

Analysis of pre-2020 submission according to the new reporting guidelines

*EEA & ETC-ATNI work
compiled by Luca Liberti*



Presentation Plan

- Status of the submissions
- Analysis of the links in H-K
- Conclusions and future plans

Status of the submissions – The Problem

- We want to see how many successful deliveries we have in previous years according to the new rules
- How do we define a successful submission?
- Dataflow H-K are tightly connected
- Submission rules and QA changed over time
- Flexible approach
- Process as many deliveries as possible
- Keep track of the issues



Status of the submissions – Definitions/1

- We take all the files delivered by a country in H-K for a given reporting year
- Each XML file is classified according to its level of compliance to the rules:
 1. The file content cannot be imported
 2. The file is imported but not processed
 3. Some of the links cannot be resolved
 4. All the links are resolved

Status of the submissions – Definitions/2

- We rank each dataflow according to the lowest score of the files included in the delivery.

Example:

two files were submitted for AQPlans one is correctly imported and linked, the other is not imported

Dataset H is considered “not imported”

Status of the submissions – Definitions/3

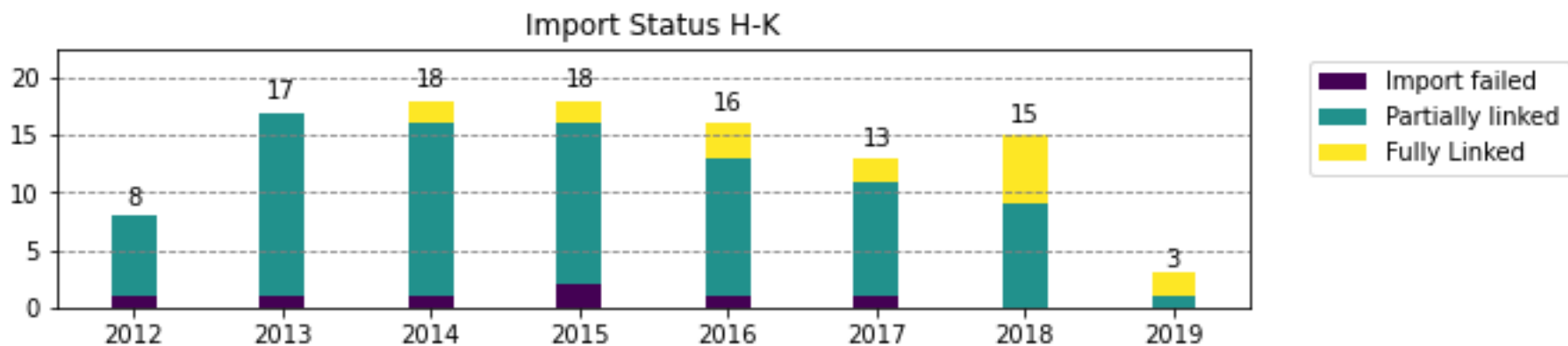
- We rank each H-K submission according to the lowest rank assigned to its dataflows
- We only try to build links if all the dataflows were imported

Example:

Dataflows H, I, J for country X 2018 are imported and fully linked and K is partially linked

Country X 2018 delivery is considered “partially linked”

Status of the submissions – Results



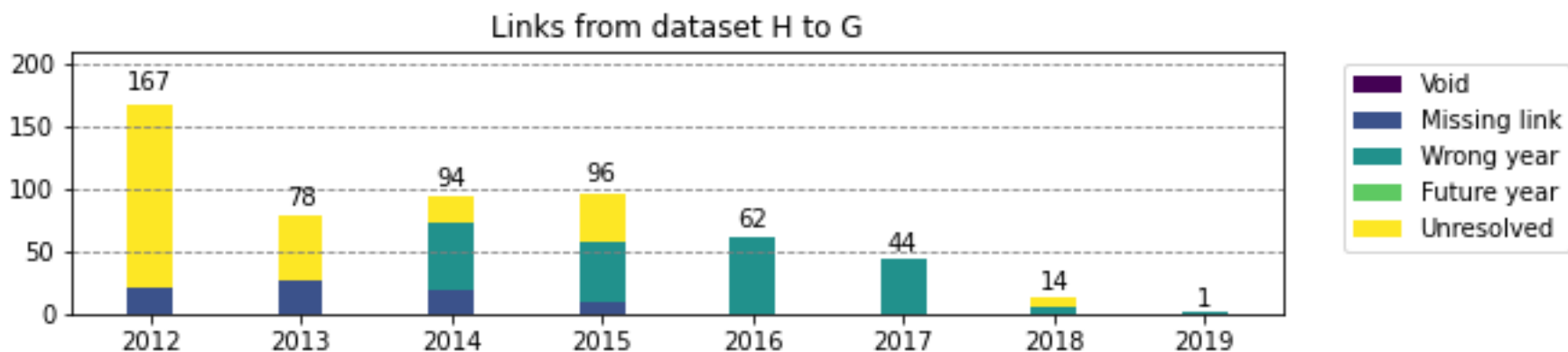
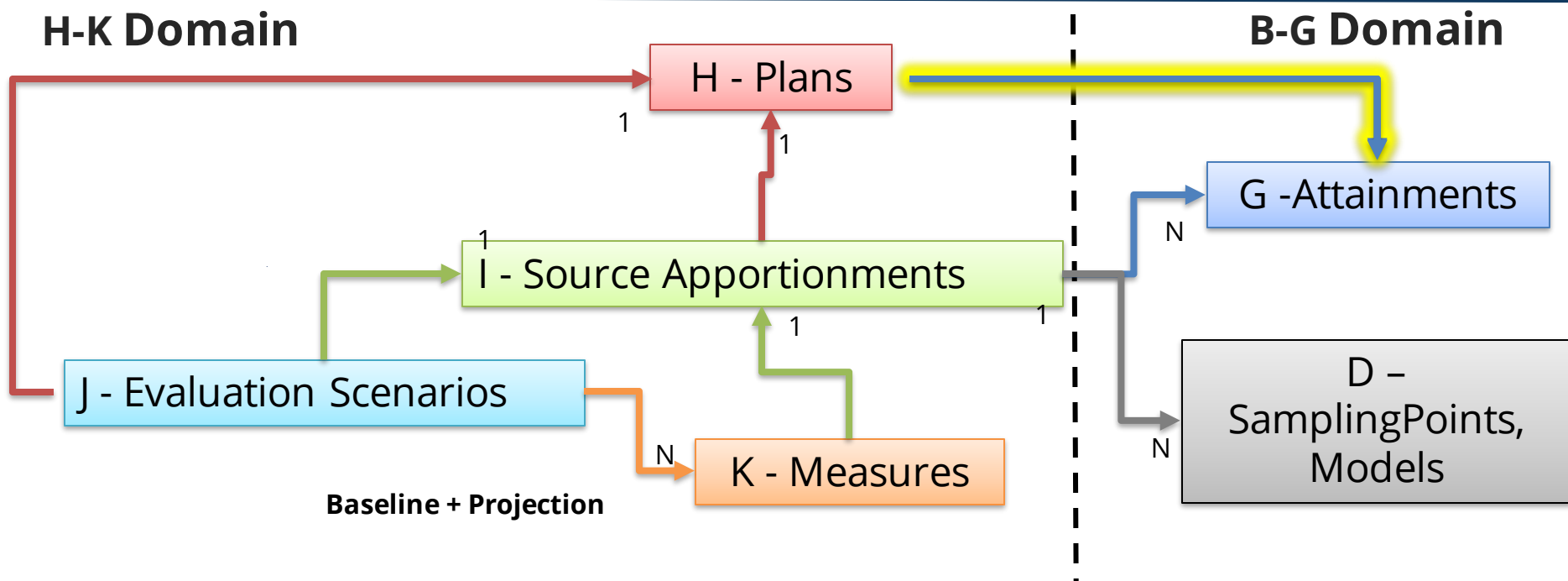
Link Analysis – Overview

- **We try to establish the links between dataflows according to latest reporting rules**
- **We apply the rules to all the submissions including the ones done in the past**
- **When the link cannot be created we identify different types of problems**

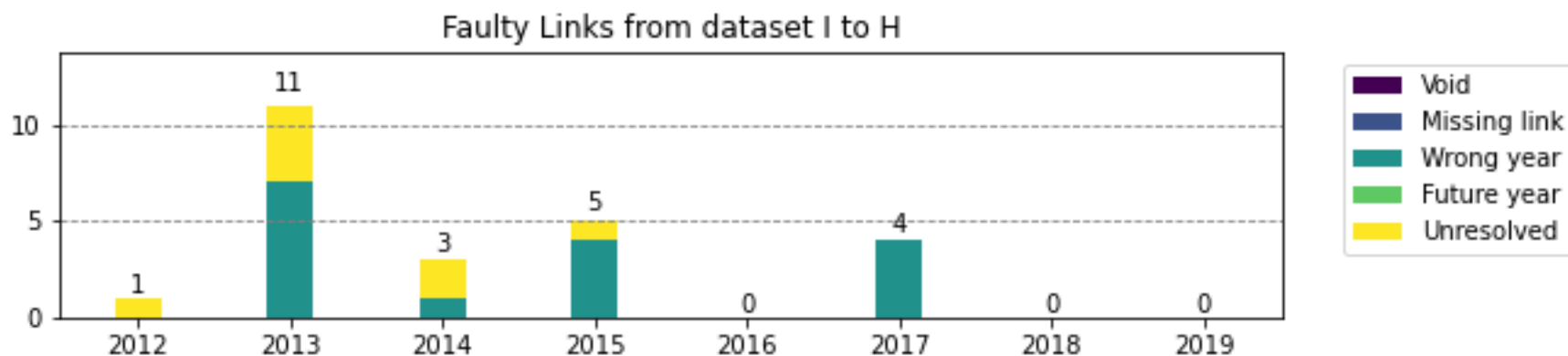
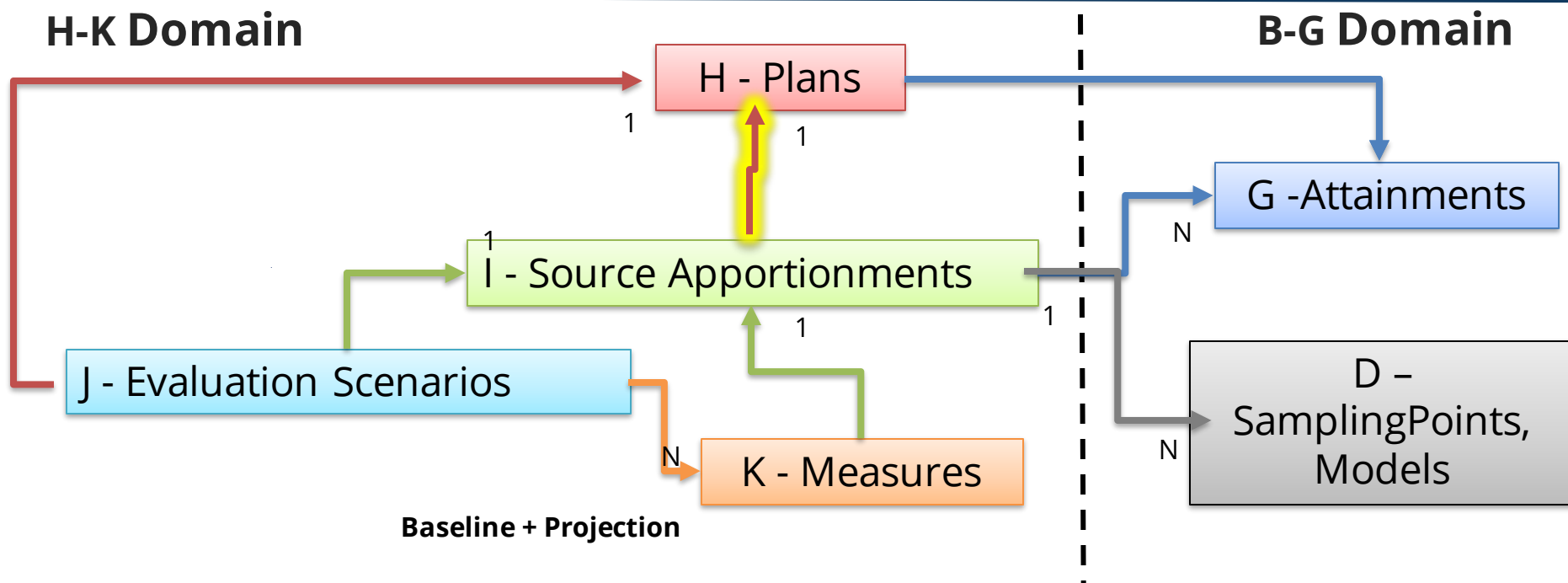
Link Analysis – Classification

- **A mandatory link is not provided or it is provided as empty XML tag: VOID**
- **When many links are expected and none is found: MISSING LINK**
- **An element linked using an identifier cannot be found: UNRESOLVED**
- **An element is found in a year different to the reporting year: WRONG YEAR**
- **An element is found in a future year : FUTURE YEAR**

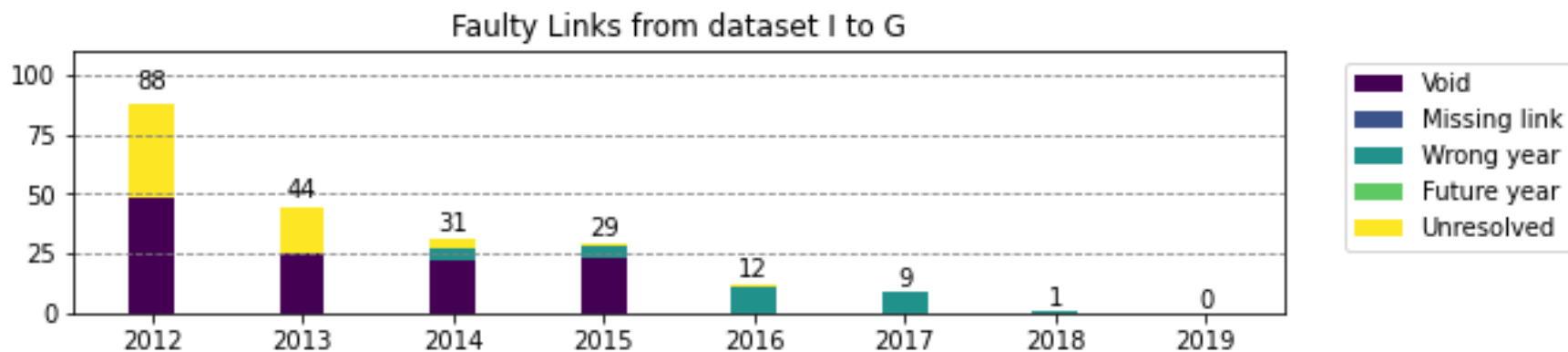
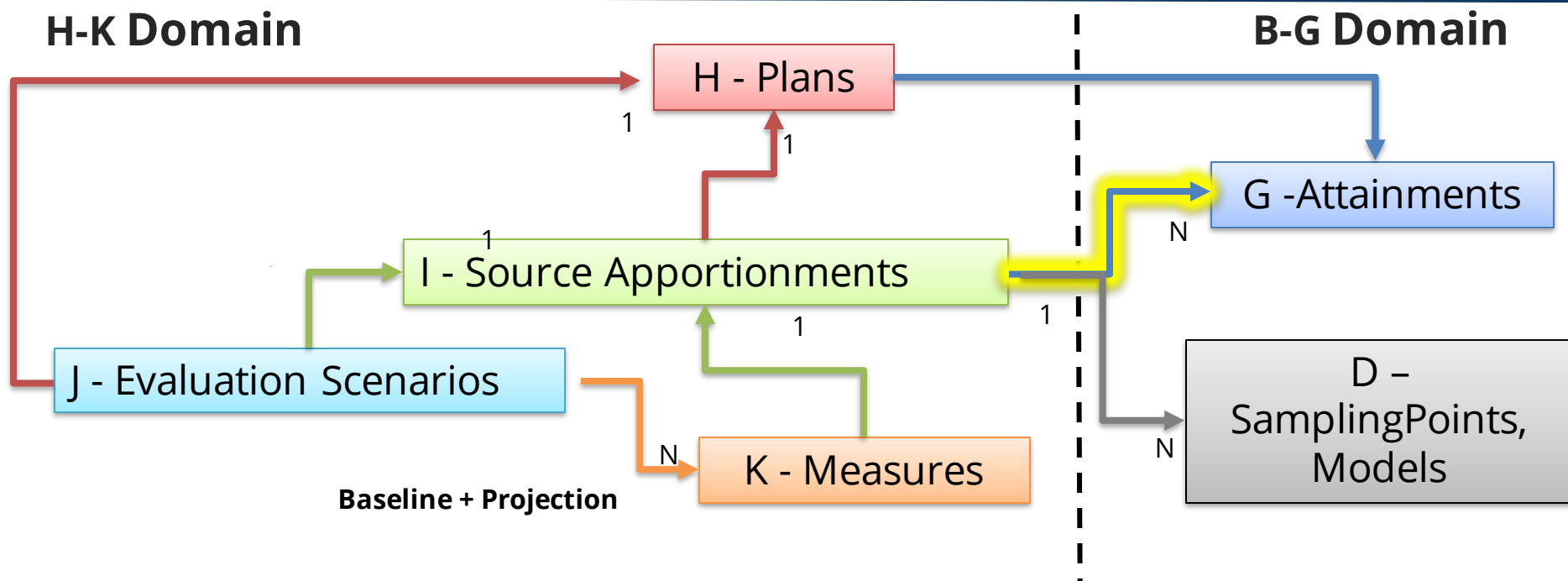
Link Analysis – H to G



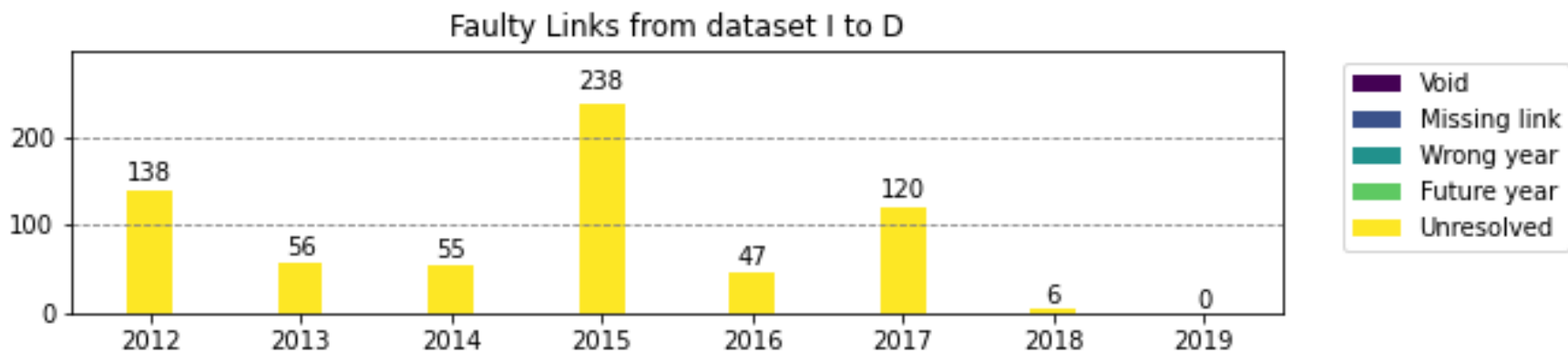
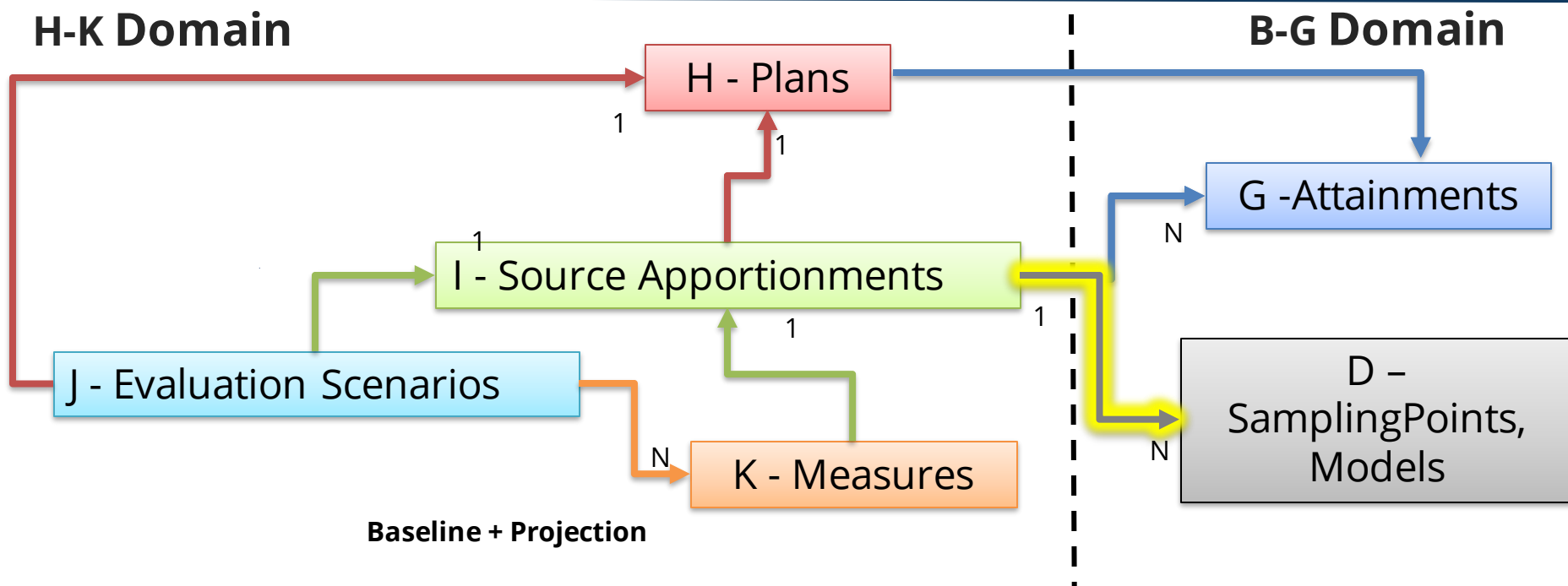
Link Analysis – I to H



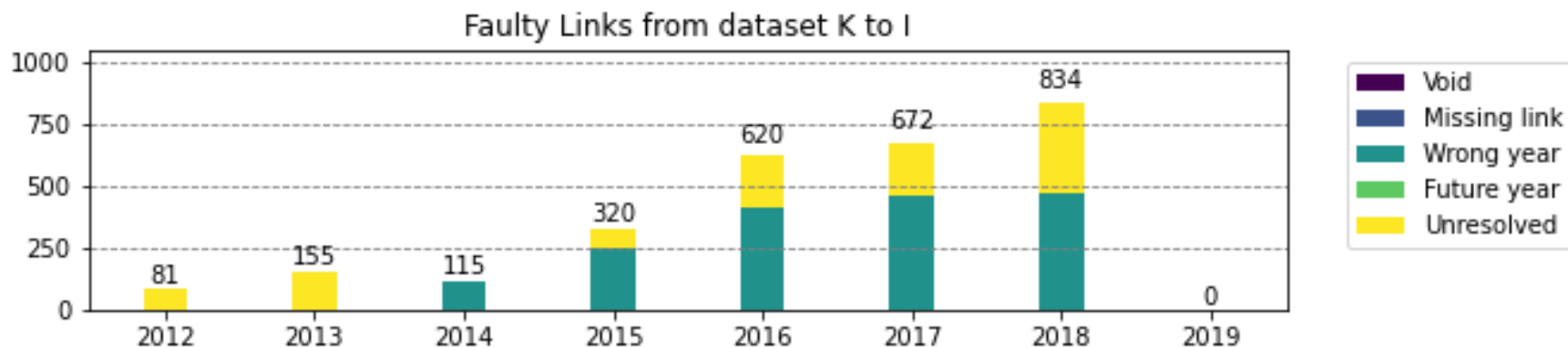
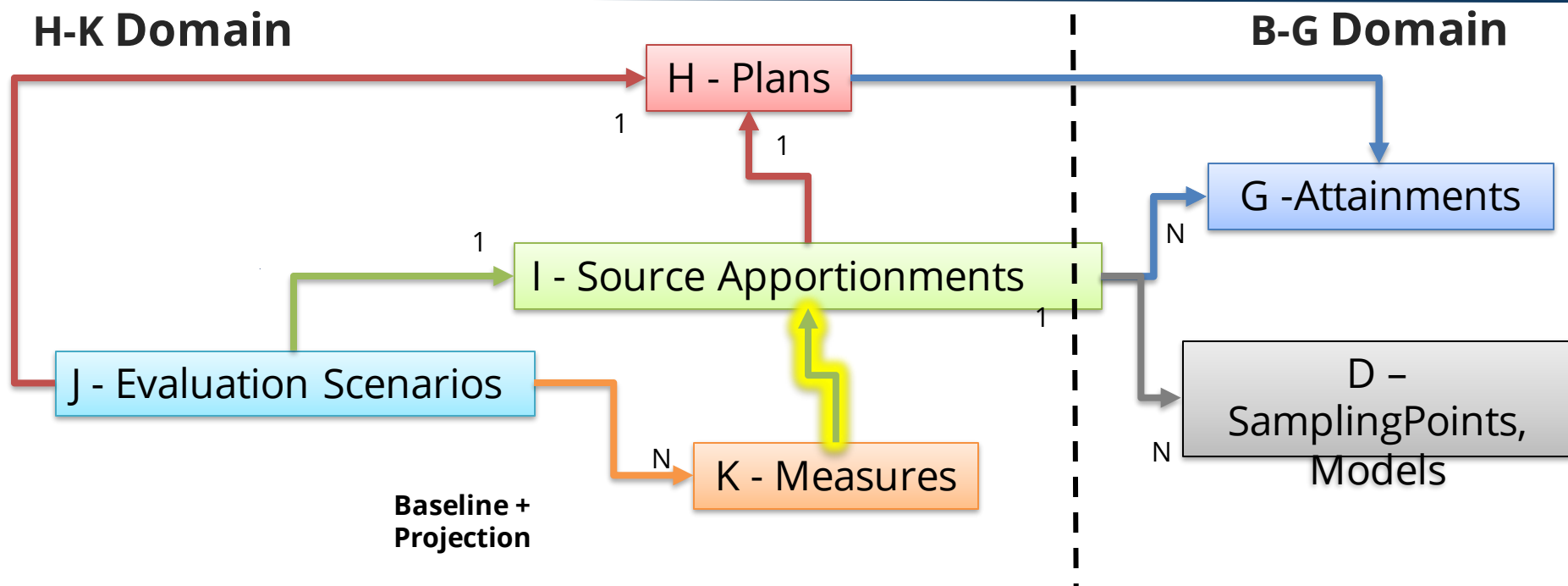
Link Analysis – I to G



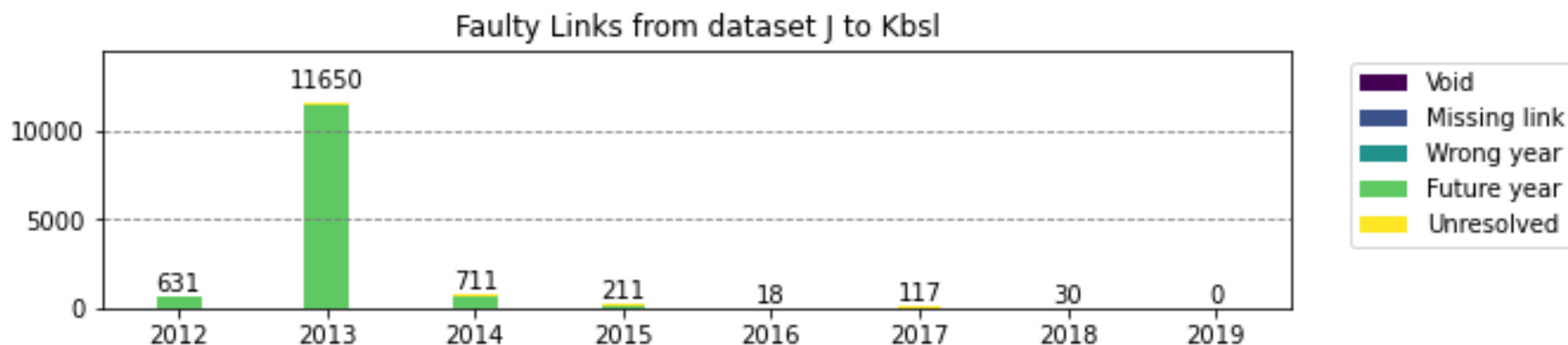
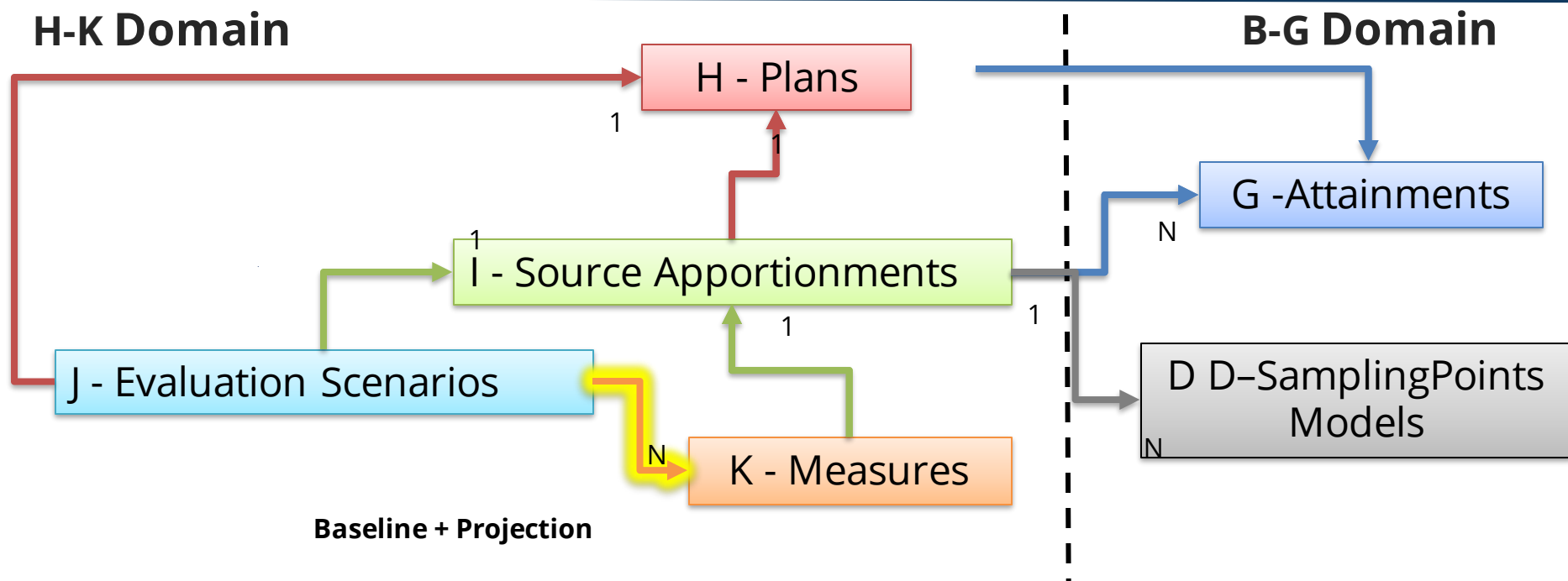
Link Analysis – I to D



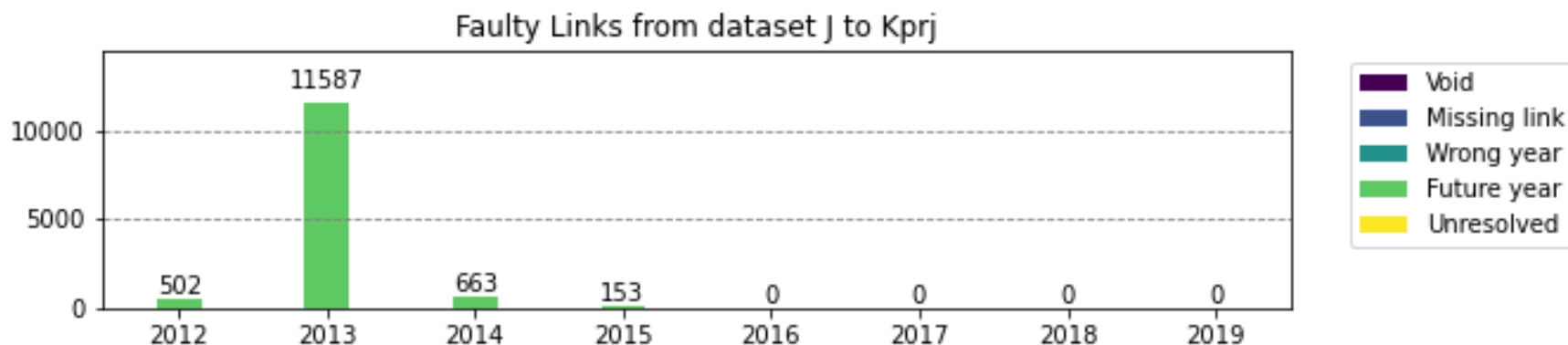
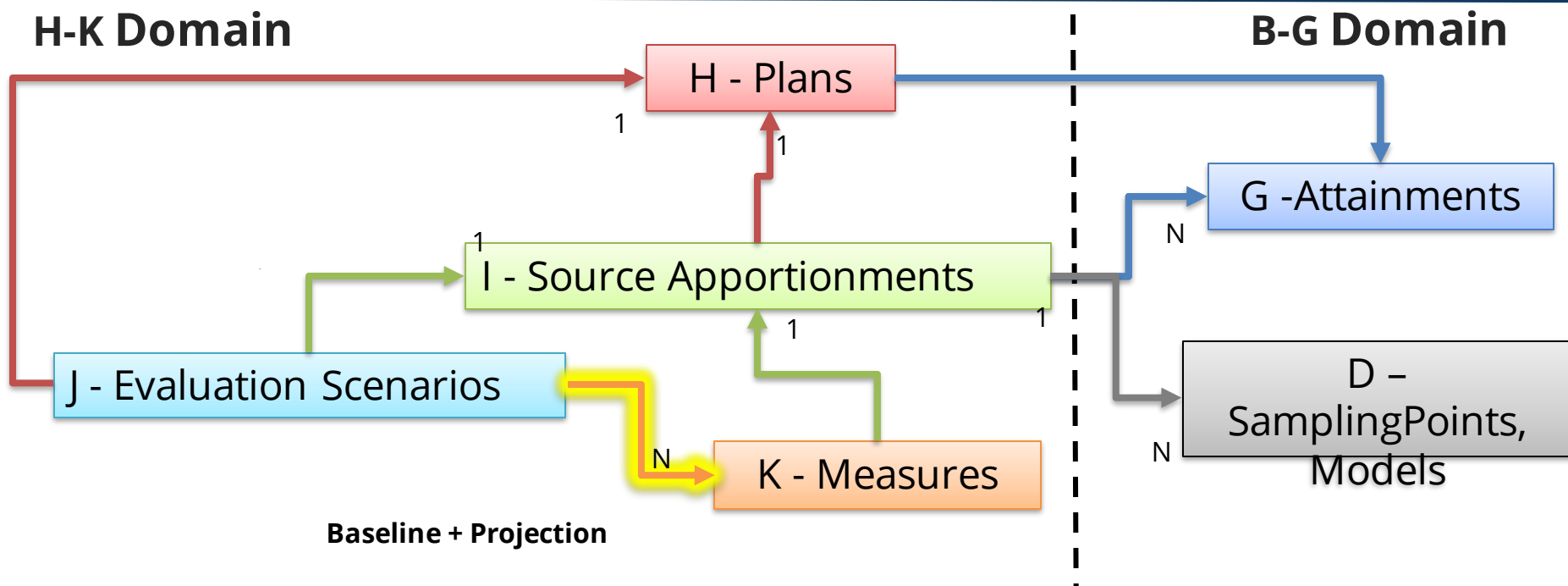
Link Analysis – K to I



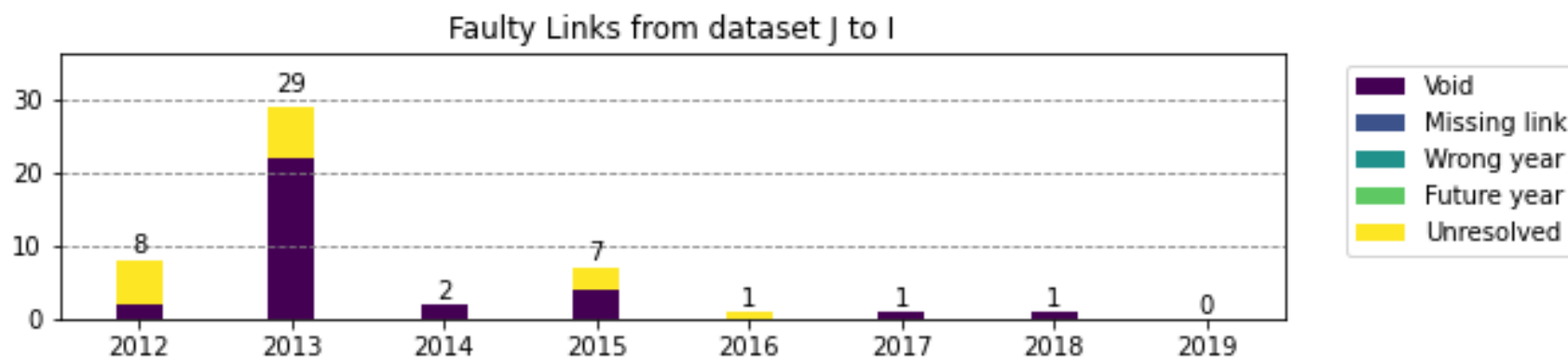
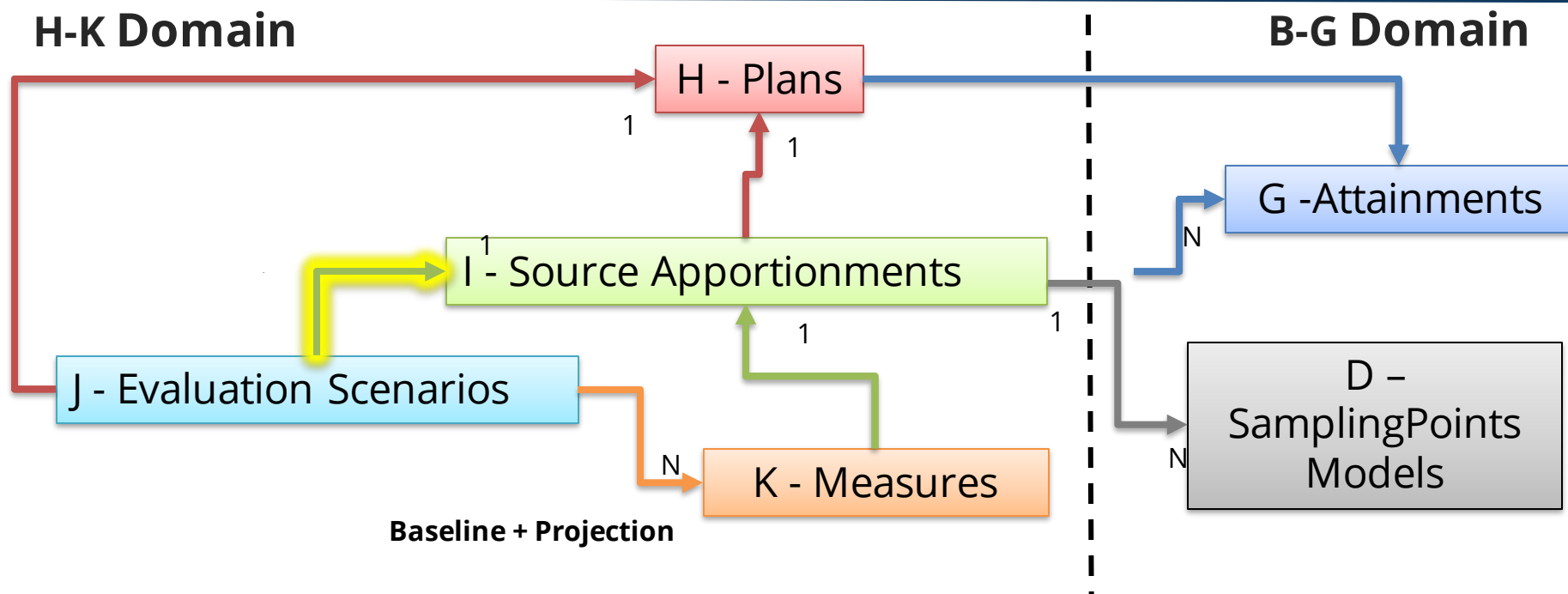
Link Analysis – J to K baseline



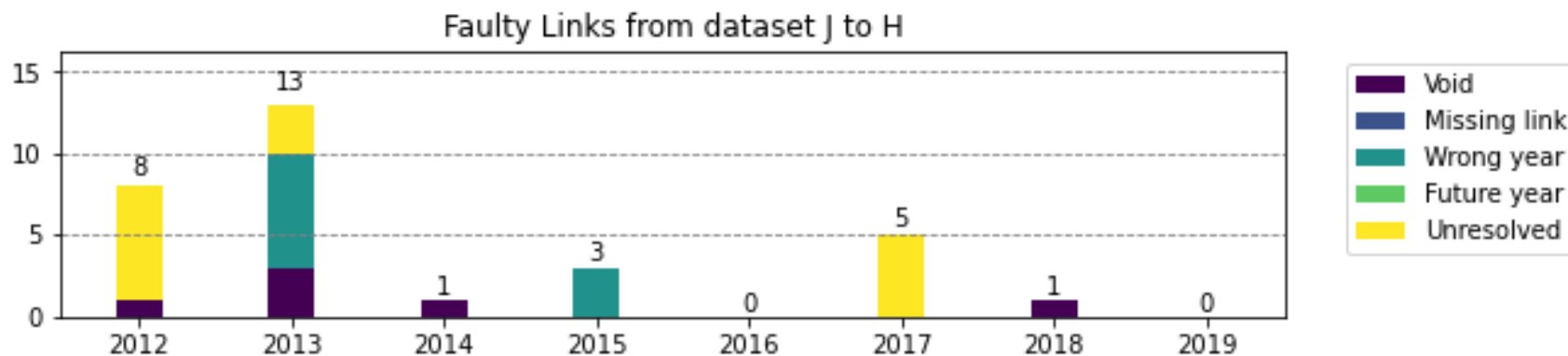
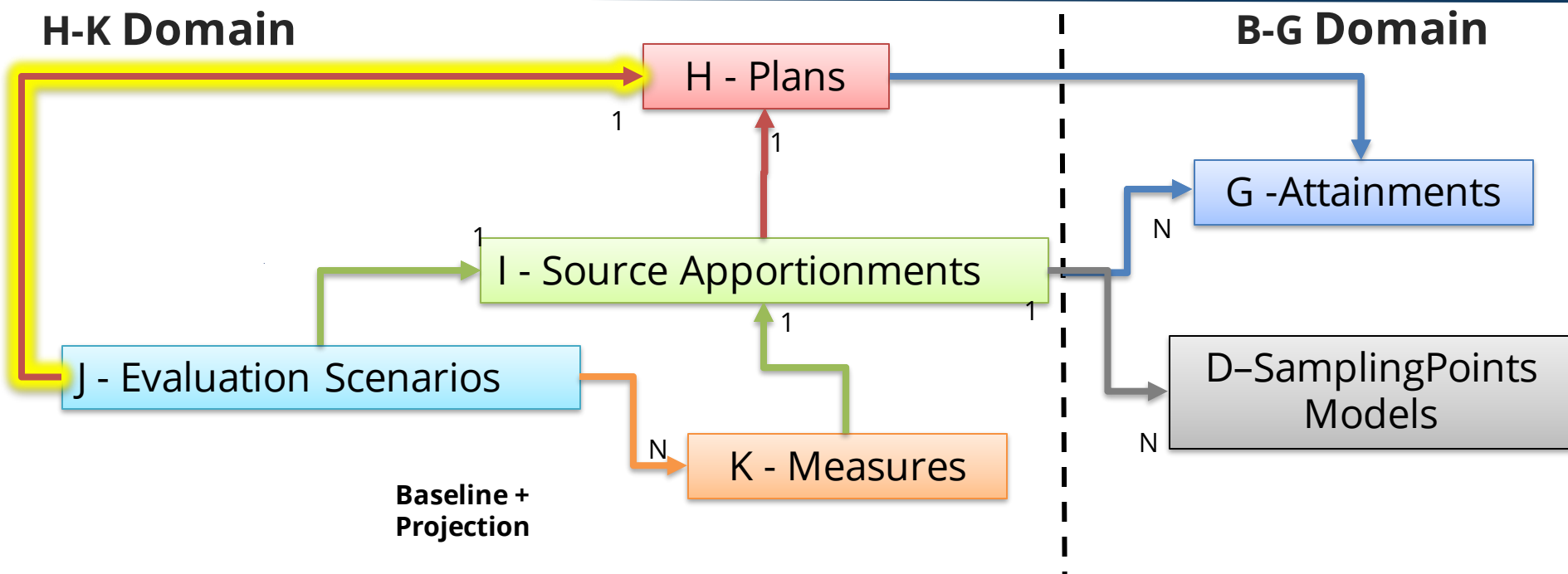
Link Analysis – J to K projection



Link Analysis – J to I



Link Analysis – J to H



Conclusion and Future Plans

- **Further improvement of QA checks**
- **We reached a point where no further improvement can be achieved without resubmission by countries**
- **We can provide Tier 2 analysis to country to target data quality issues**
- **Are the countries supposed to redeliver data?**

Thank You!!