs: 85 %



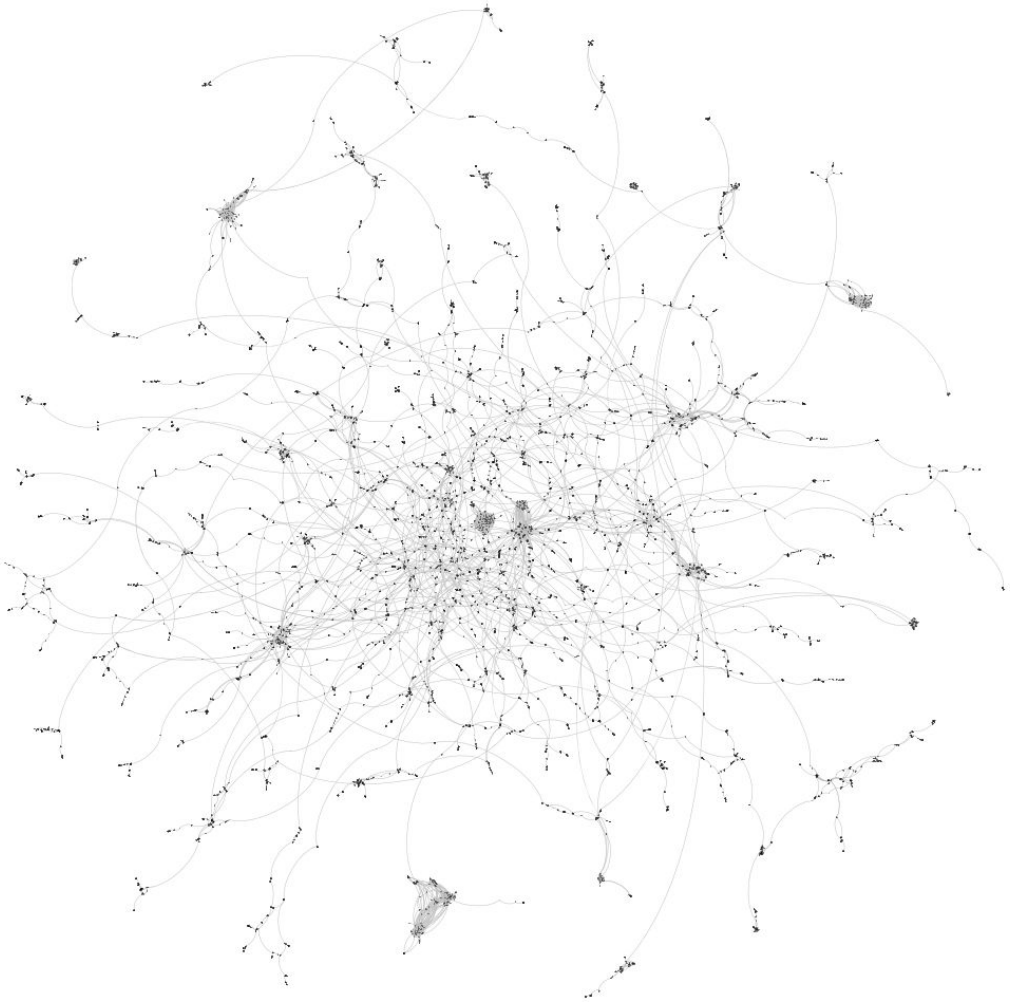
Visualization as network:

- Easy to identify wrong links
  Example
- Helps to identify problematic data
  Example

Linking results / clusters

Big cluster with ~ 9000 Datapoints (games)

Big franchises such as Super Mario, Final Fantasy, The Legend of Zelda, ...

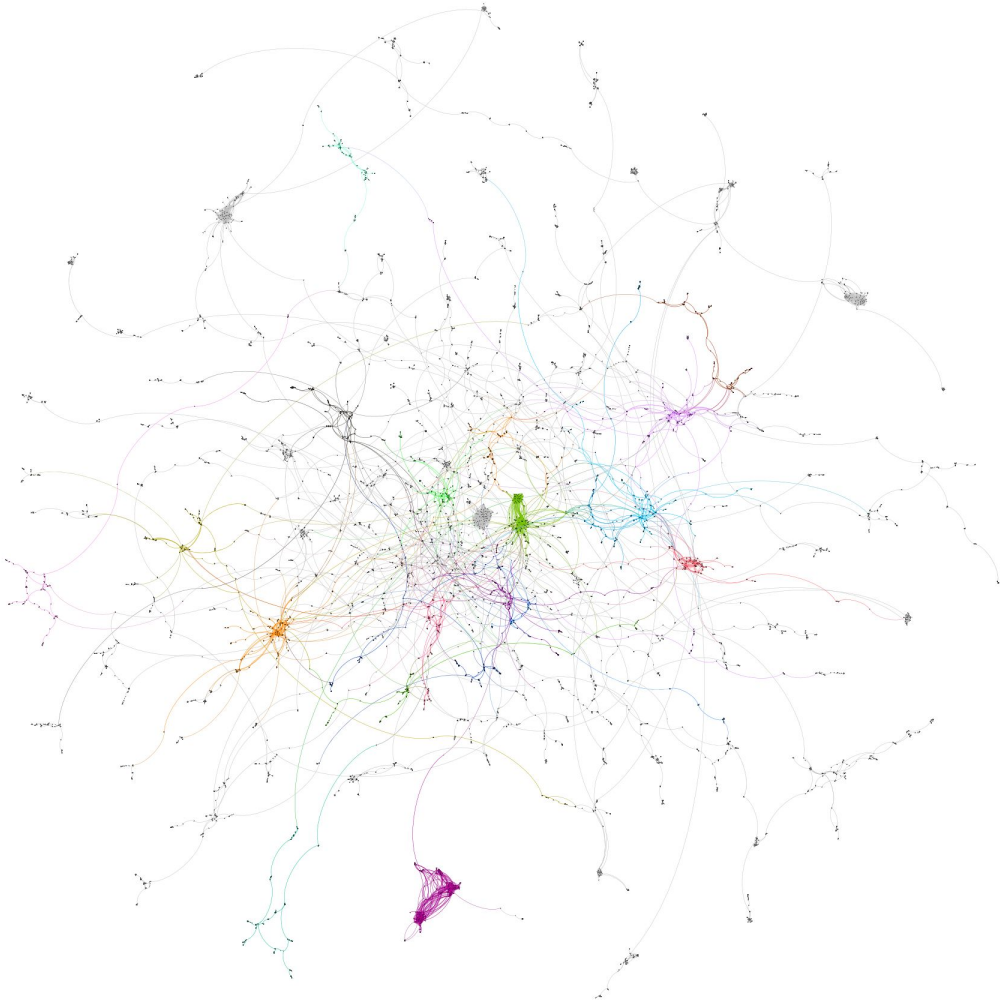


Linking results / clusters

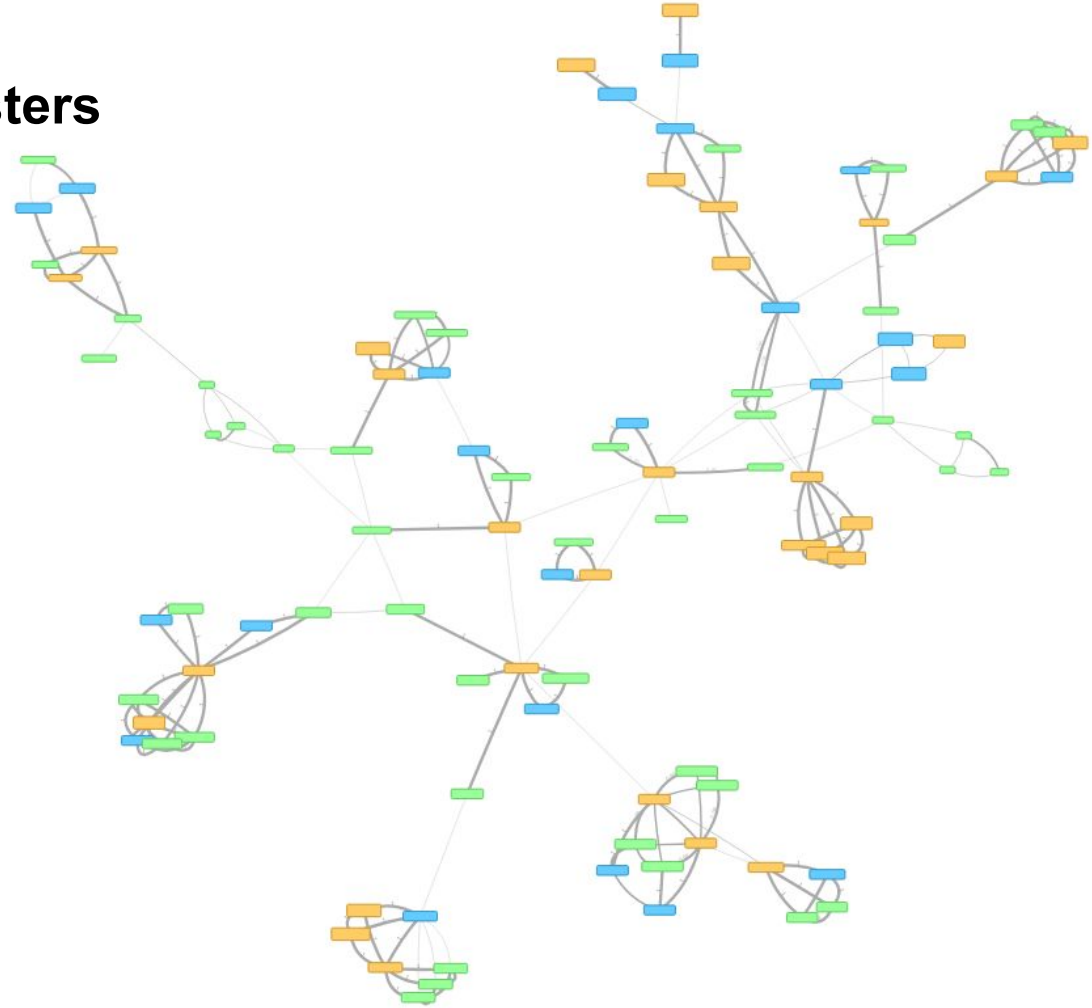
Big cluster

=> community detection algorithm

- 144 smaller clusters
- Mostly focused on a specific game/series



Linking clusters



Link

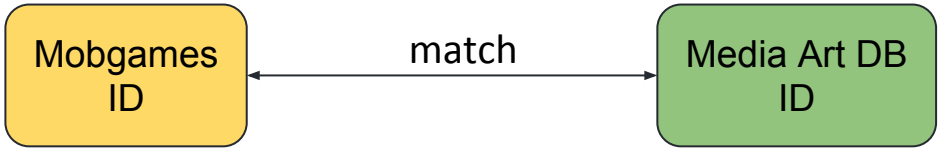
Linking results / clusters

Problems:

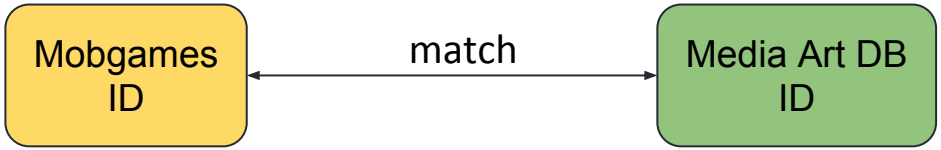
- Missing links - “We don’t know what we don’t know”
- Link ratios are not very expressive

Behind the links

Conceptual Issues



“This Mobygames ID has a match with that Media Art DB ID”



“This Mobygames ID has a match with that Media Art DB ID”

“The release information of a Mobygames ID are the same as in this Media Art DB ID”

“This game in Mobygames is the same game as that Media Art DB ID”

Thoughts about links/matches/cluster

- Links have no semantics (at this state)
- Different goals and perspectives causes heterogeneous data models
- Researcher ask for a “game”

Thoughts about links/matches/cluster

- Links have no semantics (at this state)
- Different goals and perspectives causes heterogeneous data models
- Researcher ask for a “game”

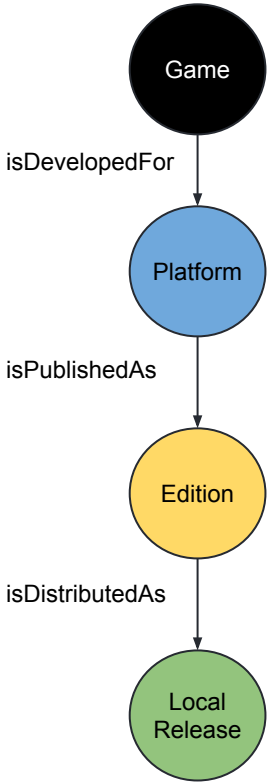
Record ≠ Game

Mobygames => New Edition = New Record

Media Art DB => New Release = New Record

GameFAQs => New Platform = New Record

Conceptual entities of video games



(Common) Game title

Nier

Platform
Game title for specific platform

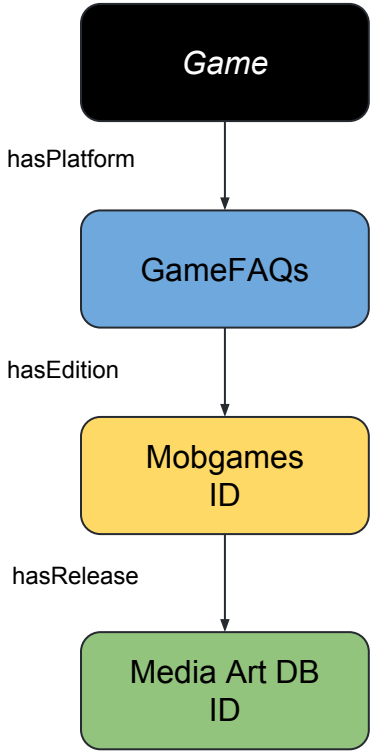
XBox 360
NieR Gestalt

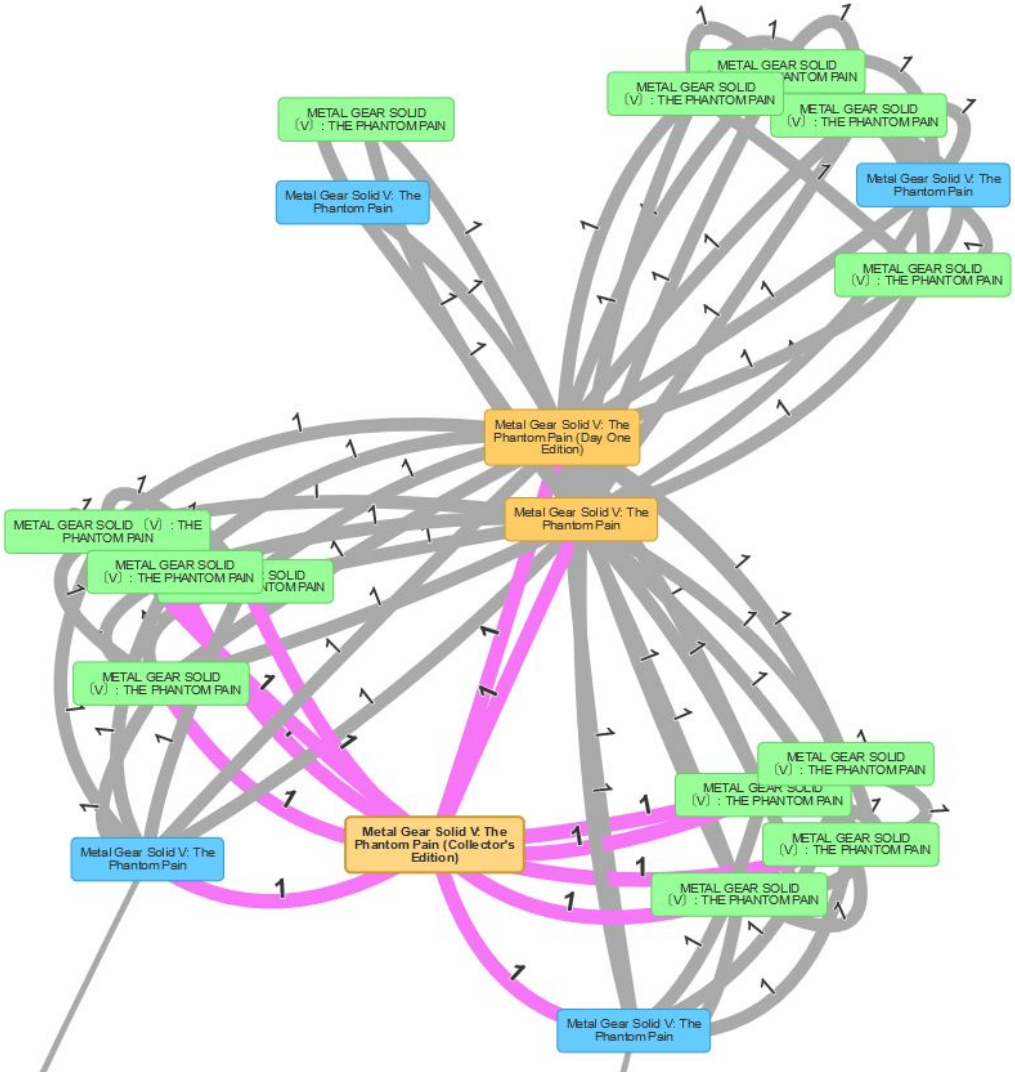
Title variations (Origin title, Edition title)
Variations in packaging (special cases, origin game + add ons, ...)

NieR Gestalt (Origin game)
NieR Gestalt (Platinum collection)

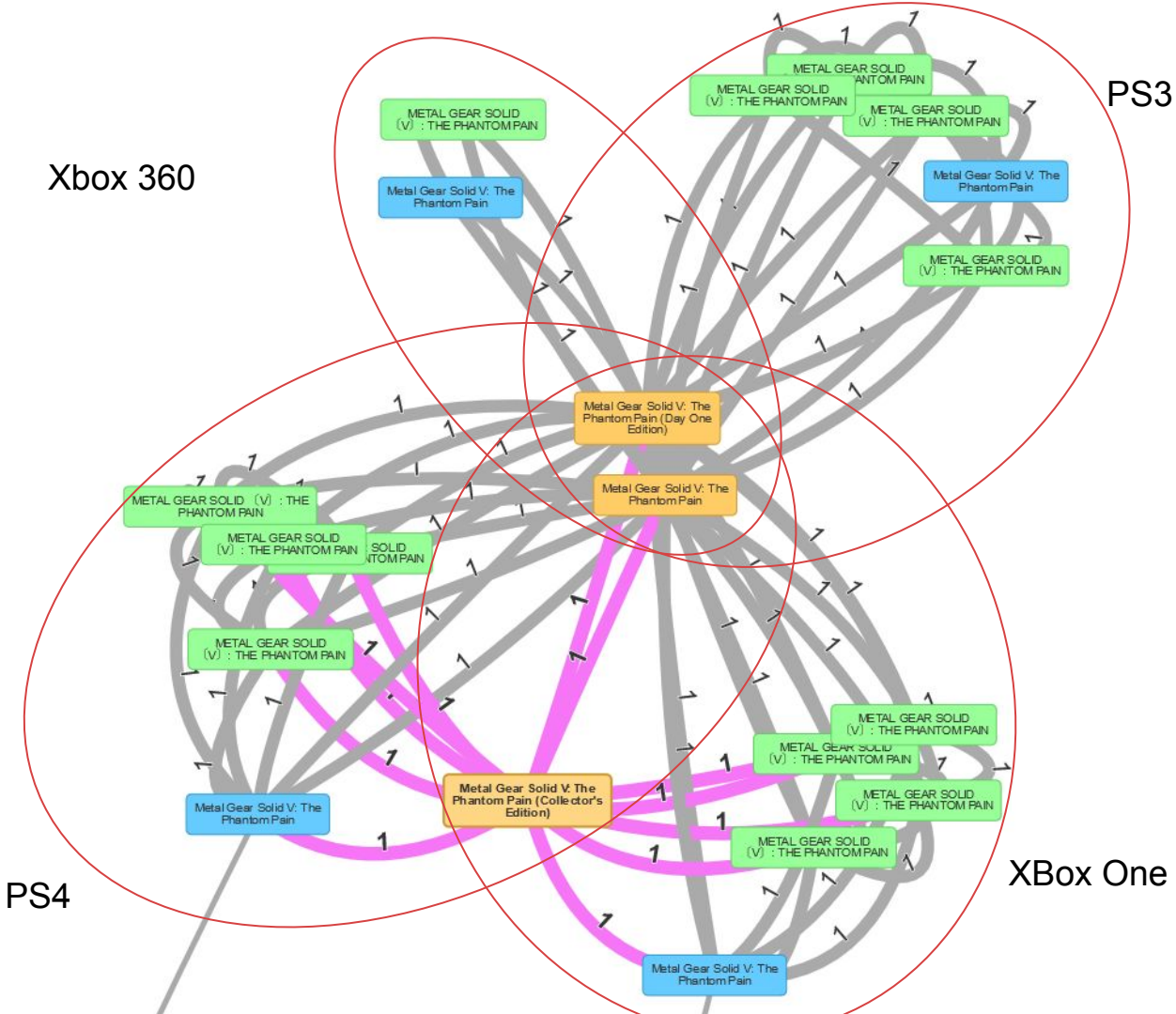
Release data (date and region/country)
Release titles (depends on language)

04/22/10
04/21/11
ニーアゲシュタルト

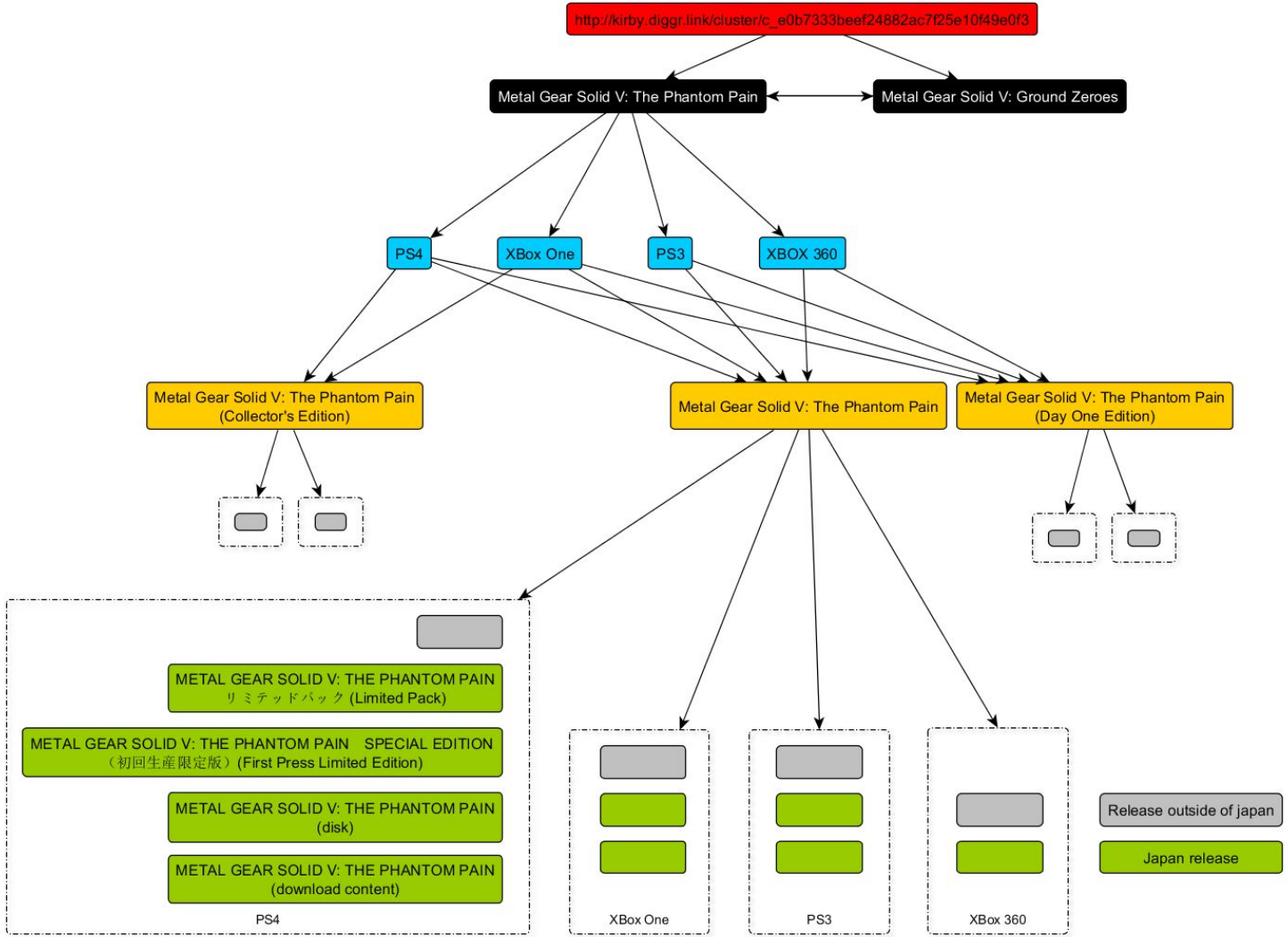

Game
hasPlatform
GameFAQs
hasEdition
Mobgames
ID
hasRelease
Media Art DB
ID

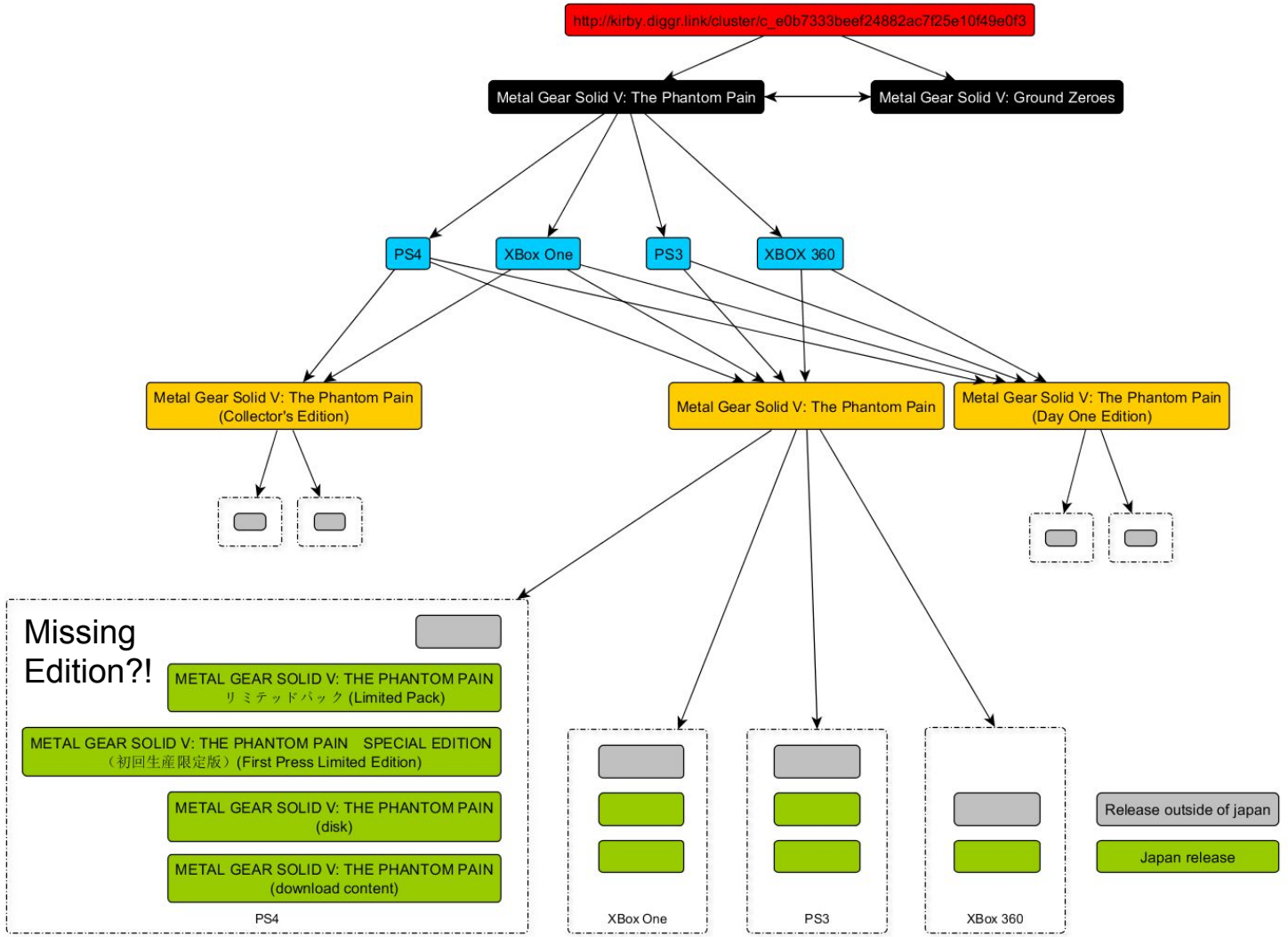

METAL GEAR SOLID (V) : THE PHANTOM PAIN
METAL GEAR SOLID (V) : THE PHANTOM PAIN
METAL GEAR SOLID (V) : THE PHANTOM PAIN
Metal Gear Solid V: The Phantom Pain
Metal Gear Solid V: The Phantom Pain
METAL GEAR SOLID (V) : THE PHANTOM PAIN
Metal Gear Solid V: The Phantom Pain (Day One Edition)
Metal Gear Solid V: The Phantom Pain
METAL GEAR SOLID (V) : THE PHANTOM PAIN
METAL GEAR SOLID (V) : THE PHANTOM PAIN
METAL GEAR SOLID (V) : THE PHANTOM PAIN
METAL GEAR SOLID (V) : THE PHANTOM PAIN
Metal Gear Solid V: The Phantom Pain (Collector's Edition)
Metal Gear Solid V: The Phantom Pain
METAL GEAR SOLID (V) : THE PHANTOM PAIN
METAL GEAR SOLID (V) : THE PHANTOM PAIN
Metal Gear Solid V: The Phantom Pain

Link


Xbox 360
PS3
PS4
XBox One
METAL GEAR SOLID (V) : THE PHANTOM PAIN
Metal Gear Solid V: The Phantom Pain
Metal Gear Solid V: The Phantom Pain (Day One Edition)
Metal Gear Solid V: The Phantom Pain (Collector's Edition)

Link


http://kirby.diggr.link/cluster/c_e0b7333beef24882ac7f25e10f49e0f3
diggr Cluster
Metal Gear Solid V: The Phantom Pain
Metal Gear Solid V: Ground Zeroes
PS4
XBox One
PS3
XBOX 360
Gamefaqs
Metal Gear Solid V: The Phantom Pain (Collector's Edition)
Metal Gear Solid V: The Phantom Pain
Metal Gear Solid V: The Phantom Pain (Day One Edition)
Mobygames
METAL GEAR SOLID V: THE PHANTOM PAIN リミテッドパック (Limited Pack)
METAL GEAR SOLID V: THE PHANTOM PAIN SPECIAL EDITION （初回生産限定版）(First Press Limited Edition)
METAL GEAR SOLID V: THE PHANTOM PAIN (disk)
METAL GEAR SOLID V: THE PHANTOM PAIN (download content)
PS4
XBox One
PS3
XBox 360
Release outside of japan
Japan release
Media Art DB

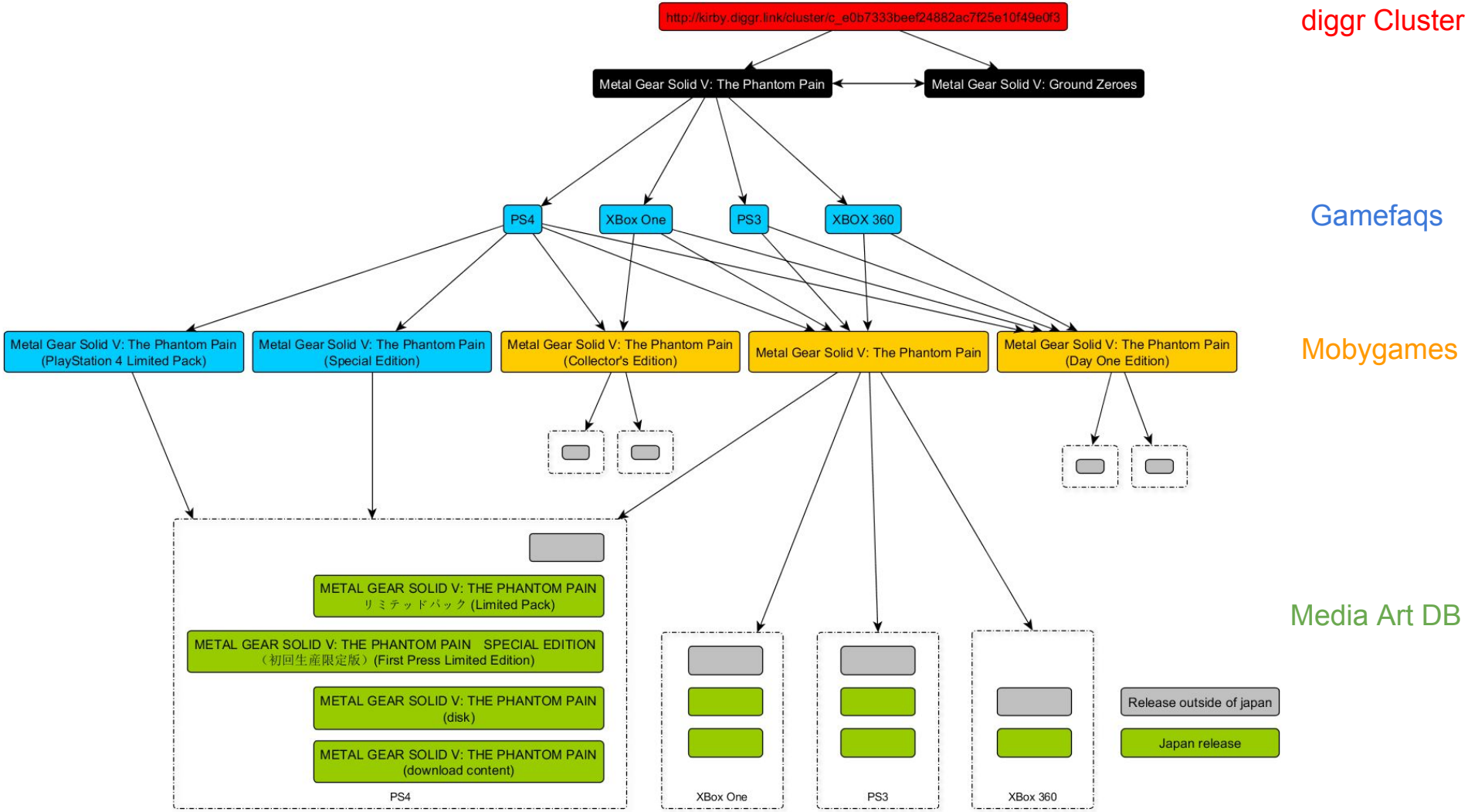


diggr Cluster

Gamefaqs

Mobygames

Media Art DB


http://kirby.diggr.link/cluster/c_e0b7333beef24882ac7f25e10f49e0f3
diggr Cluster
Metal Gear Solid V: The Phantom Pain
Metal Gear Solid V: Ground Zeroes
PS4
XBox One
PS3
XBOX 360
Gamefaqs
Metal Gear Solid V: The Phantom Pain (PlayStation 4 Limited Pack)
Metal Gear Solid V: The Phantom Pain (Special Edition)
Metal Gear Solid V: The Phantom Pain (Collector's Edition)
Metal Gear Solid V: The Phantom Pain
Metal Gear Solid V: The Phantom Pain (Day One Edition)
Mobygames
METAL GEAR SOLID V: THE PHANTOM PAIN リミテッドパック (Limited Pack)
METAL GEAR SOLID V: THE PHANTOM PAIN SPECIAL EDITION （初回生産限定版）(First Press Limited Edition)
METAL GEAR SOLID V: THE PHANTOM PAIN (disk)
METAL GEAR SOLID V: THE PHANTOM PAIN (download content)
PS4
XBox One
PS3
XBox 360
Media Art DB
Release outside of japan
Japan release

Video game data models

- clarify what you talking about
- no “one model to rule them all” -> depends on research question

Ongoing development

- add some semantics to the links (eg. belongsToSameSeries, sameTitleAs)
- create the game entity

Conclusion

By combining and using online video game databases we can

=> Build better research datasets
=> Build a video game reference dataset

But:
=> Automatic linking still has room for improvements
=> Conceptual model is required that can incorporate the different data models
=> It’s a long-term project

UNIVERSITÄT
LEIPZIG

Thank you!

https://diggr.link/
https://github.com/diggr